Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Generative Adversarial Networks for Forensic Image Synthesis and Identification

[1]*Vinaya Kulkarni,* [2]*Shital Karande,* [3]*Jagruti Patil,* [4]*Anushruti Adhikari,* [5]*Diti Jariwala,* [6]Arya Nigade
[1]*Department of Computer Engineering Bharati Vidyapeeth's College of Engineering for Women Pune, India*
vinaya.kulkarni@bharatividyapeeth.edu
[2]*Department of Computer Engineering Bharati Vidyapeeth's College of Engineering for Women Pune, India*
shital.jadhav@bharatividyapeeth.edu
[3]*Department of Computer Engineering Bharati Vidyapeeth's College of Engineering for Women Pune, India* *jagrutipatil9723@gmail.com*
[4]*Department of Computer Engineering Bharati Vidyapeeth's College of Engineering for Women Pune, India*
anushruti.adhikari24@gmail.com
[5]*Department of Computer Engineering Bharati Vidyapeeth's College of Engineering for Women Pune, India* *diti.jari@gmail.com*
[6]*Department of Computer Engineering Bharati Vidyapeeth's College of Engineering for Women* Pune, India arya.nigade2003@gmail.com

ARTICLE INFO

Received: 29 Dec 2024

Revised: 12 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

In forensic investigations and identity verification, manual facial sketching remains a time-consuming and subjective process. This paper proposes a two-phase automated system that integrates generative and deep learning techniques to overcome the limitations of traditional sketch-based recognition. In the first phase, facial sketches are synthesized from detailed textual descriptions using Stable Diffusion enhanced with ControlNet, effectively translating semantic features into visual representations. In the second phase, a deep metric learning approach using FaceNet is employed to extract embeddings from both the generated sketches and the CelebA dataset. Cosine similarity is then used to retrieve the top-matching faces from the precomputed database. The system demonstrates promising results in accurately identifying similar facial images based on sketch inputs, offering potential applications in criminal investigations, surveillance, and identity verification. Experimental results validate the effectiveness and scalability of the proposed approach.

Keywords: Sketch Generation, CelebA Dataset, Face Recognition, ControlNet, Stable Diffusion, Cosine Similarity

I. INTRODUCTION

Forensic facial sketching remains a crucial method in criminal investigations, particularly when photographic evidence of the suspect is unavailable. Traditionally, law enforcement agencies rely on trained forensic artists who manually translate eyewitness descriptions into hand-drawn facial sketches. However, this approach is time-intensive and highly subjective, relying heavily on both the memory of the witness and the artistic skill of the sketcher. As noted by Jha et al. [4], inaccuracies and delays caused by human error can significantly reduce the effectiveness of suspect identification, especially when time-sensitive action is needed.

Recent advances in deep learning and artificial intelligence have opened new possibilities for automating the sketch generation process. Generative Adversarial Networks (GANs), and more recently, Diffusion models, have shown promising results in text-to-image synthesis. GAN-based frameworks such as StackGAN and StackGAN++ employ multi-stage refinement to generate high-resolution images from textual descriptions. These models reduce the dependency on human sketch artists and enhance consistency in visual representation. Further improvements have been made using conditioning mechanisms like ControlNet, which guides the generative model based on additional structural or semantic inputs. Kumar et al. demonstrated that conditioning GANs on detailed textual prompts using architectures like StyleGAN2 can lead to highly accurate image outputs suitable for forensic reconstruction.

However, generating a sketch is only half the challenge in forensic applications. Once a facial image is synthesized, it must be matched against a large gallery of mugshots or face datasets to identify the suspect. Here, face recognition systems based on deep metric learning come into play. FaceNet, introduced by Schroff et al., remains a widely adopted framework that maps facial images into embedding vectors and compares them using cosine similarity. This approach allows for effective retrieval of similar faces even when there is a significant modality gap between sketches and photographs. The CelebA dataset, with over 200,000 facial images annotated with various attributes, serves as a reliable benchmark for training and evaluating such models.

