AI IMAGE GENERATION USING TEXT PROMPT

Brahmdeep Singh
Big Data Analytics - Apex Institute of Technology
Chandigarh University
Mohali, India
20bcs3846@cuchd.in

Abstract- Artificial intelligence (AI) has witnessed a paradigm shift in image generation techniques recently, with an emphasis on Latent ODEs and stable diffusion models within the framework of Reversible Generative Models (LoRA). This literature review focuses on significant research and writings that contribute to our growing understanding of this creative tactic. The study highlights how important reversibility is for capturing the underlying structure of data, beginning with the foundations of reversible generative models. Real Non- Volume Preserving (Real NVP) models and diffusion probabilistic models are offered as predecessors to lay the groundwork for the stable diffusion process.

This article investigates the advantages of diffusion models over Generative Adversarial Networks (GANs) in image synthesis, emphasizing the application of diffusion models to high-fidelity image production. As dynamic representations in the latent space, latent ODEs are provided as a flexible and expressive way to characterize complex interactions across time. The temporal component is further studied with insights from the application of Latent ODEs in irregularly-sampled time series and works on latent space models for dynamic networks.

With an emphasis on dependable dissemination APIs, API integration is offered as a practical means of utilizing these intricate models in various applications. The research emphasizes the changeable generation capabilities of the stable diffusion LoRA pipeline, which allow the fine-tuning of characteristics such as diffusion steps and latent space dynamics. The literature highlights the method's versatility by demonstrating its application to artistic depiction, data augmentation, and other areas.

Keywords- Artificial Intelligence, Image Generation, Stable Diffusion Models, Low-Rank Adaptive (LoRA), API Integration.

I. PROBLEM DEFINITION

The current state-of-the-art in image generation is limited by the capacity of traditional machine learning models to generate images that are both high quality and diverse. Additionally, the development of web applications that can utilize these machine learning models is often complicated by the need to integrate multiple technologies and frameworks. To address these limitations, there is a need for a web application that utilizes an AI-based image generation model and is built using a comprehensive web development stack. The MERN stack provides a full-stack development framework that combines MongoDB, Express.js, React.js, and Node.js to simplify web application development, and can be used to build a web application that can generate high-quality and diverse images using an AI-based model. Therefore, the problem to be addressed is the development of an AI-based image generation application that utilizes the MERN stack to provide a seamless and comprehensive solution for generating high-quality and diverse images.

II. PROBLEM OVERVIEW

The field of image generation has advanced significantly in recent years, with AI-based models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) being used to generate high-quality images. However, the development of web applications that can utilize these models is often complex and time-consuming, as it requires the integration of multiple technologies and frameworks. The MERN stack provides a comprehensive solution for web development, combining MongoDB, Express.js, React.js, and Node.js to provide a full-stack framework for building web applications. By integrating an AI-based image generation model with the MERN stack such as LoRA (Low-Rank Adaptation), it is possible to create a web application that can generate high-quality and diverse images with ease.

This problem overview focuses on the development of an AI-based image generation application using the MERN stack, which can simplify the process of creating web applications that generate images using machine learning models. The development of such an application has the potential to enable a wider range of developers to create image generation applications, and to accelerate the advancement of the field of image generation.

III. LITERATURE REVIEW

Image Generation Using AI (studies)

The field of image generation has seen significant advancements in recent years, with AI-based models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) being used to generate high-quality images. However, integrating these models into web applications has proven to be complex and time-consuming. In response, developers have turned to comprehensive web development stacks, such as the MERN stack, to simplify the process of building web applications.

LoRA is a fine-tuning method that was introduced in a research paper titled "Layer-wise Relevance Allocation for

Fine-grained Analysis of Transformer Models" by Jain and Wallace (2020). It focuses on understanding the contribution of different layers of a pre-trained language model in making predictions for downstream tasks.

D. P. Kingma and colleagues' "Reversible Generative Models" (2016) [1], In order to lay the foundation for comprehending the significance of reversibility in generative modeling, this paper presents the idea of reversible generative models.

L. Dinh and associates' "Real NVP Density Estimation" (2016) [2], The Real NVP (Real Non-Volume Preserving) model is a type of generative model that is based on the concepts of reversible generative models. Real NVP must be understood in order to appreciate generative modeling's reversible nature.

As stated in "Diffusion Models Beat GANs on Image Synthesis" by J. Ho et al. (2020) [3], Unlike Generative Adversarial Networks (GANs), the effectiveness of diffusion models for picture synthesis is examined in this paper. Understanding the benefits of diffusion models forms the foundation of the stable diffusion LoRA pipeline.

