International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Language Enabled Image Originator

Akarsh J[1], Dr. Devaraj Verma C [2]

[1]*Research Scholar, Department of CSE, Jain (Deemed-to-be) University*
[2]*Professor, Department of CSE, FET, Jain (Deemed-to-be) University.*

---***---

Abstract - Traditionally, forensic artists would painstakingly sketch a suspect's face from a witness's \statement in order to create forensic photographs. There are restrictions on this procedure, though. First, it depends a great deal on the interpretation of the artist, which can bring falsehoods and prejudices. It can also take a lot of time, particularly if the drawing needs to be refined repeatedly.
Image generation is the process of creating new pictures that are comparable to the people in a certain dataset. Producing visually realistic pictures that fit the input's properties is the aim of image creation. information. Machine learning employs a number of pictures generating approaches, such as auto-regressive models, variational autoencoders (VAEs), generative adversarial networks (GANs), and stable diffusion models. These models are trained on an image dataset (for instance, 5,85B CLIP-filtered image-text pairings make up the large-scale research dataset LAION 5B) and are taught to produce new pictures that are comparable to the original data. An image generating model may be used to create a series of pictures for forensic sketching, with the witness's description serving as the basis for selecting the best image. This and the necessary face feature changes may be fed into the Image-to-Image Translation model. Until an adequate picture is produced, the picture-to-Image Translation model creates a fresh series of images with alterations.

Key Words: variational autoencoders (VAEs), Generative adversarial networks (GANs),

1.INTRODUCTION

Law enforcement frequently uses the conventional technique of forensic sketching to produce a visual depiction of a suspect's appearance based on the description given by a witness or victim. The witness's capacity to remember and precisely describe certain face traits, as well as the artist's interpretation of those features, place limitations on this method. The old approach of forensic sketching might produce erroneous drawings that can have major repercussions in criminal investigations since people prefer to recall faces holistically. Owing to these drawbacks, a more precise and effective technique for generating suspects' faces is required.

The goal of this project is to use generative AI to create a forensic picture generator in order to overcome these issues. Large datasets of face photographs can teach the AI model the general structure, proportion, and symmetry of the face, enabling it to create a collection of images based on a comprehensive description supplied by witnesses or victims. After that, the user may choose the best image from the collection, and the Image-to-Image Translation Model will utilize that image to create the next set of images. Through this iterative process, the produced pictures may be improved and refined, leading to a final image that closely resembles the witness's account.

In addition to lowering the time and expense involved in manual sketching, the suggested method also has the potential to improve final image accuracy and diversity, which will facilitate the identification of suspects from a variety of racial and gender backgrounds. In the end, generative AI technology has the ability to greatly increase the precision and effectiveness of forensic drawing, which would enhance the results of criminal investigations and bring victims' rights to justice.

2. METHODOLOGY

2.1 Objectives

1) Select a suitable generative Text-to-Image Translation model that can be trained on a sizable dataset of facial pictures to understand the symmetry, proportion, and structure of the face.
2) Combine the Text-to-Image Translation model with an Image-to-Image Translation model that produces a collection of pictures according to a comprehensive description given by victims or witnesses.
3) Provide a UI that lets the user choose the best-suited picture from the created set. The picture-to-Image Translation model uses this image to create the next set of images.
4) By contrasting the produced photographs with real suspects in criminal investigations, you may assess the forensic image generator's precision and effectiveness.
5) Compare the forensic picture generator's performance against more conventional forensic sketching techniques like hand drawing or facial composites.
6) Examine the forensic image generator's possible effects on criminal investigations, taking into account the time and money savings from eliminating the need for hand drawing as well as the possibility of improved picture correctness and diversity.

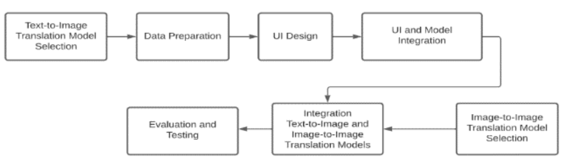

Fig –2.1: Design Overview of Language Enabled Image Originator.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

2.2 Execution

Implementing a project on forensic sketch using generative AI involves several key steps. The process of implementation is as follows:

