IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

Case Study about the Image Generator from Prompt using Stable Diffusion Model

Mr. Landage P. S.[1], Mr. Landage M.N.[2]

HOD, Computer Technology, Shivaji Polytechnic Atpadi, Atpadi, India[1]

Lecturer, Computer Technology, Shivaji Polytechnic Atpadi, Atpadi, India[2]

Abstract: In Artificial Intelligence, automatically describing what's there in a photograph or image has always been a context of study. This paper includes the implementation of Automatic Image Generator using and -STABLE DIFFUSION models. It combines recent studies of machine translation as well as computer vision. The datasets used were Stable diffusion for evaluation of the performance of the described model. Through the scores, one can apart the generated Images as good Images and bad Images. Main applications of this model include usage in virtual assistants, for image indexing, for social media, for visually impaired people, recommendations in editing applications and much more

keywords*:* Stable Diffusion, Denoise, Image Generator, Prompt Etc.

I. INTRODUCTION

Diffusion models are a type of generative model that is trained to denoise an object, such as an image, to obtain a sample of interest. The model is trained to slightly denoise the image in each step until a sample is obtained. It first paints the image with random pixels and noise and tries to remove the noise by adjusting every step to give a final image that aligns with the prompt. Image Generator models is based on encoder-decoder architecture which use input vectors for generating valid and appropriate Images. This model bridges gap between natural language processing as well as computer vision. It's a task of recognizing and interpreting the context described in the image and then describing everything in natural language such as English. Our model is developed using the two main models i.e. STABLE DIFFUSION. The encoder in the derived application is which is used to extract the features from the photograph or image and STABLE DIFFUSION works as a decoder that is used in organizing the words and generating Images. Some of the major applications of the application are self-driving cars wherein it could describe the scene around the car, secondly could be an aid to the people who are blind as it could guide them in every way by converting scene to Image and then to audio, CCTV cameras where the alarms could be raised if any malicious activity is observed while describing the scene, recommendations in editing, social media posts, and many more.

II. RELATED WORK

The application is merged with two main architectures and which describes attributes, relationships, objects in the image and puts into words. is an extractor that extracts features from the given image. STABLE DIFFUSION will be fed with the output of the and following it will describe and generate a Image. is a Stable diffusion which process the data having the input shape similar to two-dimensional matrix. Model has many layers including input layer, Convo Layer, Pooling Layer, Fully-connected layers, Softmax, and Output layers. Input layer in is an image. Image data is presented in form of 3D form of matrix. Convo Layer also known as feature extractor where it performs the convolutional operation and calculate the dot products. Image is sub layer in Convo layer that converts all negative values to zero. Pooling layer is one where the volume of the image is being reduced once the convolution layer executes. A fully-connected layer is connection layer that connects one neuron in a layer to other neuron in other layer involving neurons, biases and weights. Softmax layer is used for multi- classification of objects where using formula the objects are classified. Output layer is last layer at model and has the encoded result to be fed to STABLE DIFFUSION model, where output of previous step is fed to on-going step. STABLE DIFFUSION (Long Short-Term Memory) is an extended version of that are used to the predict the sequence based on the previous step where in it remembers all the steps and also the predicted sequence at every step. It grasps the required information from the processing of inputs as well as forget gate and also it does remove the non-required data

III. FLOW OF THE PROJECT

- Importing The Libraries
- Configuring The GPU Memory To Be Used For Training Purposes
- Load The Model

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

- Extracting Features
- Building The Stable Diffusion Model
- Running The Stable Diffusion Model
- Providing Prompt
- Generating The Image And Download With PNG, JPG etc.

III. DATA FLOW DIAGRAM

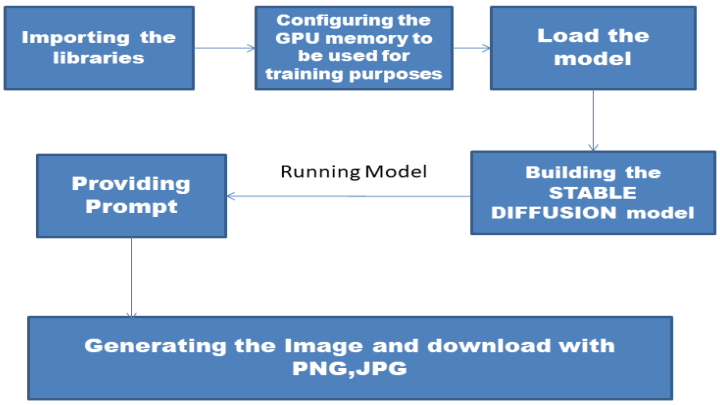

Fig: 1 Data Flow Control Diagram

IV. PROPOSED ARCHITECTURE OF STABLE DIFFUSION MODEL

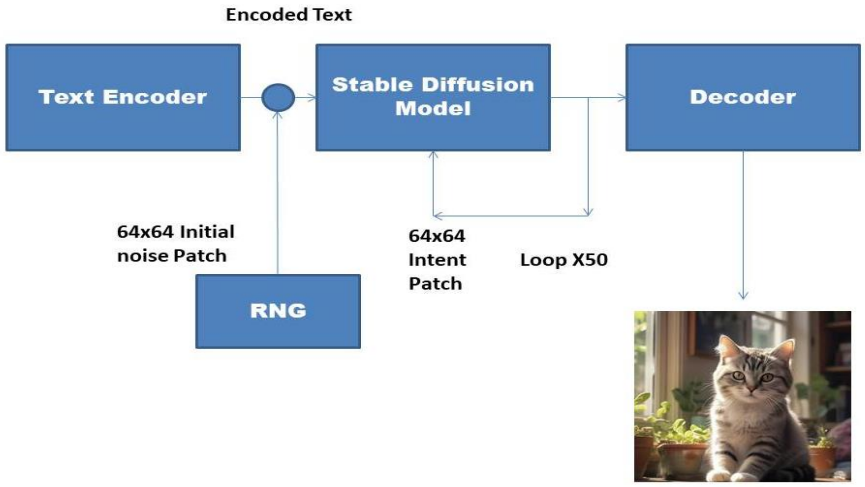

Fig: 2 Architecture of stable diffusion model using prompt

V. SYSTEM REQUIREMENTS

1. OS: Windows 8 and above, Recommended: Windows 10.
2. CPU: Intel processor with 64-bit support
3. GPU: Graphics card 4 GB
4. Disk Storage: 8GB of free disk space.