This research builds upon earlier work presented in [23], where a conceptual approach for forensic sketch generation using GANs was explored, supported by a literature-driven analysis and performance benchmarking. The current study moves beyond theory to implement and evaluate a fully functional two-phase system. The first phase employs Stable Diffusion with ControlNet for text-to-sketch generation, leveraging natural language descriptions of suspects. The second phase embeds the generated sketch using FaceNet, followed by similarity-based retrieval against a precomputed CelebA embedding database. Through this automated pipeline, the research demonstrates a scalable and reliable solution for sketch-based suspect identification, offering improvements in efficiency, accuracy, and objectivity over traditional methods.

While several studies have advanced either sketch generation [13], [14] or sketch-to-photo matching [7], few have integrated both into a single automated system for forensic use.

II. LITERATURE REVIEW

In order to enhance suspect identification, Jalan et al. [1] developed a system that leverages Generative Adversarial Networks (GANs), specifically StyleGAN, to improve suspect identification. Their approach generates facial images based on input descriptions and further enhances them using Tunable Latent GAN (TL-GAN), which adjusts the input in the latent space. By reducing user input and accelerating the process, their approach greatly enhances identification accuracy, rendering it a useful tool for real-world crime investigations. In another study, Patil and Shubhangi [2] addressed face recognition by forensic sketch through a geometrical face model. They employed an AdaBoost approach for face detection and added a geometrical model to extract discriminative facial features. An additional artificial neural network (ANN) classifier enhances precision. Although impressive results are delivered, the work points out how improved feature extraction techniques are desired to deliver still better results for forensic purposes.

Dixit and Raj [3] went the more user-centric route by creating a system with a drag-and-drop interface through which users can simply draw facial sketches. These are then compared to a criminal database with the help of the DeepFace facial recognition library. The research highlights the need to improve this process further to increase the accuracy of recognition. Jha et al. [4] combined CNNs and GANs as a hybrid model to recognize face sketches. The CNNs assist in feature extraction from forensic sketches, and the GANs produce realistic images to identify them. Training the system on a big database of sketches and real photographs significantly enhances its robustness and accuracy of sketch-to-photograph matching [4]. Wu et al. [5] proposed a new sketch-to-sketch transformation model based on conditional GANs (cGANs). The method iteratively improves early forensic sketches from ambiguous or partial descriptions to assist forensic artists in creating more precise and detailed representations, leading to greater chances of successful identification.

Building on facial sketch recognition, Reed et al. [6] investigated how age progression can influence sketch identification. Using a recurrent neural network (RNN), their system keeps up with changing facial features with age, thus it can recognize aged sketches and real-time photographs. This methodology manages to cope effectively with the vicissitudes of natural aging and is, therefore, a useful resource in the long- term prosecution of crime [6]. AttnGAN, introduced by Xu et al. [7], employs an attention mechanism to produce very detailed images from textual descriptions by highlighting the most pertinent words for each sub-region of an image. It also employs a deep multimodal similarity model to enhance the precision of matching an image to textual descriptions. Showing a breakthrough improvement on the CUB and COCO datasets, this model represents a breakthrough in text-to-image synthesis. Ouyang et al. [8] addressed the challenging problem of cross-modal face matching, especially in the case of forensic sketches and caricature sketches that tend to involve abstraction and exaggeration. To this, they introduced a mid-level facial attribute model that learns semantic properties separately for each modality and is thus robust to distortion and misalignment. They also merged these qualities with novel characteristics through Canonical Correlation Analysis (CCA) to result in robust outcomes. They also provided a new dataset with 59,000 attribute labels to enable future studies.

