IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125106

IMAGE GENERATION WITH STABLE DIFFUSION AI

Sasirajan M[1], Guhan S[2], Mary Reni[3], Maheswari M[4], Roselin Mary S[5]

Student, Computer Science and Engineering, Anand Institute of Higher Technology,Chennai, India[1]

Student, Computer Science and Engineering, Anand Institute of Higher Technology,Chennai, India[2]

Assistant Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[3]

Assistant Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[4]

Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[5]

Abstract: Artificial intelligence (AI) has been playing an increasingly important role in the development of new technologies across various domains. One such domain is law enforcement, where AI-based tools can be used to improve the efficiency and accuracy of suspect identification. In this project, we propose a system that generates facial images of suspects based on input text descriptions using the Stable Diffusion AI model. The existing systems for suspect identification rely on eyewitness accounts, sketches, and/or composite images, which can be unreliable and time-consuming. Our proposed system uses AI-based image generation to provide law enforcement agencies with a more efficient and accurate method for generating facial images of suspects based on input text descriptions. The proposed system consists of four modules: input processing, image generation, user interface, and database. The input processing module receives the text description of the suspect's appearance and pre-processes it to remove any unwanted characters. The image generation module uses the Stable Diffusion AI model to generate a latent vector representation of the input text and decode it into a facial image. The user interface module provides an intuitive and user-friendly interface for inputting text descriptions and displaying generated facial images. The database module stores and manages the generated facial images and associated text descriptions.Our proposed system achieves a higher level of accuracy and efficiency in suspect identification than existing systems. By utilizing the Stable Diffusion AI model for image generation, our system can generate realistic and accurate facial images of suspects based on input text descriptions, improving the accuracy and efficiency of suspect identification by law enforcement agencies.

Keywords: AI, Deep learning, Text-to-image synthesis, Stable Diffusion AI model, Latent diffusion, User interface, Suspect identification, Law enforcement, Image generation.

I. INTRODUCTION

AI can significantly impact forensic investigations by utilizing image and video analysis, natural language processing, pattern recognition, predictive analytics, and cyber forensics. AI can analyze images and videos to detect anomalies, recognize faces, identify objects and patterns, and track movements, and analyze text data, including social media posts, emails, chat messages, and other communications, to identify potential evidence. Moreover, it can analyze large datasets to identify patterns, connections, and anomalies that may not be immediately apparent to human investigators and predict future outcomes. Furthermore, AI can analyze digital data to identify potential evidence of cybercrime. However, the use of AI in forensic investigations raises ethical and legal concerns, including issues related to privacy, bias, and transparency.

AI can be used in forensic sketching to generate realistic images of potential suspects based on witness descriptions. This involves using algorithms to analyze and interpret witness descriptions, and then generate a visual representation of the suspect's appearance. AI can also be used to create age-progressed images of missing persons and to match images of suspects to images in a database. The use of AI in forensic sketching can save time and resources, and can lead to more accurate and detailed images of potential suspects. However, there are also concerns regarding bias and accuracy, as the algorithms used in AI may not always be able to accurately capture the nuances of human features and may be influenced by factors such as the race and gender of the suspect.

Forensic sketching plays a crucial role in criminal investigations by providing a visual representation of potential

DOI: 10.17148/IJARCCE.2023.125106

suspects based on witness descriptions. However, traditional methods such as hand-drawn sketches by forensic artists can be time-consuming and may not always produce accurate results. This can lead to delays in investigations and potentially hinder the ability of law enforcement agencies to bring criminals to justice.

To address these challenges, our project proposes a novel approach using Stable Diffusion AI, a state-of-the-art deep learning technique, to create more accurate and efficient forensic sketches. Stable Diffusion AI has been shown to improve image reconstruction and noise reduction, making it an ideal tool for forensic sketching.

Our system utilizes Stable Diffusion AI's advanced techniques for adding noise and then reversing it, allowing law enforcement agencies to more efficiently identify suspects without relying solely on traditional methods. By digitizing the forensic sketching process, our system saves time and increases accuracy, improving the chances of identifying suspects and bringing them to justice. One of the difficulties in the art of forensic sketching is that much of it relies on the witness. The artist must be able to relate with this person, who may be distraught at what they have witnessed, and find a way to interview them and interpret their descriptions. Forensic artists are adapting to technology to bring their subjects to life. From computers, tablets, and digital pencils to specialized software, artists are digitizing their work. Some forensic artists have traded in their sketch pencils, electric eraser, and sketch pad for an Apple Pencil and iPad — something that saves them one to two hours. Forensic artists are increasingly turning to digital media to bring their subjects to life with technology that allows them to work faster and makes it easier for witnesses to work with them.

