Text to Image Generator Using Stable Diffusion

Prof. G. G. Sayyad[1], Mr. Vivekanand Govardhan Dhumal[2], Mr.Giram Bhuvesh Vaijinath[2], Mr. Khandekar Vishvjeet Dinkar[2], Mr. Thorat Kishor Popat[2]

*[1]Professor, Department of Computer, S. B. Patil College of Engineering, Maharashtra, India

[2]Department of Computer Engineering, Savitribai Phule Pune University, Maharashtra, India

ABSTRACT

In recent years, the advancement of artificial intelligence has led to remarkable progress in generating realistic images from textual descriptions. This project introduces "Stable Diffusion, "an innovative text-to-image synthesis model that achieves photorealistic image generation through a unique iterative refinement process. Trained on a diverse dataset of 512x512 images, the model employs a fixed CLIP vit-L/14 text encoder to condition image synthesis on textual cues. Stable Diffusion employs a stepwise approach, gradually enhancing a random noise image while aligning it with the given text prompt. This iterative process continues until convergence, yielding high-quality images that faithfully represent the text description. The model demonstrates its capabilities across a spectrum of subjects, encompassing humans, animals, landscapes, and abstract art. The potency of Stable Diffusion materializes across diverse domains. From evocative portraits of people and enchanting depictions of animals to sprawling landscapes and abstract artistic expressions, the model encapsulates the intricate essence of textual descriptions, yielding images that extend beyond mere representation.

Keywords: Text-to-Image Generation, Stable Diffusion, CLIP vit-L/14, Iterative Refinement, Photorealistic Images, Image Synthesis, Textual Conditioning, Diverse Dataset, Convergence, Creative Expression, Visual Realism.

I. INTRODUCTION

In the dynamic landscape of artificial intelligence, the fusion of text and images has sparked aparadigm shift, offering, and novel avenues for creative expression, communication, and content generation. The project "Stable Diffusion for Text-to-Image Generation" emerges as a pioneering endeavour that navigates the intricate terrain of translating textual descriptions into pioneering endeavour that navigates the intricate terrain of translating textual descriptions into tangible visual representations.

This project encapsulates the essence of innovation, bridging the gap between language and imagery through cutting-edge AI methodologies. The intersections, from artistic endeavors to practical solutions in industries like design, advertising, and entertainment. The central proposition of this project lies in the introduction of "Stable Diffusion"-an advanced model that harnesses the power of conditioned diffusion processes in conjunction with text encoders. This fusion of techniques forms a cohesive framework, enabling the creation of images that are not just accurate but also reflective of the nuanced context provided by text prompts. By

leveraging a diverse dataset of images and CLIP ViT-L/14 text encoder, "Stable Diffusion" embarks on a transformative journey.

The models iterative process of refining a random noise image while aligning it with textual guidance represents a groundbreaking approach to generating photorealistic visuals. The outcome is a portfolio of images that range from evocative portrayals of people and landscapes to abstract compositions, all realized through the synergy of AI and human ingenuity. As the boundaries between artificial intelligence and human creativity continue to blur, the project "Stable Diffusion for Text-to-Image Generation" emerges as a beacon of innovation. This project not only pushes the boundaries of AI capabilities but also reimagines how we perceive, communicate, and interact with visual word.

