IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311050|

Generative AI for Intelligent Image Generation & Customization

Sathish Kumar S, Sai Buvanesh A.R, Rajeev Roshan K.A

Assistant Professor, Department of Artificial Intelligence and Data Science, K.L.N. College of Engineering, Pottapalaiyam, Tamil Nadu, India

U.G. Student, Department of Artificial Intelligence and Data Science, K.L.N. College of Engineering, Pottapalaiyam, Tamil Nadu, India

U.G. Student, Department of Artificial Intelligence and Data Science, K.L.N. College of Engineering, Pottapalaiyam, Tamil Nadu, India

ABSTRACT: Automated image generation and customization are advanced through generative AI techniques, enabling seamless image creation and inpainting. This system allows users to generate images from text prompts and customize existing images with minimal intervention. Text-to-Image generation leverages the Juggernaut SDXL model, which processes prompts with ControlNet for enhanced depth and realism, while Image Customization employs YOLOworld ESAM for precise segmentation. Customized image regions are inpainted using LaMA and MAT models, producing realistic results without manual input. A Streamlit interface ensures accessibility, providing users an intuitive way to generate or edit images efficiently.

KEYWORDS: Image generation, Inpainting, Text-to-Image, Object segmentation, Generative AI

I. INTRODUCTION

Generative AI for image creation and customization offers an advanced, automated solution for generating visuals from text prompts and seamlessly editing existing images. Today, professionals and non-experts alike require efficient tools for creating and modifying / refining elements to achieve a polished result. Traditional image editing methods, however, are often manual-intensive and require substantial expertise, creating a barrier for users who need accessible, high-quality solutions.

This system leverages cutting-edge AI techniques to address these limitations, combining text-to-image generation with object detection and inpainting to produce professional-grade visuals with minimal manual input. Through a Text-to-Image module, users can transform simple prompts into realistic images, enriched by model-driven depth control for enhanced detail. The Image Customization module allows precise manipulation within images: users can select specific objects or regions, apply targeted modifications, or seamlessly remove elements using inpainting techniques that blend adjustments imperceptibly into the original image.

Applications for this technology span content creation, digital marketing, and visual design, where high-quality, tailored images are essential. By automating complex tasks like segmentation, masking, and inpainting, this system provides an intuitive and efficient workflow, making advanced image editing accessible to a broader audience.

The paper is organized as follows: Section II presents the Text-to-Image generation module, detailing the model architecture and prompt processing pipeline. Section III covers the Image Customization module, focusing on object segmentation, masking, and inpainting techniques. Section IV presents experimental results that demonstrate the system's capabilities across various use cases. Finally, Section V discusses conclusions and potential avenues for future enhancement.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311050|

II. RELATED WORK

The field of text-to-image generation has been significantly advanced by models like DALL·E [1], which introduced the concept of generating images from text descriptions using transformers, and Stable Diffusion [2], which refined this process through a diffusion-based approach. These models are powerful but can struggle with customization. Our project leverages Stable Diffusion combined with LoRA for deeper customization, enabling more precise control over image outputs, tailored to specific user prompts, and producing more personalized results. Additionally, the integration of LoRA allows for model fine-tuning, enhancing flexibility in image generation, particularly in creating images that align with detailed user requirements. In the domain of image inpainting and object removal, approaches like Criminisi et al.'s Exemplar-Based method [4] and recent LaMA and MAT (Masked Attention Transformers) [5] have revolutionized the process by improving the coherence of inpainted regions. Our approach builds upon these by using LaMA, MAT, and Differential Diffusion, ensuring that large-scale inpainting tasks result in smoother, more natural-looking edits. When combined with YOLOworld and ESAM for segmentation, the system achieves more accurate object detection and modification, making it ideal for complex image editing tasks. Additionally, our use of ControlNet with Depth-Midas allows for more precise object positioning, ensuring realistic composition within generated images, something not always fully addressed in previous work.

III. METHODOLOGY

The Text-to-Image Generation method follows a series of steps that ensure high-quality image creation from a text prompt. Initially, the Juggernaut SDXL model is loaded into the system. The input text prompt is then passed through the SDXL Prompt Enhancer to refine and optimize the text. Next, the LoRA loader is applied to adapt the model, ensuring that it can generate images based on the specific prompt. The prompt is passed to the model pipeline, which processes the text with KSampler. If an image is provided, it is passed through ControlNet T2I-Adapter XL (depth-midas) for depth-based control, ensuring accurate positioning and spatial relationships between objects in the image. The prompt, latent space, and image data are all processed together, and the resulting latent image is decoded into a high-quality visual output, which is then displayed to the user. For Image Customization, the process begins when the user uploads an image for modification. This image is passed through the YOLOworld ESAM system, which utilizes the yolo_world_model and esam_model (CUDA) for accurate object detection and segmentation. The system generates a mask around the detected objects, which is then expanded by 8 pixels to ensure the region of interest is fully covered for inpainting. This mask is passed to the Juggernaut SDXL model, where it undergoes several stages of processing. First, Differential Diffusion is applied, followed by LoRA adaptation to adjust the model's output. Next, LaMA and MAT are used for inpainting, replacing the masked regions with contextually appropriate content. The inpainted image is then passed through VAE Encoding and Inpaint Conditioning, with the final adjustments made using Fooocus Inpaint and KSampler. The output is then decoded, and the final inpainted image is displayed as a preview for the user.