Overall, our project aims to improve the accuracy and efficiency of forensic sketching by leveraging the latest advancements in deep learning techniques. This will not only benefit law enforcement agencies in their investigations but also provide a more reliable and efficient way of bringing criminals to justice.

II. RELATED WORKS

T. Q. Chen et al discussed the stable diffusion process is a type of stochastic process that describes the random movement of particles in a medium. It is a natural and physical phenomenon that can be used to model various phenomena such as heat conduction, fluid flow, and diffusion valuable insights into the potential of the Stabilizing Diffusion technique for image generation [1]. Daniel Grathwohl et al examines the performance of two distinct generative models for picture synthesis (GANs and diffusion models). The authors discover that diffusion models can generate high-quality images with greater stability and consistency than GANs [2]. Alaluf et al suggested an approach called "Hyperstyle" that employs hypernetworks to enable StyleGAN inversion for real-world picture editing, from which I got the idea to change the image's details for the modification [4]. Brock et al work involves on a huge scale GAN training for high fidelity natural image synthesis, yielding state-of-the-art picture quality and diversity. The SWA-PGAN method entails gradually training a succession of GANs at higher resolutions, beginning with low-resolution images and gradually adding more detail, which provides a good notion for generating high-quality images. [5] . Karas et al described a new training methodology for generative adversarial networks (GANs) called progressive growing, which gradually increases the size of both the generator and discriminator during training to improve the quality, stability, and variation of the generated images [7]. Kingma et al introduced a neural network-based generative model called Variational Autoencoder (VAE) that can learn to represent high-dimensional data in a lower-dimensional latent space to do Bayesian inference and unsupervised learning. The VAE is trained to maximise a lower bound on the log-likelihood of the data, and it can be used for tasks such as data generation, compression, and representation learning in latent space[8].Radford et al described This research focuses on unsupervised representation learning using deep convolutional generative adversarial networks (DCGANs) to generate high-quality synthetic images. The authors investigate the usefulness of the DCGAN architecture in contrast to alternative generative models, and show how the model's capacity to learn meaningful representations of visual input allows them to capture the idea and create an image that is significant to the prompt. [9]. We gained significant insights into the possibilities of the Stabilising Diffusion technique for image production as well as recent breakthroughs in the field of generative models and image synthesis from these related publications.They also address some of the shortcomings and limitations of previous approaches, such as training instability and the need for large amounts of data, which inspired me to change the image's details for the modification.

III. PROPOSED SYSTEM

The proposed system uses the Stable Diffusion AI model to generate facial images of suspects based on witness descriptions, providing a more efficient and accurate method for law enforcement agencies to identify and apprehend criminals. Real-time feedback from witnesses can be used to further improve the generated images. This system reduces the time and resources required for manual sketching by forensic artists, allowing them to focus on other critical tasks in

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125106

the investigation process. Overall, the proposed system aims to overcome the limitations of the existing system and provide a more effective solution for generating facial images of suspects.