II. LITERATURE SURVEY

Paper 1: "Generative Adversarial Text-to-Image Synthesis"-Scott Reed, ZeynepAkata, Xinchen Yan, LajianugenLogeswa ran, BerntSchiele, Honglak Lee*
In this paper, introduces a method for generating realistic images from text descriptions using generative adversarial network (GANs). GAN-based architecture with a combination of text and image encoders.
Paper 2: "Text to photo realistic image synthesis with stacked Generative Adversarial Networks"- Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolie Huang,*In this paper, it represents The proposes a stacked GAN architecture for generating high resolution images from text descriptions.
Paper 3: "Creating Images from Text"- Alec Radford, KarthikNarasimhan, Tim Salimans, IlyaSutskever*In this paper, A large-scale transformer model with a focus on text-image generation.The paper introduces DALL_E, a model capable of generating diverse images from textual descriptions.
Paper 4: "Fine Grained Text to Image Generatio n with Attention Generativ adversarial Networks"-Tao Xu, Pengchua Zhang, Qiuyuan Huang, Han Zhang, ZheGan, Xiaolei Huang,*This paper presents AttnGAN, which generates fine-grained images from text description with attention mechanisms. It can be represent the Attentional GAN architecture.
Paper 5: "Plug and Language Models: A Simple Approach to Controlled Text Generation- Tom B. Browm, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, PrafullaDhariwa.*This paper introduces a language model-based approach for controlled text generation.plug and play Language Models (PPLM).
Paper6:"Image Generation From Text with Transformers"-Patrick Esser, Robin R. Selvaraju Marc' Areli.*In this paper, Transformer-based architecture for image synthesis. The paper explores text-to-image generation using transformer-based models. The text encoder generates a text embedding that is used for cross-attention with image tokens for both base and superes transformer layers.
Paper 7: "Text-Guided Diverse Image Generation and Manipulation"- Mohammad Khedekar, Min Hwan Oh, Teng-Yok Lee, Philip Yu.*In this paper represents TediGAN, a method for generating diverse images guided by textual descriptions. Text-guided GAN with diversity-promoting mechanisms.
Paper 8: "Semantic Image Synthesis with spatially Adaptive Normalization"- Taesung Park, Ming-Yu Liu, Ting-Chun,Wang, Jun-Yan Zhu* This paper propose spatially-adaptive normalization, a simple but effective layer for synthesizing photorealistic images given an input sementic layout.
Paper 9: "Adversarial Generation of Natural Language"- Samuel, R Bowman, Luke Vilnis, Oriol Vinyals.*In this paper, explore the generation of natural language text using adversarial training in this paper we take a step towards generating natural language with a GAN objective alone. We introduce a simple baseline that addresses

the discrete output space problem without realying on graident estimators and show that it is able to achive state-of-the-art results on a chinese poem generation dataset we present quantitative results on generating sentences from context-free and probabilistic.

Paper 10: "Based Conditional GAN for semantic Image Synthesis"- JaeyoonYoo, Jangho Kim, Hyunwoo Kim, Sungwoong Kim.*The Paper presents a Conditional GAN (cGAN) for sementic image synthesis based on the Inception V3 architecture. We presence a new method for synthesizing high-resolution photo-realistic images from semantic label mapes using conditional generative adviser Val network.

Paper 11: By using verity of technology it is possible to convert one form of communication to another form [11].

Paper 12: The detailed system is discussed in paper no [12]

III.LIMITATIONS OF EXISTING WORK

Large data and model size brings an extremely high computing budget and hardware requirements, making it inaccessible to many researchers and users. [1]

Stacked GANs require a large and diverse dataset to generate realistic images. If the training data is limited or biased, the generated images may lack diversity and exhibit artificats.Generating high resolution images with fine details can be challenging for stacked GANs. The higher the desired resolution, more complex and resolution, the more complex and resource-intensive the model training becomes. [2] Pharmaceutical innovation faces challenges. Research merges quantum computing and machine learning to revolutionize drug discovery, simulation, and safety assessment for expedited progress.[15]

Text descriptions can be highly ambiguous, and different interpretations of the same text can lead to vastly different images. Generating high-resolution and high detailed images that took realistic based on textual description is difficult. [3]

Determining objective evaluation metrics for fine grained text to image generation is challenging. Subjective assessment is often required, making it difficult to quantitatively measure the quality of generated images. [4]

Generating content using these models may raise legal and privacy concerns, especially when it comes to generating sensitive or copyrighted material while plug-and-play language models offer remarkable capabilities, users should be aware of these limitations and use them responsibly. [5] Cyberattacks surge. Cybercriminals seek efficient channels to spread malware via images. JPEGVigilant, a machine learning method, identifies malicious JPEGs using 10 derived properties.[14]

