JOURNAL OF DYNAMICS AND CONTROL

VOLUME 8 ISSUE 10

IMAGEGAFTER: AI-DRIVEN DATA AUGMENTATION FOR COMPUTER VISION

Madhuri Barhate, Rutuja Diwate, Hitesh Tolani, Diksha Teware, Ashwanti Gaikwad, Shravan Kalzunkar

Vishwakarma Institute of Technology, Pune, India

ISSN: 1672-6553

IMAGEGAFTER: AI-DRIVEN DATA AUGMENTATION FOR COMPUTER VISION

Madhuri Barhate, Rutuja Diwate, Hitesh Tolani, Diksha Teware, Ashwanti Gaikwad, Shravan Kalzunkar*
Vishwakarma Institute of Technology, Pune, India

***Corresponding Author: tewarediksha03@gmail.com**

ABSTRACT: ImageGafter introduces a Python-based software that leverages large language models (LLMs) to address the persistent issue of image scarcity in large-scale deep learning models. The software takes an input image and uses advanced algorithms to generate coherent and descriptive text representations of the visual content. By seamlessly integrating cutting-edge AI technologies, it bridges the gap between visual and textual data, tackling the challenge of limited annotated images in deep learning. The project goes beyond text generation by creating a diverse set of prompts using Generative AI. These prompts are designed to capture various features of the input image, and are then used by Generative AI to create new images, effectively expanding the dataset. This artificial dataset growth provides a richer source of information for training deep learning models. By combining LLM-driven image-to-text generation with prompt-based image synthesis, the project offers an innovative solution to image scarcity, paving the way for more robust and accurate deep learning models.

KEYWORDS: Generative AI, Graphical user interface, LLM, Prompt generation, Image generation.

1. Introduction

Super-intelligent artificial machines are part of this age. Computer vision has been married with natural language processing to do something new. Here is a piece of Python-based software, focusing on the problem of the scarcity of images in large deep learning models. Generative AI makes our software take an image input and produces the most advanced algorithm possible for the coherent description of the visual content.

By seamlessly integrating cutting-edge AI technologies, this software not only bridges the gap between visual and textual data but also addresses the critical issue of limited annotated images in the realm of deep learning. The project doesn't stop at text generation; it extends its capabilities to create a diverse set of prompts using Generative AI. These prompts are carefully crafted to encapsulate various aspects and features of the input image.

The intention is to expand the dataset artificially, providing a richer source of information for training large deep learning models. Through the innovative use of Generative AI in both text and prompt generation, our project aims to revolutionize the conventional approaches to handling image scarcity, paving the way for more robust and accurate deep learning models

In essence, ImageGafter forms a holistic solution to the challenge of image scarcity by not only describing images with text but also dynamically generating a set of prompts to augment the available dataset. By addressing this fundamental bottleneck in the training of deep learning models, our Python software endeavors to contribute to the advancement of AI applications that heavily rely on visual data for their learning and decision-making processes.

2. Existing Systems

Many GUI-based image-to-image generation systems exist in the market, but they often have the following limitations, which our project aims to address:

- Limited Control Over Individual Images: Users cannot customize individual images generated in a batch. For example, if 4 images are generated from a single source, the user cannot modify just image 2 using a specific prompt.[1]
- Restricted Number of Images: Most systems limit the number of images that can be generated from a single image, typically capping it at 4.[2]
- Non-Intuitive UI: Many systems have cluttered or unintuitive user interfaces, making them harder to navigate.[3]
- Difficulty in Reproducibility: Replicating results is challenging due to a lack of transparency in text prompts, making it harder to recreate outputs exactly.[4]

3. Proposed system and main functionalities

ImageGafter is an advance Image-to-Image tool which blends the power of Image Captioning, LLMs and Diffusion models. At the heart of ImageGafter's capabilities is the ability to similar images from a given image and provide a simple yet powerful GUI for better human to generative ai models. Features like Generative AI based image generation make augmenting image datasets a walk in the park.

Initially, a caption is generated for the given image which can then be tailored according to user's demands. After which number of image captions where n is the number of images to be generated which will be specified when the image is given as input. These prompts are then visible to the user, and the user may tailor the as per his need.

After n prompts have been generated, the user can again fine-tune them after which these prompts are then passed to diffusion model for image generations. The generated images and their respective image generation prompt are visible to user in the GUI.

These AI-generated images help diversify datasets by creating new variations of existing images, covering a wider range of features such as different angles, lighting, styles, or objects. This enhanced variety reduces overfitting by exposing the model to more scenarios, improving its ability to generalize and recognize patterns in real-world data. As a result, the model becomes more robust and accurate, especially when dealing with previously unseen inputs. By expanding the dataset with diverse, high-quality images, AI-generated images play a key role in improving overall model performance.

 ISSN: 1672-6553

In essence, ImageGafter eliminates the need for manually applying transformations like rotation, flipping, scaling, or color adjustments to diversify data. Instead, it automates the process of generating new images from a given input, streamlining image augmentation. This is particularly advantageous for large datasets with imbalanced image distribution, where manual augmentation would be a time-consuming and laborious task.