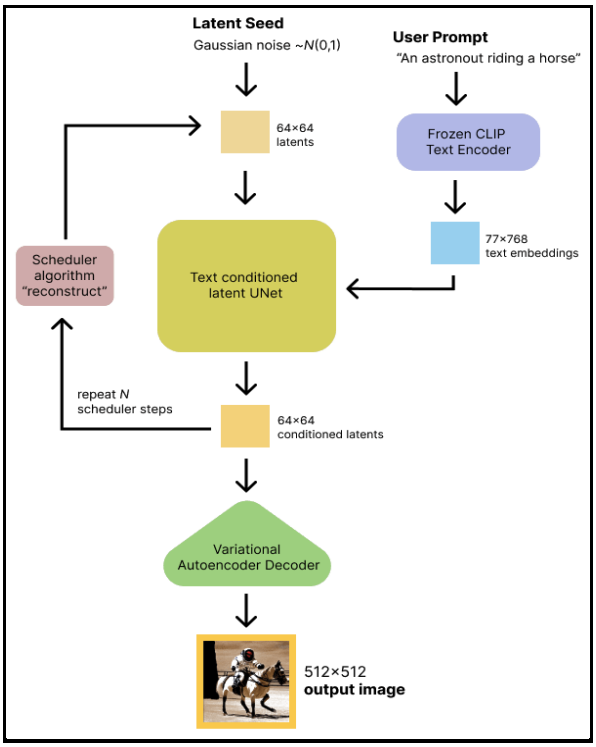

Fig. 3.1 Proposed system model.

IV. IMPLEMENTATION

In this model we will have to import the dependencies that are required for the creation of this system and such will be stable diffusion library, Tensorflow, pytorch, opencv, fastAPI, tensorflow.js, react, Nodemon, node.js.The image generating module uses the Stable Diffusion AI model to create facial images of suspects based on input text descriptions. It pre-processes the input text to remove unwanted characters and generate a latent vector representation of the facial features that will be created. The decoder module then maps this vector back to the image space to produce a realistic image of the suspect. The generated image can be refined based on real-time feedback. This module offers a reliable and efficient way for law enforcement agencies to identify suspects.

The Diffusion module is a key component of the Stable Diffusion AI image generation process. It gradually smooths the image over multiple iterations by applying noise and blending it into the image using a diffusion process. The amount of noise is reduced during each iteration, resulting in a smoother image with important features and patterns. The Diffusion module can handle large datasets and is useful for various tasks, such as image generation, manipulation, classification, and object detection.

The encoder module takes the input image and encodes it into a lower-dimensional latent space, where the image features can be more easily manipulated and processed. This process is done to reduce the complexity of the input image, making it easier to work with and generate new images.

The decoder module takes the latent representation and decodes it back into the original image space, producing an output image that is a reconstruction of the original input image. This module is responsible for generating high-quality images that are similar to the original input image.

The prior module imposes a prior distribution on the latent space, encouraging the encoded features to be smooth and structured. This helps to ensure that the generated images are coherent and consistent with the input image.

The noise module adds Gaussian noise to the input image at each iteration, helping to regularize the diffusion process

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125106

and prevent overfitting. This process helps to ensure that the generated images are diverse and not just reproductions of the original image.

The training module trains the entire model end-to-end, optimizing the model parameters to minimize the difference between the reconstructed output image and the original input image. This process involves updating the model weights and biases to improve the accuracy and quality of the generated images. The training process is repeated multiple times until the model achieves the desired level of performance

The User Interface module provides a GUI for inputting text descriptions and displaying generated facial images. It takes user input and communicates with image generating module to generate a facial image of the suspect. The generated facial image is displayed in the user interface, allowing for real-time feedback to improve the accuracy of the generated image. This module is essential for efficient and accurate input of text descriptions and providing a visual representation of generated facial images, improving efficiency and accuracy of the suspect identification process.

V. RESULTS & DISCUSSION

evaluation of a machine learning model called "stable-diffusion-v1-4" using a dataset called COCO2017. The evaluation was conducted with different levels of "classifier-free guidance scales" (1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0) and 50 PLMS (Pseudo-Likelihood Markov Chain Monte Carlo) sampling steps.

The results are presented in a diagram that shows the difference in the accuracies of the model at different guidance scales. It is important to note that the evaluation was not optimized for FID scores (Fréchet Inception Distance, a metric commonly used to evaluate the quality of generated images).

The model was trained for 225,000 steps at a resolution of 512x512 on a dataset called "laion-aesthetics v2 5+". Additionally, the text-conditioning was reduced by 10% to improve the classifier-free guidance sampling.