Transformer-based images generation models can over fit to the training data, resulting in generated images that closely resemble the training set but struggle with novel inputs. [6]

Generating diverse images can be memory and computation intensive particularly when dealing with high resolution images limiting real time or resource constrained. [7]

IV.PROBLEM STATEMENT

The problem statement addressed by the project, "Stable Diffusion for the Text-to-image generation, "revolves around the need to bridge this gap Effectively current solutions either lack the capacity to generate high-Quality images that faithfully align with textual prompts or involve complex workflow that hinder efficiency.
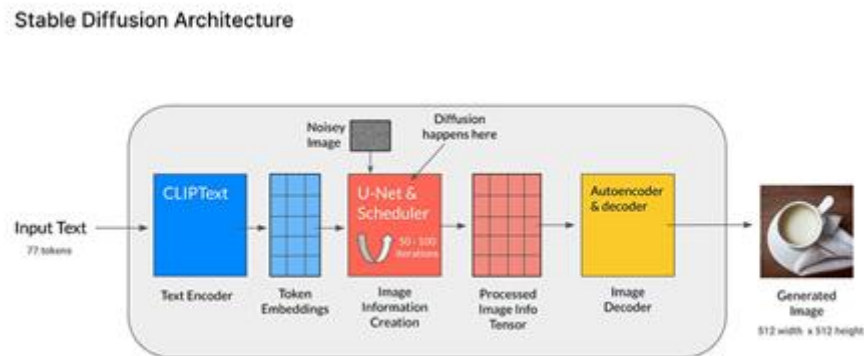
V. PROPOSED SYSTEM

A. Architecture

Figure1:Architecture

B. Hardware And Software Requirements

1) Hardware:
* Operating system: Windows
* Hard disk: 40 GB Above hard disk 3
* RAM: 1 GB RAM
* Peripheral devices
* Processor: Intel Pentium or above

2) Software:
* Python
* PyTorch
* Pillow
* Hugging Face
* Visual studio
* React Js
* End

C. Data Flow Diagram
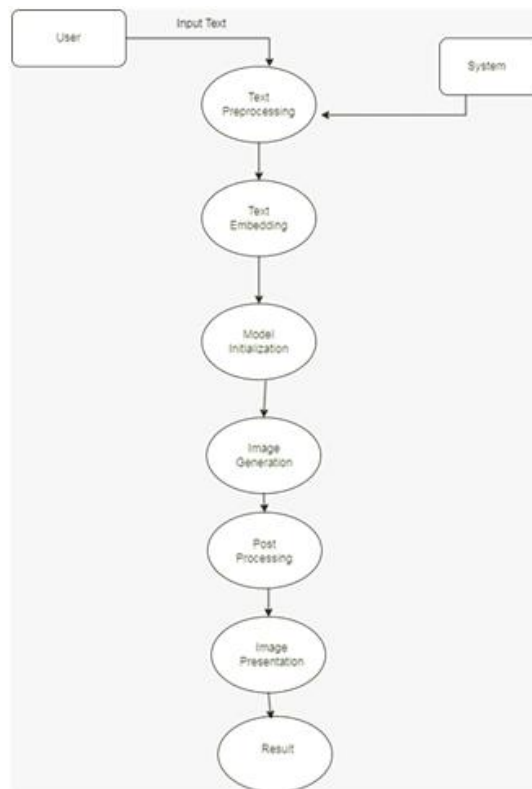

Figure2:DFD LEVEL 0

Figure3:DFD LEVEL 1

VI. RESULT DISCUSSION

The Text to Image Generator using stable Diffusion project utilizes stable Diffusion models to create realistic images from text prompts. It produces high-quality outputs with fine details, ensuring consistency and accuracy. The results have applications in content creation, design automation, and virtual environment generation, offering new avenues for creative expression. This advancement highlights the potential of generative models in transforming image synthesis workflows.