Chen, Y., et al. (2018) [4], "Latent ODEs for Irregularly-Sampled Time Series "In this article, latent ODEs are introduced and their use in modeling time series data with irregular sampling is discussed. It establishes the foundation for comprehending how ODEs are applied in hidden spaces.

P. J. Hoff and colleagues' "Latent Space Models for Dynamic Networks" (2012) [5], Insights into the application of latent spaces for modeling dynamic processes—which are pertinent to the temporal elements of LoRA models— are offered by this paper's exploration of latent space models in the context of dynamic networks.

J. Sohl-Dickstein et al. (2015) [6], "Diffusion Probabilistic Models" The theoretical underpinnings and practical uses of diffusion probabilistic models are covered in this paper. It offers a more thorough comprehension of the picture creation process' diffusion.

N. Shazeer et al. (2016) [7], "Training Generative Adversarial Networks with Limited Data" A common question in the setting of stable diffusion models is how to train generative models with minimal data. This work tackles these issues.

IV. Background

- Artificial intelligence (AI) has a long history dating back to the mid-20th century. The early pioneers of AI were mostly computer scientists and mathematicians who developed early machine learning algorithms, symbolic reasoning systems, and expert systems.
- One of the earliest and most influential AI programs was the Logic Theorist, developed by Allen Newell and Herbert A. Simon in 1955. The program used symbolic logic to prove mathematical theorems and was a significant breakthrough in the field.
- In the 1960s and 1970s, researchers began developing machine learning algorithms, which enabled computers to learn from data. One of the most famous algorithms developed during this time was the perceptron, developed by Frank Rosenblatt in 1957, which is a type of artificial neural network.
- In the 1980s, expert systems became popular, which were computer programs that used knowledge bases and rules to make decisions. These systems were used in a variety of applications, such as medical diagnosis, financial planning, and oil exploration.
- In the 1990s and 2000s, researchers began developing more advanced machine learning algorithms, such as deep learning, which uses artificial neural networks with many layers to learn from data. These algorithms have been used in a wide range of applications, such as image and speech recognition, natural language processing, and autonomous vehicles.
- Today, engineers continue to work on improving AI algorithms and developing new applications for the technology. They are also working on addressing ethical and social issues related to AI, such as bias, privacy, and the impact of automation on the workforce.

V. Proposed Framework

User Interface: React was used to create a dynamic and responsive interface for users. Size, colour, texture, and other parameters can all be entered via the user-friendly interface. The created image can be previewed and downloaded directly from the interface. A scalable and efficient platform for creating web applications, Node.js and Express.js form the basis of the backend server. In order to generate and return the photos, the server communicates with the database and the AI algorithms. The application's AI algorithms employ deep learning strategies like Stable Diffusion and LoRA to create one-of-a-kind, high-resolution photographs. Libraries like TensorFlow and Keras are used to include the AI algorithms into the backend server.

The programme supports multiple image file types, including PNG, JPEG, and SVG, for saving and downloading. The server controls all data transfers, guaranteeing safe and efficient file storage. Protection: Multiple layers of protection, including encryption, strong authentication, and data validation, are built into the programme from the ground up. The application can be deployed to many cloud platforms to guarantee scalability and availability, including Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

Stable Diffusion and LoRA

1. Stable Diffusion

In the context of AI, "stable training methods" generally refer to techniques or approaches that improve the stability and convergence of deep learning models during training. The stability of training is crucial for preventing issues like exploding or vanishing gradients, mode collapse, or training divergence.

2. LoRA (Low Rank Adaptation Model)

The LoRA method involves computing relevance scores for each layer of the language model based on the gradients flowing through that layer during training. These relevance scores indicate the importance of each layer in the overall prediction process. By considering the relevance scores, one can gain insights into how different layers of the model contribute to the task at hand.

During fine-tuning using LoRA, the relevance scores are used to guide the learning process. Layers with high relevance scores are given more importance and updated more during the fine-tuning process, while layers with low relevance scores are updated less. This allows the fine-tuning process to focus on the most relevant parts of the model and potentially improve its performance on the target task.

VI. Methods

6.1. Object-Oriented Framework

Data Layer:

The data layer is responsible for handling data models and database interactions using MongoDB. MongoDB is a popular NoSQL database that stores data in JSON-like documents, making it an ideal choice for working with JavaScript-based applications like MERN. In this layer, the data models define the structure of the data that will be stored in the database, including images and associated metadata. Mongoose, a MongoDB object modeling tool, is used to define the data models and interact with the database.