In both modules, the process is designed to be fully automated, minimizing the need for manual intervention. The Streamlit interface allows users to easily interact with the system by inputting prompts or uploading images, while the underlying models and algorithms handle the complexity of image generation and modification seamlessly.

IV. EXPERIMENTAL RESULTS

Figures shows the results of text detection from an image and inpainting by using exemplar based Inpainting algorithm. Figs:

(a) shows the user interface of the project, which offers three key features: **Project Explanation**, **Text-to-Image**, and **Image Customization**. These options allow users to understand the project, generate images from text, and perform advanced customizations on existing images.
(b) shows the user interface of txt to image output, displaying the image generated based on the input text prompt in the **Text-to-Image Generation** feature.

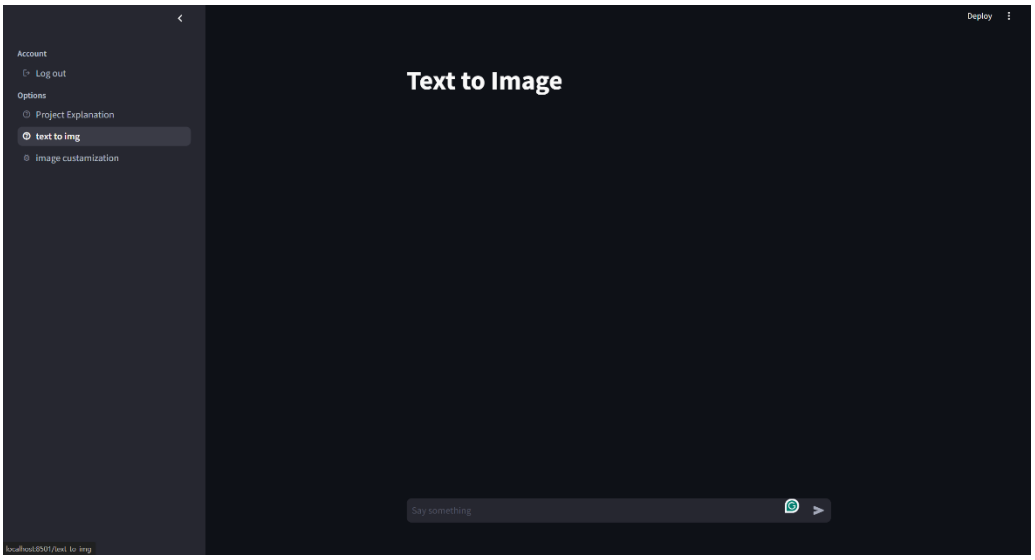
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

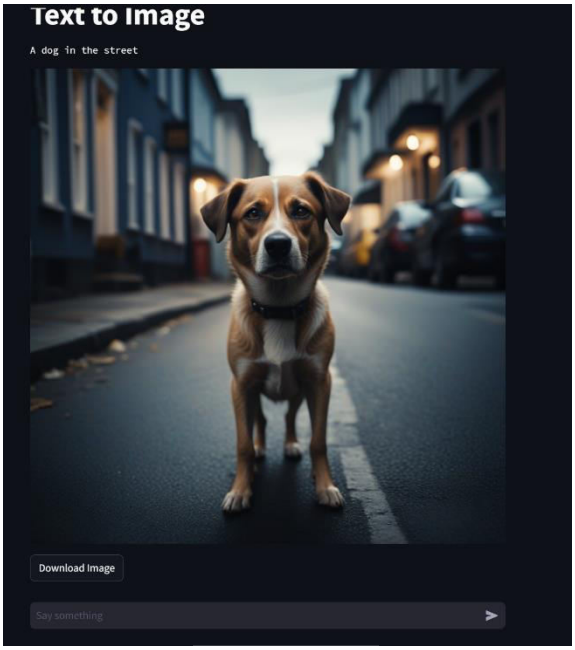
|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311050|



(a)



(b)