1) **Data Collection:** Gather a dataset consisting of facial images and corresponding textual descriptions or sketches. This dataset should include both real facial images and corresponding forensic sketches or textual descriptions of those images.
2) **Data Preprocessing:** Preprocess the facial images by resizing them to a consistent size and normalizing the pixel values. Preprocess the textual descriptions by tokenizing, removing stopwords, and converting words to numerical representations (e.g., word embeddings or one-hot encodings).
3) **Model Architecture Selection:** Choose an appropriate generative AI model architecture, such as a conditional generative adversarial network (cGAN) or a variational autoencoder (VAE). The model should take the textual description or forensic sketch as input and generate a corresponding facial image.
4) **Training:** Train the generative AI model on the preprocessed dataset. During training, optimize the model's parameters using techniques like backpropagation and gradient descent. The model should learn to generate realistic facial images based on the textual descriptions or forensic sketches.
5) **Evaluation:** Evaluate the performance of the generative AI model by generating facial images from textual descriptions or forensic sketches and assessing their quality. Use metrics like perceptual similarity, structural similarity index (SSIM), or human evaluation to measure the similarity between the generated images and real facial images.
6) **Iteration and Fine-tuning:** Based on the evaluation results, finetune the generative AI model by adjusting hyperparameters, modifying the architecture, or increasing the training duration. Iterate this process until satisfactory results are achieved.
7) **Deployment:** Once the generative AI model performs well, deploy it as a usable tool for generating forensic sketches from textual descriptions or vice versa. Create a user-friendly interface where users, such as forensic artists or investigators, can input text or sketches and receive corresponding facial images as output
8) **Ethical Considerations**: Take into account ethical considerations throughout the implementation process, such as privacy, data protection, and potential biases in the generated images. Ensure that the project adheres to responsible AI practices and respects legal and ethical guidelines.

3. SYSTEM ARCHITECTURE:

3.1 Generative AI:

1) A subset of artificial intelligence approaches centered on employing machine learning algorithms to create or generate new content—such as text, photos, music, and even videos—is referred to as generative AI, or generative modeling.
2) In contrast to conventional AI models, which are tailored for certain tasks, generative models leverage patterns found in existing data to produce novel and creative content.
3) The technique of creating new pictures that resemble those in a given dataset is known as image generation. Machine learning employs a number of methods for generating images, such as auto-regressive models, variational autoencoders (VAEs), generative adversarial networks (GANs), and stable diffusion models. These models learn to produce new pictures that are similar to the input data by being trained on a collection of images (for instance, LAION 5B, a large-scale dataset for research purposes consisting of 5,85B CLIP-filtered image-text pairings).
4) This, together with the face traits that require modification, may be fed into the Image-to-Image Translation model. Until an adequate picture is produced, the picture-to-Image Translation model creates a fresh series of images with alterations.

3.2 Stable Diffusion:

Image processing techniques such as "stable diffusion" are used to minimize noise and smooth out the image while maintaining key characteristics. Stable diffusion can be applied to forensic sketching in order to improve and enhance the image's realism. During this procedure, a filter is applied to the image that selectively smooths it out while keeping key elements like skin texture, facial features, and other distinctive qualities. As a result, it is simpler to identify the culprit using a crisper, more solid picture. The following figure shows how a stable diffusion model works step-by-step.

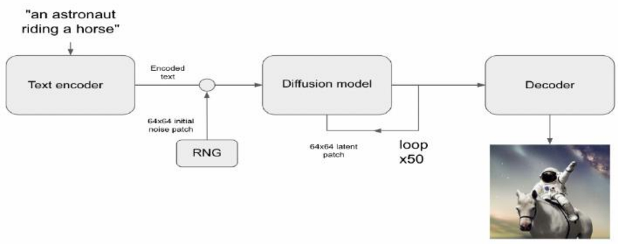

Fig -3.1: Flow Chart of Stable Diffusion Model.

4. RESULTS OBTAINED

4.1 Snapshots of Execution

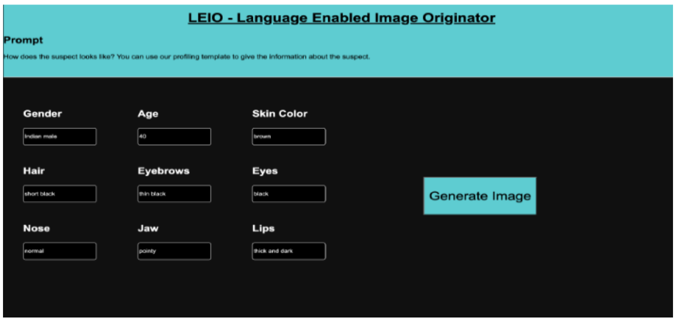

Fig -4.1: Interface for Input.

Figure - 4.2: Execution Status in Terminal.

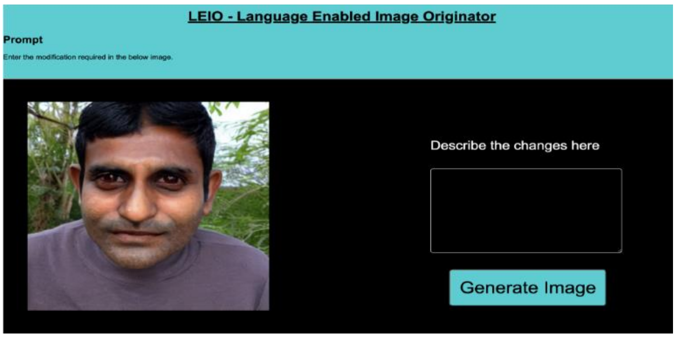

Figure - 4.3: Generated Image.