API Layer:

The API layer is responsible for handling HTTP requests and responses using the Express.js framework. Express.js is a popular Node.js web application framework that simplifies building web applications by providing a set of tools for handling HTTP requests, routing, and middleware. In this layer, the endpoints for creating and retrieving images are defined, along with any authentication and authorization logic needed to secure the application.

Safetensors is a new simple format for storing tensors safely (as opposed to pickle) released by Hugging Face and that is still fast (zero-copy). For its efficiency, many stable diffusion models, especially Lora models are released in safetensors format. You can find more its advantages from huggingface/safetensors and install it via pip install.

6.2. Data Processing

Dataset Transformation:

Prepare your dataset for the specific downstream task you want to fine-tune the language model for. this involves tasks such as data collection, cleaning, and preprocessing common preprocessing steps include tokenization, lowercasing, removing stop words, and handling special characters.

Tokenization:

Convert the raw text data into a sequence of tokens that the language model can understand. Tokenization splits the text into individual units, such as words or subworlds, which serve as input to the model. It's important to use the same tokenization scheme as the one used during pretraining to maintain consistency.

Data Encoding:

Convert the tokenized text into numerical representations that can be processed by the language model. This typically involves mapping tokens to unique integer IDs and creating input tensors or arrays.

Train-Validation Split:

Split the dataset into training and validation subsets. The training set is used to fine-tune the language model, while the validation set helps in monitoring the model's performance and tuning hyperparameters.

Fine-tuning with LoRA:

During the fine-tuning process, you train the language model on the task-specific dataset using backpropagation and gradient descent. You compute the relevance scores using LoRA based on the gradients flowing through each layer during training. The relevance scores guide the learning process by allocating more updates to layers with higher relevance scores. Examples of hyperparameters include learning rate, regularization strength, number of hidden layers, and batch size.

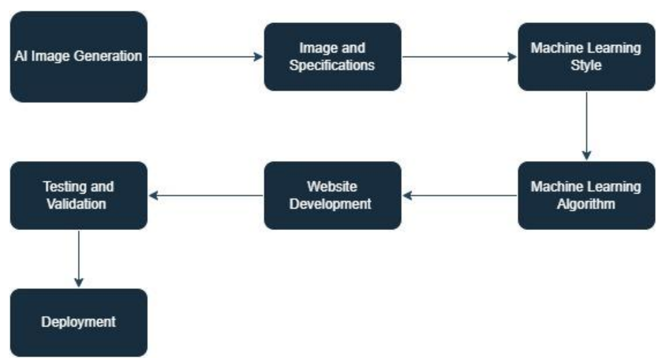

Fig – Flowchart of the program

Evaluation:

Once the fine-tuning process is complete, evaluate the performance of the fine-tuned model on a separate test set. This provides an unbiased measure of how well the model generalizes to new, unseen data.

Hyperparameter Tuning:
Hyperparameter tuning is the process of selecting the optimal combination of hyperparameters for a machine learning model to achieve the best performance on a given dataset.

VII. RESULTS

We started with two distinct prompts and ran them through the AI model, which resulted in the generation of the two images that are presented below.

Here, we have taken a Trained Tags of Model which belongs to HiPoly 3D LoRA Model – v2.0 Showcase. Which is model full of anime poses, faces, body structure and mainly focusing on cartoon aspects of the image.

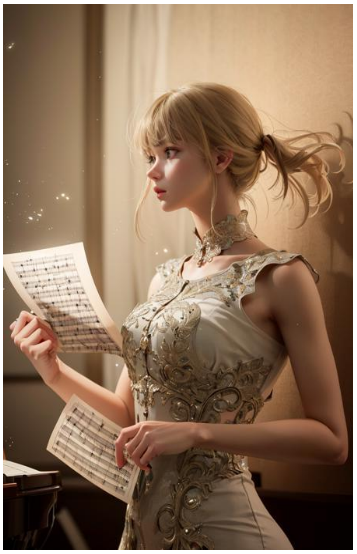

Fig – AI image Produced

Prompt:
{
<lora: hipoly3DModelLora_v20:0.5>, 3d, realistic, masterpiece, best quality, 1girl, Musician, holding an instrument, passionate expression, sheet music, stage or practice room background, musical notes floating around, rim lighting, side lighting, cinematic light, ultra-high res, 8k uhd, film grain, best shadow, delicate, RAW, light particles, detailed skin texture, detailed cloth texture, beautiful detailed face, intricate details, ultra detailed

Negative prompt
[: EasyNegative:0.5], extra fingers, fewer fingers, face paint, mole.
}

Next, we take another Model namely – Gemini Mix which focuses on producing images which are more realistic, high-resolution images which present realism and adds life to the images. The low sampler and CFG gives more realistic results for the picture generated. This model also includes poses, faces, body structure / animal structure and additionally gives flexibility for more than one type of artistic approach to the images being generated. The negative prompts are also an excellent examples of how well an image can be produced and how to negate the bad aspects on the image.