The steps for system are as follows:

1) The first step involves edit the config.json file for setting the gemini and stable diffusion api keys and other settings for these models such as number of samples.

2) Then click on the upload image button, enter the path of image or directory in case of multiple images in the alert box.

3) The process of Image generation will start and a loading screen will appear. After the image generation is completed the images along with their prompt will be displayed in the GUI screen.

4. System Design

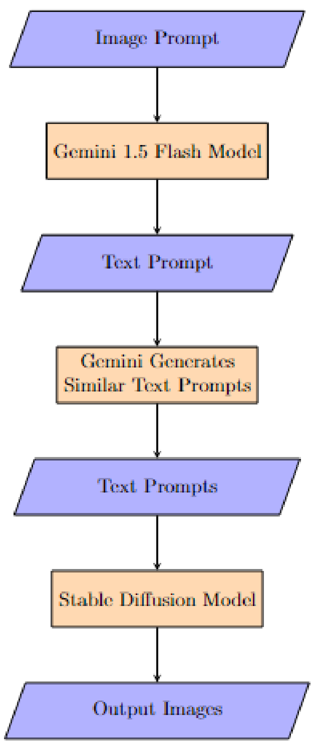


Fig - Process Flow

Fig. 1: System Architecture

ImageGafter allows to augment your image datasets seamlessly, Initially, input your gemini api key, adjust the prompt generation settings and input your stable diffusion api key in the config.json file. Subsequently, After your image has been given as input, prompt generation and image

ISSN: 1672-6553

generation takes place.

Input Image:

Enter the path of image which needs to be augmented. After selecting the number of samples to be generated. The image is automatically preprocessed by the software such as image dimensions being resized. You may also choose to enter the path of a directory instead of a single image.

Prompt Generation:

The software generates n number of image captions where n is the number of images to be generated which will be specified when the image is given as input. These prompts are then visible to the user, and the user may tailor the as per his need.

Image Generation:

After the user has customized the prompts as per his need, the prompts are then sent to a text to image model (Stable Diffusion Night Vision XL) for generating images.

The user may also customize the following parameters:

1. Negative Prompt: Entering a negative prompt will exclude the specified elements from the image generation process. Used for refining the images.
2. Guidance Scale: This value dictates how strictly should the generated image adhere the prompt text.
3. Scheduler: The field is used to control the noise in the image generated by the model.

Image Saving:

The user specifies an output directory to save the images. By default, images are stored in the 'outputs' folder of the current directory. Each image is named using an UUID (Universally Unique Identifier) to prevent overwriting of images stored in the same directory.

4. Screen layouts

A. Image Upload

Fig 2: Upload Image

ISSN: 1672-6553

B. Enter path of image file

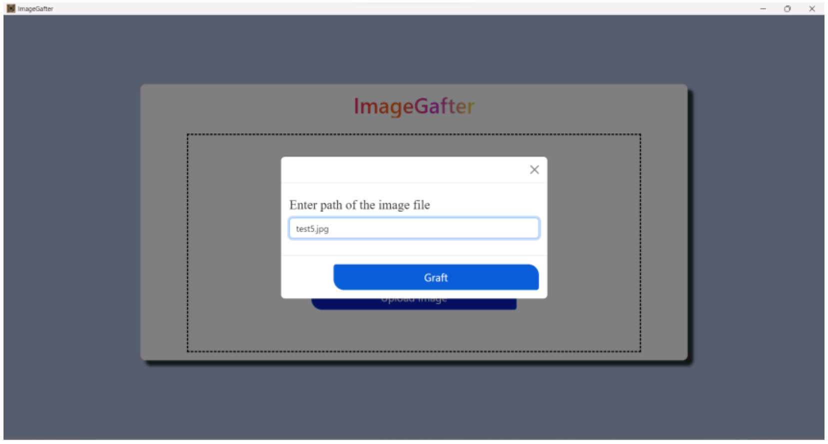


Fig 3: Image path

C. Generating similar images

Fig 4: Image generation

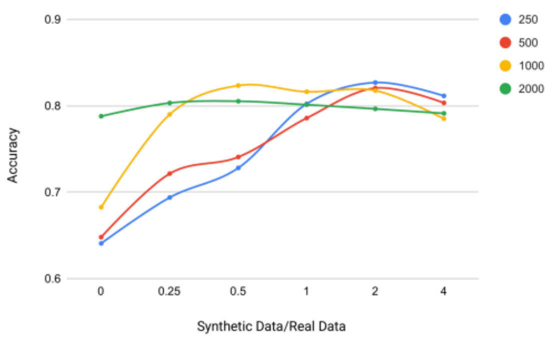
D. Generated Images from given image with text prompts

Fig 5: Results

ISSN: 1672-655

5. Comparative analysis

Below are the graphs and tables for a comparative analysis of models based on accuracy. The analysis includes models trained exclusively on real image data, a mix of synthetic (AI-generated) and real data, and models trained solely on synthetic data.