Thakare et al. [9] suggested a CNN-based method to identify forensic face sketches using both global and local features with a fusion model. This approach improves accuracy in recognition and provides a stable solution to the difficulties of law enforcement agencies while identifying suspects. Galea and Farrugia [10] tackled the challenges of forensic sketch- mugshot gallery matching using a deep learning method optimized for small data. By transferring a pre-trained face recognition model through transfer learning and employing a 3D morphable model to generate synthetic images and sketches, they enhanced accuracy and minimized error rates. This novel framework has potential for enhancing criminal identification via deep learning methods. Zhao et al. [11] addressed the problem of generating photorealistic facial images from sketches, thus confronting this challenge as face hallucination. This GAN-based model uses both the outline and attribute vectors to better contribute to an image that attains realistic completeness with respect to outlines and attributes. The generator part of the network includes a feature extraction and downsampling-upsampling layer connected through skip connections to maintain performance with fewer layers while ensuring that the generated face matches the desired attributes through the discriminator.

Kumar et al. [12] explored generating face images from text descriptions using a two-stage StackGAN architecture for sketch refinement. Unlike prior work focused on simpler subjects like birds or flowers, their model interprets detailed facial attributes such as hair color and expressions to produce realistic images. Tested on the CelebA dataset [21], their approach achieved an inception score of 4.04, demonstrating its effectiveness in translating complex textual descriptions into high-resolution, realistic face images. This method has potential applications across art, entertainment, and education.

Zhang et al. [13] presented the problem of generating high- quality images with GANs. The authors introduce StackGAN, a two-stage approach in which Stage I generates a low- resolution image using a text description, while Stage II refines that image into a high-resolution photorealistic output. Additionally, StackGAN-v2 introduces a multi-stage architecture with multiple generators and discriminators, improving stability and image quality. Some proposed models have reportedly outperformed existing methods in realistic image synthesis and are considered significant advancements in the realm of text-to-image synthesis.

Ayanthi [14] examined the task of generating photorealistic facial images from text descriptions using StyleGAN2. The developed framework uses BERT embeddings, which map text to the latent space of StyleGAN2, enabling the production of aligned facial images at a resolution of 1024x1024 pixels corresponding to input descriptions. This model performs better than earlier approaches, achieving a 57% similarity to ground truth images and a face semantic distance of 0.92. Promising results are achieved in both image quality and alignment stages.

Ramzan et al. [15] raised the challenge of generating realistic images from text descriptions in their research. Based on this work, a Recurrent Convolutional Generative Adversarial Network (RC-GAN) has been proposed that can generate semantically consistent images by converting visual concepts into pixels. Trained on the Oxford-102 flowers dataset, their model achieves an inception score of 4.15 and a PSNR value of 30.12 dB, demonstrating its ability to generate realistic flower images from captions. Future work may involve expanding this model into multiple datasets for wider applications.

III. PROPOSED MODEL

In this paper, we present the detailed implementation of our forensic face sketching and identification system, which integrates advanced generative models and facial recognition techniques. Our system pipeline is divided into two key phases: automated sketch generation and sketch-based face identification. We leverage state-of-the-art models and frameworks to achieve high accuracy and efficiency.

To implement this system, several resources and tools are utilized. The CelebA dataset serves as the reference gallery for identification, containing over 200,000 celebrity face images with annotated attributes. For sketch generation, the model uses Stable Diffusion in combination with ControlNet (specifically, the lllyasviel/control_v11p_sd15_scribble variant), which allows conditioning the generation on edge maps or scribbles. The implementation leverages the TediGAN repository for handling prompt-based image manipulation and sketching. The entire pipeline is executed in Google Colab, with intermediate outputs and datasets stored and accessed via Google Drive for ease of collaboration and resource management.

For the identification phase, the system uses the FaceNet architecture, specifically the InceptionResnetV1 model pretrained on VGGFace2, which is known for producing robust facial embeddings. Face detection and alignment are performed using MTCNN, and similarity is calculated using cosine similarity between embeddings. The Python environment is supported with libraries such as facenet-pytorch, matplotlib, pickle, and numpy for efficient computation, embedding storage, and result visualization.

A. Environment and Setup

The initial step in developing the proposed system involves setting up the necessary environment and preparing the dataset for processing. The CelebA (CelebFaces Attributes) dataset, a large-scale face attributes dataset with over 200,000 celebrity images, was downloaded from Kaggle. Due to its size, the dataset was extracted and uploaded to Google Drive, allowing seamless integration with Google Colab and avoiding repeated data transfers.