AI has revolutionized the field of image generation, bringing forth a new era of realism and visual fidelity. With advanced algorithms and neural networks, AI models have the ability to produce stunningly realistic pictures that challenge the boundaries of imagination. Through the marvels of AI, virtual brushstrokes come alive, breathing realism into pixels and transforming the realm of generated pictures.

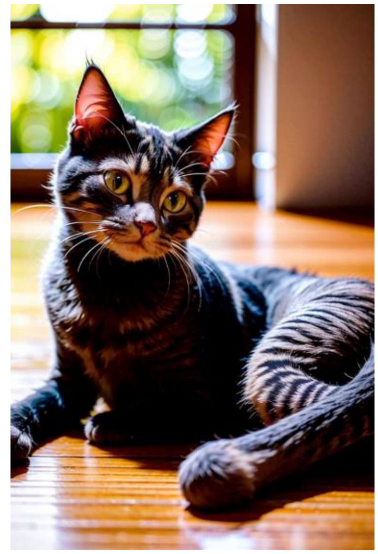

Fig – AI image Produced

Prompt:
{
<(photorealistic:1.4), (extremely intricate:1.2)>, < (exquisitely detailed skin), cinematic light, ultra-high res, 8k uhd, intricate details, Movie Light, Oil Painting Texture, film grain, dreamlike, perfect anatomy, best shadow, delicate, RAW>, no humans, (((cat, rainbow cat))), photo background, animal focus, animal, (((lying on the floor))), photo inset, (((closed eyes))), cat with nine tails, sleep, light brindle,

Negative prompt
(Worst quality, normal quality, low quality:1.7), watermark, signature, text, freckles, low res, monochrome, grayscale, bad proportions, ugly, dirt on face
}

The outcomes of employing Stable Diffusion have been encouraging. Artificial intelligence (AI) picture generating models have the potential to significantly impact many fields and fields of application, including the arts and design, product visualization, and advertising, with further development.

VIII. CONCLUSION

Effect of Fine tuning can be seen in this image where each image is 500 steps of another step. Trained with 9 images for one clip

("female game character, in a steampunk city, 4K render, trending on art station, masterpiece".)
You can see that with 2500 steps, you already get somewhat good results.
By the nature of LoRA, one can interpolate between different fine-tuned models by adding different matrices

The MERN stack AI Image Generation website project is a promising solution to generate high-quality images using artificial intelligence and machine learning techniques. The project utilizes a MERN stack framework and involves preprocessing, segmentation, and postprocessing techniques to enhance the quality of the generated images. With further development and refinement, the project has the potential to become a valuable tool for a wide range of industries and applications.

It's worth mentioning that the LoRA method is focused on relevance scoring and does not explicitly affect the data processing steps. The data processing pipeline for fine-tuning language models remains similar to traditional approaches, while LoRA comes into play during the training phase to allocate updates to different layers based on their relevance.

IX. FUTURE WORK

While DALL-E has showed promising results in producing high-quality graphics in response to textual prompts, there remains potential for advancement in this area. Generative models that can produce images at a higher rate and with higher quality could be one area of research. Adding sensory data, like as video or audio, might be explored further to produce more lifelike and varied results.

DALL-E can create graphics in response to textual prompts, although it may not always grasp the whole semantic context of the text. It would be interesting to see how picture generating models could benefit from a deeper comprehension of language and context in the future. This may involve the use of knowledge graphs and semantic networks in addition to more advanced language models.

Dataset Generation: Artificial intelligence model outputs are highly sensitive to the standard of the training data used to create them. Creating novel and varied datasets for training picture creation models could be a focus of future research. Methods of crowdsourcing to collect more diverse and representative photos, or methods of synthetic data synthesis to enrich existing datasets, are two examples.

While the theoretical prowess of picture generating models like LoRA and Stable Diffusion is undeniable, there is still a great deal of room for improvement in their practical applications. Art, design, advertising, and product visualization are just few of the potential areas where picture creation models could be applied in the future. The integration of such models into preexisting software programmers or the creation of new tools and interfaces for their usage by non-experts are both viable options.