Mixing Ratio	Train Size			
	250	500	1000	2000
0	0.6408	0.6480	0.6826	0.7878
0.25	0.6938	0.7215	0.7810	0.8032
0.5	0.7281	0.7407	**0.8233**	**0.8052**
1	0.8022	0.7857	0.8162	0.8012
2	**0.8268**	**0.8204**	0.8175	0.7963
4	0.8113	0.8032	0.7849	0.7911

Fig 9: Analyzing mixing ratio vs. accuracy [5]

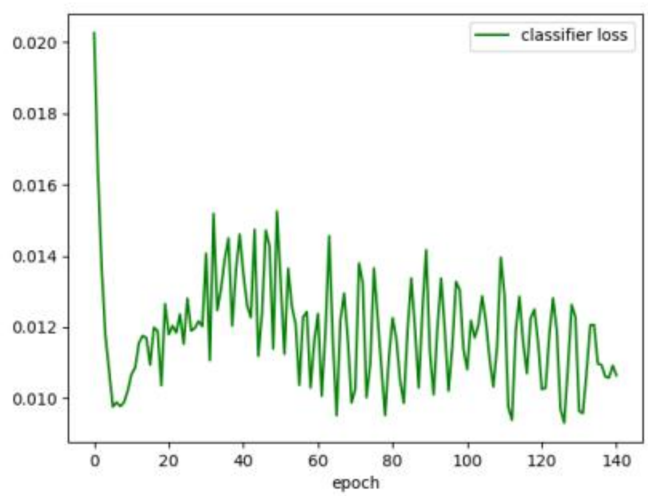


Fig 10: Classifier loss for 1000 real images [5]

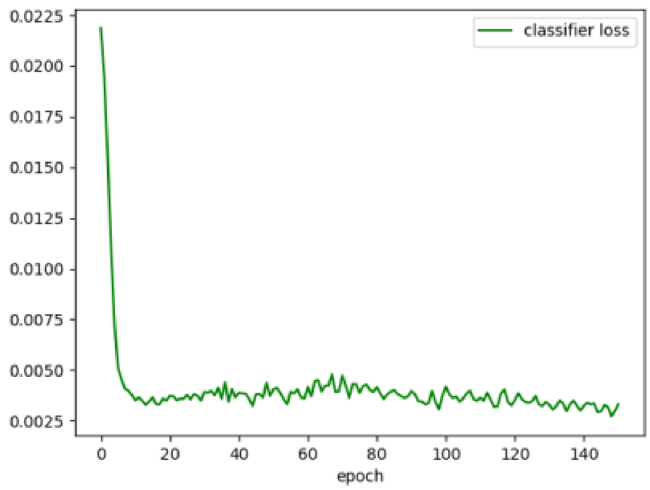


Fig 11: Classifier loss for 500 real and 500 synthetic images [5]

Dataset Size	Real Data Accuracy	Synthetic Data Accuracy
250	0.64076	0.70920
500	0.64796	0.76318
1000	0.68258	0.75662
2000	0.78781	0.79304

Table 1: Accuracies of classifiers trained solely on GAN-generated data [5]

6. Conclusion

ImageGafter is a sophisticated Image-to-Image tool that leverages the synergy of Image Captioning, Large Language Models (LLMs), and Diffusion models to facilitate efficient image dataset augmentation. By generating tailored captions for input images, users can specify the number of variations needed, which are then refined and passed to a diffusion model for image synthesis. This automated process enhances dataset diversity by producing variations in angles, lighting, styles, and objects, thereby reducing overfitting and improving model generalization. Ultimately, ImageGafter streamlines the image augmentation workflow, eliminating the need for manual transformations and significantly optimizing the handling of large, imbalanced datasets.

References

1. Aditya Ramesh, Mikhail Pavlov, Gabriel Goh. Scott gray, et al, “Zero-Shot Text-to-Image Generation”, arXiv, 2021.
2. Nanda Kumar, Manoj Kumar M, Hari Hara Sudhan Vand Santosh R, “Text to Image Generation using AI”, IJCRT, 2023.
3. Akanksha Singh, Ritika Shenoy, Sonam Anekar and Prof Sainath Patil,” Text to Image using Deep Learning”, IJERT, 2021.
4. Philip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, “Image-to-Image Translation

ISSN: 1672-655

5. with Conditional Adversarial Networks”, arXiv, 2018
6. David Liu and Nathan Hu, "GAN-Based Image Data Augmentation", Standford CSS29, 2020.
7. Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, et al, “Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding”, arXiv, 2022.
8. Jonathan Ho, Ajay Jain and Pieter Abbeel, “Denoising Diffusion Probabilistic Models”, NIPS, 2020.
9. Lvmin Zhang, Anyi Rao and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models”, IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
10. Omri Avrahami, Dani Lischinski, and Ohad Fried, “Blended Diffusion for Text-driven Editing of Natural Images”, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
11. Gafni, O., Polyak, A., Ashual, O., Sheynin, S., Parikh, D., Taigman, Y. (2022). "Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors”. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds) Computer Vision – ECCV 2022. ECCV 2022. Lecture Notes in Computer Science, vol 13675. Springer, Cham.