Once uploaded, the dataset was programmatically accessed in Google Colab using the drive module from google.colab. A basic data cleaning routine was implemented to remove corrupted files, ensure consistent image dimensions, and filter out low-quality samples. This step was crucial to avoid inconsistencies during face detection and embedding generation in the later stages.

Google Colab was chosen as the development platform due to its cloud-based GPU support, compatibility with Python-based machine learning libraries, and ease of integration with Google Drive. Colab enables real-time code execution, visualization, and resource sharing, making it ideal for prototyping and executing deep learning pipelines without requiring local GPU infrastructure. Libraries such as torch, facenet-pytorch, matplotlib, and PIL were pre-installed or installed via pip to build and run the model seamlessly within Colab.

B. Phase 1: Sketch Generation Phase

The sketch generation component of the system is designed to synthesize realistic facial sketches based on descriptive textual prompts and structural outlines. This phase was entirely implemented on Google Colab, a cloud-based platform that offers high-performance GPUs, allowing the model to be trained and executed without the need for local hardware. Prior to execution, the runtime environment was configured by setting the hardware accelerator to GPU mode, which ensured compatibility with computationally demanding tasks such as image synthesis and model loading. The CelebA dataset, previously uploaded to Google Drive, was mounted into Colab's file system, enabling efficient access to reference images and storage of generated outputs. Essential libraries such as diffusers, transformers, and facenet-pytorch were installed directly using pip commands to support the functioning of the sketch generation pipeline.

The core model used for sketch generation is Stable Diffusion, a latent diffusion model (LDM) that generates images by progressively denoising a latent representation. To enhance its control and precision over structural features, Stable Diffusion is integrated with ControlNet, specifically the lllyasviel/control_v11p_sd15_scribble model. ControlNet allows the diffusion process to be conditioned not only on textual descriptions but also on external visual inputs, such as edge maps or scribbles. This dual-conditioning mechanism ensures that the generated sketches are both semantically accurate and structurally aligned with the intended layout.

The generation process begins with the input of a descriptive text prompt—such as "a young woman with wavy hair and round glasses"—and a corresponding scribble or edge map. The scribble acts as a spatial constraint and is processed through ControlNet, which injects this structural information at multiple stages of the diffusion model's internal layers. The text prompt is tokenized and encoded into a latent embedding via a language encoder (typically CLIP), which is fused with the image features using cross-attention mechanisms. Initially, the model starts from a latent representation of pure Gaussian noise. Through a series of denoising steps, guided jointly by the text and structural map, the model refines this noise into a coherent sketch image in the latent space. The final latent output is decoded back into an image using a pretrained Variational Autoencoder (VAE), resulting in a high-quality sketch that conforms to both the scribble's geometry and the prompt's semantic content.

This sketch generation approach contrasts with traditional Generative Adversarial Networks (GANs), which rely on a generator-discriminator framework where the generator tries to create realistic images while the discriminator attempts to distinguish between real and fake samples. While GANs have shown remarkable performance in face generation, they often struggle with fine-grained control and can be unstable during training. In contrast, diffusion models like Stable Diffusion operate by learning to reverse the process of adding noise to data, resulting in more stable training and better fidelity in high-resolution outputs. Furthermore, the incorporation of ControlNet provides a modular way to guide the generation process with structural inputs, which is difficult to achieve reliably with standard GAN architectures.