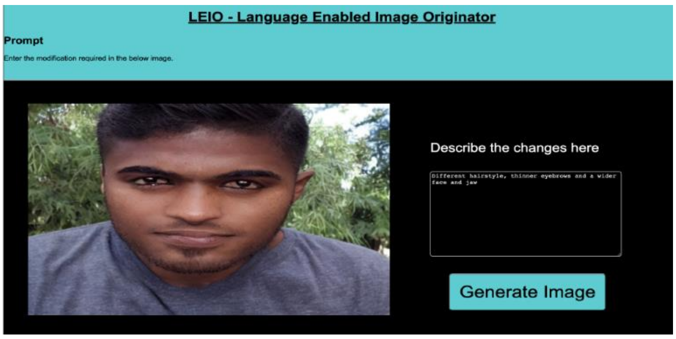

Figure - 4.4: Image Generated After Modification.

Here, we investigated how well a stable diffusion model could produce lifelike portraits of individuals using the user's spoken remarks. While certain oddities were seen in the produced pictures, the stable diffusion model showed outstanding results. We discovered that the specificity of the given description closely correlated with the produced photos' accuracy. The produced photos closely matched the descriptions supplied by users when they were specific and thorough. But vague or less precise descriptions frequently produced pictures that had little errors or discrepancies.

We used an image-to-image stable diffusion model to handle any unwanted alterations or abnormalities in the resulting pictures. The required adjustments were effectively performed to the produced photos by this model. But we also saw instances when the image-to-image model unintentionally made modifications that weren't consistent with what the user had described. We included a negative prompt mechanism in order to lessen these unwanted changes. We were able to steer the image-to-image model away from producing certain undesirable alterations by giving it negative hints. This method worked well to stop unwanted changes to the pictures.

Furthermore, we incorporated an iterative modification procedure that enabled users to indicate the necessary adjustments to be made to the produced photos until the target image was achieved. This iterative procedure turned out to be more productive and efficient than the conventional manual forensic drawing techniques All things considered, our small research showed how stable diffusion models can produce convincing pictures of suspects based on spoken descriptions. The precision of the description, the use of an image-to-image model, negative prompts, and iterative improvements helped us to overcome these difficulties and provide good results, even if certain oddities and unwanted alterations were noticed.

5. CONCLUSION

The stable diffusion model showed promise in producing pictures that closely matched the given descriptions. We did, however, come across a few oddities and unwanted alterations in the produced photographs. However, we were able to address these problems and get acceptable outcomes by using the image-to-image stable diffusion model, negative prompts, specificity in descriptions, and an iterative revision process.

Our method's combination of the text-to-image and image-to-image models turned out to be quite important. The image-to-image model effectively performed the required adjustments to improve the authenticity of the generated pictures, while the text-to-image model faithfully recorded the spoken descriptions and created the first images. We were able to overcome obstacles and raise the produced photos' accuracy thanks to this combo.

The continuous improvement in model architectures and training techniques has resulted in increasingly realistic and detailed images. However, challenges remain, particularly in ensuring the accuracy, diversity, and ethical use of generated images. Addressing these challenges involves not only technical innovation but also careful consideration of the social and ethical implications. As we look to the future, the integration of multimodal AI systems that seamlessly combine text, image, and even audio generation holds exciting promise. By harnessing the synergy between different types of data, we can unlock new levels of creativity and utility, paving the way for AI to become an even more integral part of our everyday lives.

REFERENCES :

1) S Nikkath Bushra and K Uma Maheswari. Crime investigation using dcgan by forensic sketch-to-face transformation (stf)-a review. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pages 1343–1348. IEEE, 2021.
2) Deng-Ping Fan, Ziling Huang, Peng Zheng, Hong Liu, Xuebin Qin, and Luc Van Gool. Facial-sketch synthesis: a new challenge. Machine Intelligence Research, 19(4):257–287, 2022.
3) Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bj¨orn Ommer. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10684– 10695, 2022.
4) Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, and In So Kweon. Text-toimage diffusion model in generative ai: A survey. ArXiv preprint arXiv:2303.07909,2023.
5) Y. Kataoka, T. Matsubara and K. Uehara, "Image generation using generative adversarial networks and attention mechanism," 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 2016, pp. 1-6, doi: 10.1109/ICIS.2016.7550880.
6) Ramzan, Sadia, Muhammad Munwar Iqbal, and Tehmina Kalsum. 2022. "Text-to-Image Generation Using Deep Learning" *Engineering Proceedings* 20, no. 1: 16. https://doi.org/10.3390/engproc2022020016
7) S. Göring, R. R. Ramachandra Rao, S. Fremerey and A. Raake, "AVrate Voyager: an open source online testing platform," 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, 2021, pp. 1-6.doi: 10.1109/MMSP53017.2021.9733561
8) H. Dong, S. Yu, C. Wu, and Y. Guo. Semantic image synthesis via adversarial learning. In Proceedings of the IEEE International Conference on Computer Vision, pages 5706–5714, 2017.
9) I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, pages 2672–2680, 2014.