VI. LIBRARIES USED

• Tensorflow: It's an open-source library that supports deep learning using Python etc. frameworks.
• Pillow: Pillow is a Python Imaging Library (PIL) that adds support for opening, manipulating, and saving images.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

• Numpy: To work with arrays, Numpy library is used.
• Matplotlib: Library to create static and animated visualizations in Python framework.

VII. RESULTS

○ Project Sample source code:

Chatbot.py

```
C: > Users > omkar > OneDrive > Desktop > rushi > Chatbot.py > generate_image
29    def generate_image(pipe, prompt):
47        print(f"Generated image size: {image.size}")
48
49        # Save the image with a filename based on the prompt
50        image_path = f"{prompt.replace(' ', '_')}.png"
51        image.save(image_path)
52        print(f"Image saved as {image_path}")
53
54        return image
55
56    # Function to handle user input and chatbot logic
57    def chatbot():
58        print("Chatbot: Hi! I can generate images from your text descriptions. Type 'exit' to stop.")
59
60        # Load the model
61        pipe = load_model()
62
63        while True:
64            # Get user input
65            user_input = input("You: ")
66
67            if user_input.lower() == 'exit':
68                print("Chatbot: Goodbye!")
69                break
70
71            # Generate image based on user input
72            image = generate_image(pipe, user_input)
73
74            if image is not None:
75                # Save the image with a filename based on the prompt
76                output_path = os.path.join(os.getcwd(), f"{user_input.replace(' ', '_')}.png")
77                image.save(output_path)
78                print(f"Chatbot: Image generated and saved as {output_path}")
79            else:
80                print("Chatbot: No image generated. Please try a different prompt.")
81
82    if   name   == "  main  ":
```

○ Run the Source code:

```
C:\Windows\System32\cmd.e
Microsoft Windows [Version 10.0.22631.4751]
(c) Microsoft Corporation. All rights reserved.

C:\Users\omkar\OneDrive\Desktop\rushi>python chatbot.py
Chatbot: Hi! I can generate images from your text descriptions. Type 'exit' to stop.
Loading Stable Diffusion model...
Using cuda for model inference.
Couldn't connect to the Hub: (MaxRetryError('HTTPSConnectionPool(host=\'huggingface.co\', port=443): Max retries exceede
d with url: /api/models/CompVis/stable-diffusion-v1-4 (Caused by NameResolutionError("<urllib3.connection.HTTPSConnectio
n object at 0x000002A6FC9B4170>: Failed to resolve \'huggingface.co\' ([Errno 11001] getaddrinfo failed)"))'), '(Request
 ID: 21eac314-1705-45b6-8a4f-6133fdcd7cdc)').
Will try to load from local cache.
Loading pipeline components...: 100%|                                                  | 7/7 [00:16<00:00,  2.37s/it]
Model loaded successfully on cuda
You: cat
Generating image for prompt: 'cat'
100%|                                                                                  | 50/50 [04:02<00:00,  4.84s/it]
C:\Users\omkar\AppData\Local\Programs\Python\Python312\Lib\site-packages\diffusers\image_processor.py:147: RuntimeWarnin
g: invalid value encountered in cast
  images = (images * 255).round().astype("uint8")
Image generated successfully.
Generated image size: (512, 512)
Image saved as cat.png
Chatbot: Image generated and saved as C:\Users\omkar\OneDrive\Desktop\rushi\cat.png
You:
```

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

- Project Output:

VIII. CONCLUSION AND FEATURE WORK

The model has been successfully trained and tested to generate the valid captions for the loaded images. The proposed model is based on Stable Diffusion to generate the Images where Stable diffusion works as an encoder some of the future enhancements would include describing the captions based on multiple targets. The generated caption should be in variety of languages.

REFERENCES

[1] Using Stable Diffusion with Python by Andrew Zhu (Shudong Zhu) and Matthew Fisher.
[2] Mastering Digital Art with Stable Diffusion, which provides self-study tutorials with working code
[3] R. Subash November 2019: Automatic Image Captioning Using Convolution Neural Networks and LSTM.
[4] Seung-Ho Han, Ho-Jin Choi (2020): Domain-Specific Image Caption Generator with Semantic Ontology.
[5] Pranay Mathur, Aman Gill, Aayush Yadav, Anurag Mishra and Nand Kumar Bansode (2017): Camera2Caption: A Real-Time Image Caption Generator
[6] Simao Herdade, Armin Kappeler, Kofi Boakye, and Joao Soares (june 2019): Image Captioning: Transforming Objects intoWords.
[7] Manish Raypurkar, Abhishek Supe, Pratik Bhumkar, Pravin Borse, Dr. Shabnam Sayyad (March 2021): Deep learning-based Image Caption Generator
[8] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan (2015): Show and Tell: A Neural Image Caption Generator
[9] Jianhui Chen, Wenqiang Dong, Minchen Li (2015): Image Caption Generator based on Deep Neural Networks