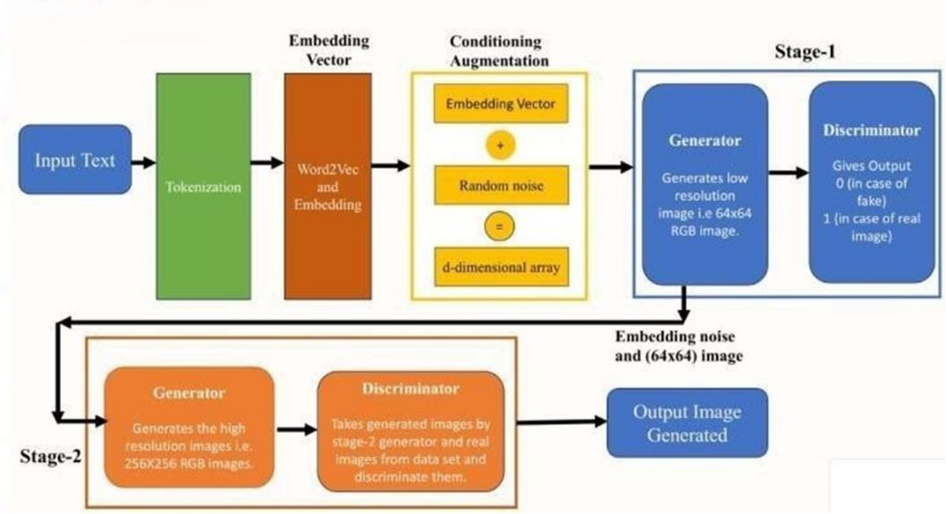

Embedding Vector
Conditioning Augmentation
Stage-1
Input Text
Tokenization
Word2Vec and Embedding
Embedding Vector
+
Random noise
=
d-dimensional array
Generator
Generates low resolution image i.e 64x64 RGB image.
Discriminator
Gives Output 0 (in case of fake) 1 (in case of real image)
Embedding noise and (64x64) image
Stage-2
Generator
Generates the high resolution images i.e. 256X256 RGB images.
Discriminator
Takes generated images by stage-2 generator and real images from data set and discriminate them.
Output Image Generated

Fig. 1. Diagrammatic View of Phase 1 - Sketch Generation

C. Embedding Phase

The embedding phase is a critical component of the face identification system, as it transforms high-dimensional image data into compact, discriminative representations known as facial embeddings. These embeddings capture the essential features of a face in a lower-dimensional vector space, where similar faces are positioned closer together. This representation facilitates efficient similarity comparison and retrieval, which is essential for identifying matches between sketches and real images in the CelebA dataset.

To generate these embeddings, all face images from the CelebA dataset were first preprocessed using the Multi-task Cascaded Convolutional Neural Network (MTCNN). This step involved detecting and aligning facial regions to a consistent scale and orientation. Such preprocessing ensures that the embeddings are not affected by background noise, varying face positions, or inconsistent cropping, which could otherwise reduce the model’s effectiveness.

Once the face regions were isolated, they were passed through the InceptionResnetV1 model, a robust convolutional neural network architecture provided by the FaceNet framework. This model, pretrained on the VGGFace2 dataset, outputs a fixed-length 512-dimensional embedding vector for each face. These embeddings are designed to be highly discriminative: the Euclidean or cosine distance between vectors of the same identity is minimized, while those of different identities are pushed further apart. The model achieves this by employing a triplet loss function during training, which optimizes the embedding space to reflect facial similarity.

After embedding generation, the feature vectors were stored using Python’s pickle module along with their corresponding image labels. This approach allows for rapid loading and lookup during the identification phase, eliminating the need to recompute embeddings at runtime. The stored embeddings form a searchable database against which the sketch embeddings can later be compared.

This phase significantly enhances computational efficiency, as searching through 200,000 full-size images would be infeasible in real-time. Instead, the system operates in a reduced-dimensional embedding space, enabling fast and accurate similarity comparisons. Additionally, the embeddings are invariant to minor changes in facial expression, lighting, and pose, making them ideal for matching against the more abstract representations found in generated sketches.