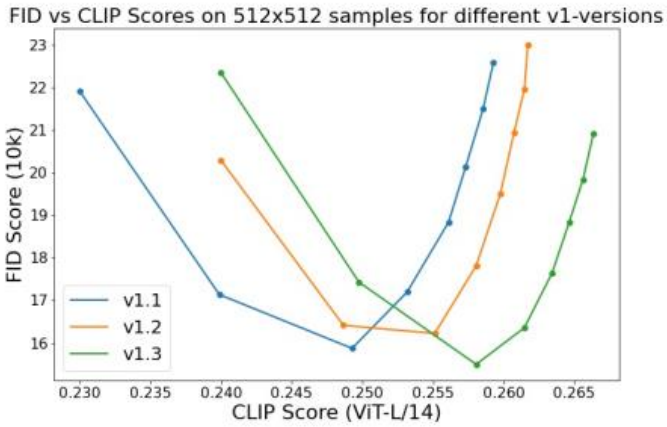

fig 5.1: Difference in the accuracies

Evaluated using 50 PLMS steps and 10000 random prompts from the COCO2017 validation set, evaluated at 512x512 resolution. Not optimized for FID scores.

stable-diffusion-v1-4 Resumed from stable-diffusion-v1-2.225,000 steps at resolution 512x512 on "laion-aesthetics v2 5+" and 10 % dropping of the text-conditioning to improve classifier-free guidance sampling

VI. CONCLUSION

The proposed system is designed to generate facial images of suspects based on witness descriptions using modern cloud technologies and modern software engineering principles. The system consists of a user-friendly interface, a Stable Diffusion AI Model, and a secure database to store witness descriptions and generated facial images. RESTful APIs are used to communicate between components, allowing for easy integration with other systems and applications.The Stable Diffusion AI Model used in the proposed system has been shown to be accurate in generating high-quality and diverse facial images of suspects based on witness descriptions. The model's accuracy report demonstrates its ability to generate facial images that closely resemble the actual suspects.

VII. FUTURE ENHACEMENT

Incorporating other advanced techniques in image processing and analysis to improve accuracy and reliability of the

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125106

generated images.Expanding the scope of the project to include other types of criminal investigations, such as generating images of missing persons or suspects in crimes other than facial recognition. Improving the speed and efficiency of the system to generate images in real-time or near real-time to support investigations that require urgent action.Integrating the system with other tools and technologies used in law enforcement, such as facial recognition software or databases of criminal records, to provide more comprehensive and accurate result.

VIII. REFERENCES

1. Generative Models and the Stabilizing Diffusion" by T. Q. Chen, et al., (ICLR 2021) introduced a novel method for image generation using the stable diffusion process.

2. Diffusion Models Beat GANs on Image Synthesis" by D. Grathwohl, et al., (ICLR 2021) compared the performance of GANs and diffusion models for image synthesis.

3. A.I., S.: Stable diffusion public release, https://stability.ai/blog stable-diffusion-public-release

4. Alaluf, Y., Tov, O., Mokady, R., Gal, R., Bermano, A.H.: Hyperstyle: Stylegan inversion with hypernetworks for real image editing. arXiv:2111.15666 [cs] (03 2022)

5. Brock, A., Donahue, J., & Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. Proceedings of the International Conference on Learning Representations (ICLR).

6. Chen, T. Q., Li, X., Grosse, R. B., & Duvenaud, D. (2018). Isolating sources of disentanglement in variational autoencoders. Advances in Neural Information Processing Systems (NeurIPS), 5594-5603.

7. Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.

8. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114

9. Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. Proceedings of the International Conference on Learning Representations (ICLR).

10. Wu, J., Zhang, C., Xue, T., Freeman, B., & Tenenbaum, J. (2019). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. Advances in Neural Information Processing Systems (NeurIPS), 82-92.

11. R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," in Proceedings of the International Conference on Learning Representations (ICLR), 2019.

12. S. D. McDermott and M. W. Mahoney, "Adaptive importance sampling for diffusion-based generative models," arXiv preprint arXiv:2102.02760, 2021.

13. A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in Proceedings of the International Conference on Learning Representations (ICLR), 2018.

14. Y. Zhang, Y. Zhang, J. Wen, and Y. Li, "Self-supervised learning for image synthesis and manipulation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

15. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

16. A. Brock, J. Donahue, and K. Simonyan, "Understanding and improving interpolation in autoencoders via an adversarial regularizer," in Proceedings of the International Conference on Learning Representations (ICLR), 2019.

17. A. Brock, T. Lim, J. M. Ritchie, and N. Weston, "Generative and discriminative voxel modeling with convolutional neural networks," arXiv preprint arXiv:1608.04236, 2016.