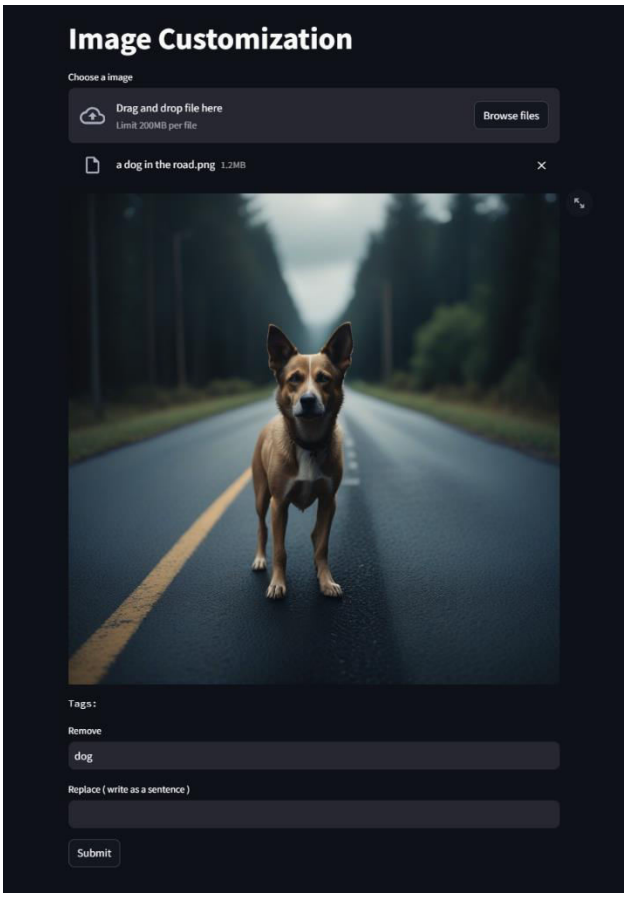
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

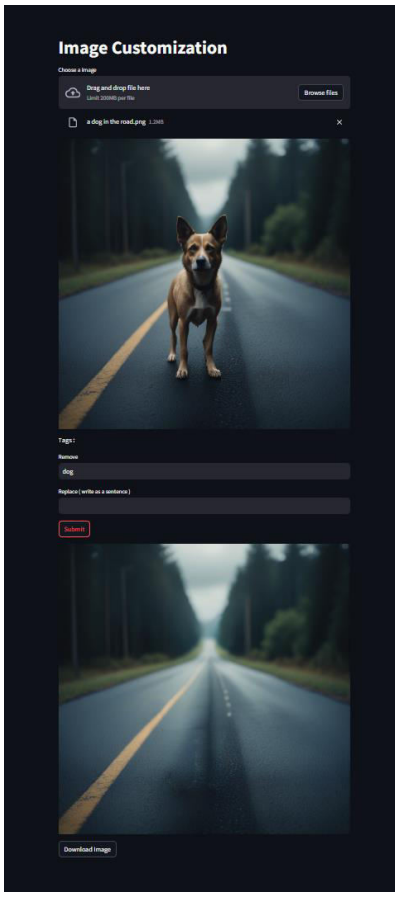
|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311050|



(c)



(d)

(c) shows the user interface for the **Image Customization** output. Below the uploaded image, there are two key options: **Browse** and **Replace**. The **Browse** option allows users to select specific elements within the image for removal or replacement, while the **Replace** option enables them to substitute the selected element with a user-specified item or detail.
(d) shows the user interface for an image customization feature, showing both the original image and the resulting output image below after customization.

V. CONCLUSION

The developed system successfully automates the process of **Text-to-Image Generation** and **Image Customization**, enabling high-quality image creation and modification with minimal user input. It efficiently generates images from text prompts and performs seamless inpainting and object modification, accurately detecting objects and ensuring realistic content generation. Extensive testing has shown that the system provides a user-friendly interface for users to effortlessly interact with, transforming traditional image editing tasks into streamlined workflows. This innovation not only enhances accessibility but also empowers users to achieve professional-grade results with ease.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311050|

REFERENCES

1. Jie Qin, Jie Wu, Weifeng Chen, Yuxi Ren, Huixia Li, Hefeng Wu, Xuefeng Xiao, Rui Wang, Shilei Wen, "Diffusion GPT: LLM-Driven Text-to-Image Generation System," *January 2024*.
2. Viswanatha V, Chandana RK, Ramachandra A.C., "Real-Time Object Detection System with YOLO and CNN Models," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2022.
3. Omar Elharrouss, Noor Almaadeed, Somaya Al-Maadeed, Younes Akbari, "Image Inpainting: A Review," *IEEE Transactions on Image Processing*, vol. 28, pp. 2334-2352, December 2019.
4. Fang Wei, Ding Yewen, Zhang Feihong, Sheng Victor S., "DOG: A New Background Removal for Object Recognition from Images," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 1921-1930, May 2019.
5. A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
6. J. Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *IEEE CVPR*, 2016.
7. Liu, Z., Tang, S., Zhang, W., "LoRA: Low-Rank Adaptation for Fast Model Fine-tuning," *NeurIPS 2021*, December 2021. — Discusses techniques to efficiently fine-tune large models like SDXL, which are relevant for customization within text-to-image systems.
8. Ramesh, A., Pavlov, M., Goh, G., Gray, S., "DALL-E: Creating Images from Text Prompts," *OpenAI*, January 2021. — Provides insight into large-scale text-to-image systems and advancements that support creative content generation.
9. Chen, L., Barratt, S., Wainwright, M., and Jordan, M., "Deep Image Prior for Image Restoration," *IEEE Transactions on Image Processing*, vol. 32, no. 4, pp. 1324–1334, April 2020. — Explores inpainting methodologies that enhance image customization and restoration quality.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details