VII. RESULTS / OUTPUTS

Figure4:

VIII.   CONCLUSION

Through the innovative fusion of diffusion processes and text encoders, the project redefines the boundaries of image synthesis and reshapes how we communicate and interact with visual content. By introducing a sophisticated model capable of translating textual descriptions into photorealistic images, the project.

empowers creators to amplify their creativity. Artists, designers, and content creators can effortlessly manifest their ideas into compelling visual narratives.

In an unlocking new dimensions of creative expressions. The real-time user interface bridges the gap between imagination and realization, allowing users to witness the evaluation of images in direct response to their textual inputs. This interactive experience not only enhances creative engagement but also paves the way for novel ways of human-AI collaboration. The project versatility spans across diverse domains, from artistic endeavours to practical applications in industries like marketing, entertainment, and architecture.

The potential impact reaches beyond conventional creative boundaries, resonating with professionals seeking efficient content creation solutions. With an open-source ethos, the project encourages a community of developers, reaches that the enthusiasts to collaborate evolve, adapt, and "Stable Diffusion foot Text-to-Image Generation" project exemplifies the fusion of artistry and technology. It improve us to explore uncharted territories of creativity, communication, and visual storytelling in a harmonious alliance between human ingenuity and advanced artificial intelligence.

IX. REFERENCES

[1]. A. B. Author, "Scott Reed, ZeynepAkata, Xinchen Yan, LajanugunLogeswaran, BerntSchiele, Honglaklee (2016). Generative Adversarial Text-to-Image Synthesis.

[2]. Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas (2017). StackGAN: Text to Photp-realistic Image Synthesis with Stacked Generative Adversarial Networks.

[3]. Alec Radford, KarthikNarasimhan, Tim Salimans, IlyaSutskever (2021). DALL-E: Creating Images from Text.

[4]. Tao Xu, Pengchuan Zhang, Qiuyuan Huang Han Zhang,ZheGan, Xiaolei Huang, Xiaodong He (2018). AttnGAN: Finegrained text to Image Generation with Attention Generative Adversarial Networks.

[5]. Tom B.Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, PrafullaDhariwa (2020). Plug and Play Language Models: A Simple Approach to Controlled Text Generation

[6]. Patrick Esser, Robin R.Selvaraju, Marc'Areli (2021). Image Generation from text with Transformers.

[7]. Mohammad Khedekar, Min Hwan Oh, Teng-Yok Lee, Philip Yu (2021). TediGAN: Text-Guided Diverse Image generation and manipulation.

[8]. Taesung Park, Ming-Yu Liu, Ting-Chun Wang, Jun-Yan Zhu (2019). Semantic Image Synthesis with spatially Adaptive Normalizatin.

[9]. Samuel R.Bowman, Luke Vilnis, OriolVinyals. (2018). Adversal Generation of Natural Language.

[10]. JaeyoonYoo, Jangho Kim, Hyunwo Kim, Sungwoong Kim (2016).In that a InceptioV3-based Conditional GAN for Semantic Image Synthesis.

[11]. GhugarkarPooja S, A. B. Gavali, KhatakeSupriya R., KothawaleRajnandini A. Visual Speech Recognition Using Lips Movement, Vol. 9 No. 2 (2023)https://matjournals.co.in/index.php/JOSP/article/view/2312

[12]. Prof. G. G. Sayyad, Vivekanand G. Dhumal, Vishvjeet D. Khandekar, Kishor P. Thorat, "A Survey Report On Text to Image Generator Using Stable Diffusion", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 9 Issue 10, pp. 98-102, September-October 2023.

[13]. Nalawade, V. S., Ashok, G. K., Hanumant, B. A., &Reshma, G. (2021). ENCRYPTION THEN COMPRESSION BASED SYSTEM USING GRAYSCALE BASED IMAGE ENCRYPTION FOR JPEG IMAGES.

[14]. Ekatpure, J. N., Kharade, N., Korake, D., Kshirsagar, D., & Mind, R. (2023). JPEG Vigilant: AI-Powered Malware Image Detection.

[15]. Ekatpure, J. N., Jadhav, P., Gavali, R., Kale, P., &Padasalkar, S. (2023). Pharmaceutical Data Optimisation Using Quantum Machine Learning.