It's important to note that LoRA, specifically designed for fine-grained analysis in language models, may not be directly applicable to image generation tasks. However, there may be other relevance allocation techniques or approaches specific to image generation tasks that can be explored.

REFERENCES

[1] Pivotal tuning for latent-based editing of real Roich, Daniel and Mokady, Ron and Bermano, Amit H and Cohen-Or, Daniel ACM Transactions on Graphics (TOG)

[2] The World Bank. Global bilateral migrations, 2011. URL https://databank. worldbank.org/reports.aspx?source=global-bilateral-migration#.

[3] The World Bank. World development indicators, 2020. URL https:// databank.worldbank.org/source/world-development-indicators.

[4] Jakob Bijak and Arkadiusz Wisniowski. Statistical modeling of international migration flows. 2011.

[5] John Blad. Gated recurrent unit, 2 2018. URL https://en.wikipedia.org/ wiki/Gated_recurrent_unit#/media/File:Gated_Recurrent_Unit,_base_ type.svg.

[6] Mary Branch, Thomas Coleman, and Yuying li. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM Journal on Scientific Computing, 21, 12 1999. doi: 10.1137/

S1064827595289108.

[7] David R. Brillinger. Some aspects of modern population mathematics. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 9(2):173-194, 1981. ISSN 03195724. URL http://www.jstor.org/stable/3314611.

[8] François Chollet et al. Keras. https://keras.io, 2015.

[9] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.

[10] Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation, 2020. URL https: //github.com/seancurrent527/ugrad_thesis.

[11] H. S. B. Filho, F. B. de Lima Neto, and W. Fusco. Migration and social networks: An explanatory multi-evolutionary agent-based model. In 2011 IEEE Symposium on Intelligent Agent (IA), pages 17, April 2011. doi: 10.1109/IA.2011.5953616.

[12] An image is worth one word: Personalizing text-to-image generation using textual inversion

[13] Zuguang Gu, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. circlize implements and enhances circular visualization in r. Bioinformatics, 30: 28112812, 2014.

[14] Saeed Heravi, Denise R. Osborn, and C.R. Birchenhall. Linear versus neural network forecasts for European industrial production series. International Journal of Forecasting, 20(3):435 446, 2004. ISSN 0169-2070. doi: https: //doi.org/10.1016/S0169-2070(03)00062-1. URL http://www.sciencedirect. com/science/article/pii/S0169207003000621.

[15] Tim Hill, Leorey Marquez, Marcus O'Connor, and William Remus. Artificial neural network models for forecasting and decision making. International Journal of Forecasting, 10(1):5 15, 1994. ISSN 0169-2070. doi: https://doi.org/10. 1016/0169-2070(94)90045-0. URL http://www.sciencedirect.com/science/ article/pii/0169207094900450.

[16] Kurt Hornik, Maxwell Stinchcombe, Halbert White, et al. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359366, 1989.

[17] Larry Lin, Kathleen M Carley, and Shih-Fen Cheng. An agent-based approach to human migration movement. In Proceedings of the 2016 Winter Simulation Conference, pages 35103520. IEEE Press, 2016.

[18] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115133, 1943.

[19] UN News. Migration and the climate crisis: the UN's search for solutions, 7 2019. URL https://news.un.org/en/story/2019/07/1043551.

[20] Department of Economic and Population Division Social Aairs. International migration flows to and from selected countries: 2015 revision. 2015. URL https://www.un.org/en/development/desa/population/ migration/data/empirical2/docs/migflows2015documentation. pdf.

3800. doi: https://doi.org/10.1016/S0304-3800(02)00064-9. URL http://www. sciencedirect.com/science/article/pii/S0304380002000649.

[22] Hye Jin Park, Woo Seong Jo, Sang Hoon Lee, and Beom Jun Kim. Generalized gravity model for human migration. arXiv preprint arXiv:1805.10422v2, 9 2018.

[23] E. G. Ravenstein. The laws of migration. Journal of the Statistical Society of London, 48(2):167235, 1885. URL https://www.jstor.org/stable/pdf/ 2979181.pdf.

[24] Caleb Robinson and Bistra Dilkina. A machine learning approach to modeling human migration. CoRR, abs/1711.05462, 2017. URL http://arxiv.org/abs/ 1711.05462.

[25] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261272, 2020. doi: https://doi.org/10.1038/s41592-019-0686-2.