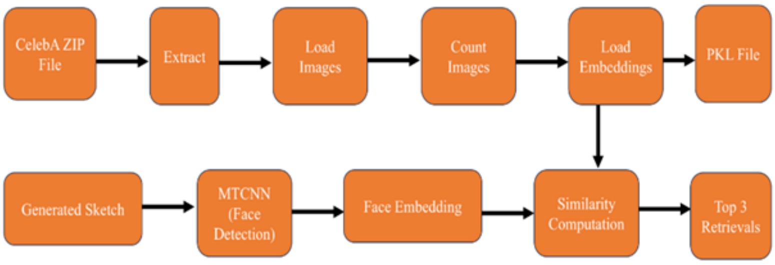


Fig. 2. Diagrammatic View of Phase 2 - Sketch Identification

D. Phase 2: Sketch Identification Phase

The final phase of the system—identification—aims to match the generated sketch image with the most similar faces from the CelebA dataset based on deep feature similarity. The process integrates precomputed facial embeddings with a runtime-generated sketch embedding, comparing them using cosine similarity to retrieve the top three closest matches. The overall pipeline is visually represented in Fig. X (reference to the uploaded diagram), which illustrates the sequential stages of data handling, embedding, and similarity-based retrieval.

The process begins with the CelebA ZIP file, which contains the large-scale dataset of celebrity face images. These images are extracted and subsequently loaded into memory using standard Python libraries such as os, cv2, and PIL. During loading, all images are resized and verified for consistency. The total number of valid images is then counted, primarily for memory and performance optimization. Following this, the dataset undergoes a face detection and embedding phase, where the aligned face crops are passed through InceptionResnetV1 (from the facenet-pytorch library) to extract 512-dimensional feature embeddings. These embeddings are saved to a. pkl (pickle) file, enabling rapid access during runtime without recomputing features.

Simultaneously, when a generated sketch is provided as input, it undergoes face detection using the MTCNN detector. MTCNN performs three stages of processing: proposal generation, refinement, and landmark localization. It extracts the primary face region from the sketch and aligns it for consistency with the original CelebA images. Once the face region is extracted, it is passed through the same InceptionResnetV1 model to produce its corresponding feature embedding.

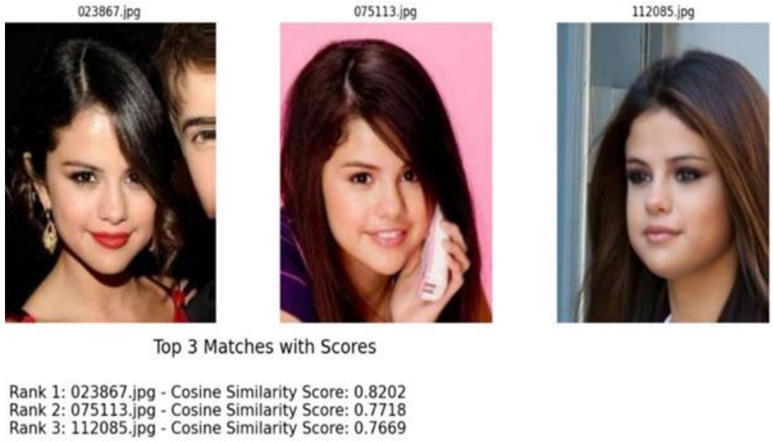
At this point, the precomputed embeddings from the CelebA dataset are loaded from the .pkl file. The newly generated sketch embedding is compared against all these embeddings using cosine similarity, a mathematical metric that quantifies the angular distance between two vectors. The core logic of similarity computation uses efficient vectorized operations with numpy or sklearn.metrics.pairwise.cosine_similarity, depending on implementation. A similarity score is computed between the sketch and every image embedding in the dataset. The embeddings are then sorted based on similarity, and the top 3 most similar faces are selected.

These top matches are visualized using matplotlib, displaying the sketch alongside the top retrievals for easy comparison. The output is both qualitative (image display) and quantitative (similarity scores), providing transparency in the system's decision-making. The entire identification phase runs on Google Colab, leveraging its GPU-accelerated runtime to significantly reduce computation time during embedding and similarity search.

The modular and preprocessed structure of the identification system ensures that results are retrieved quickly and accurately, with high scalability across larger datasets. The use of deep embeddings rather than pixel-level comparison or handcrafted features gives the system robustness to noise, variations in style (e.g., sketch vs. real), and facial distortions, thereby making it highly suitable for practical applications such as law enforcement and identity verification from forensic sketches.

IV. RESULTS

The results of this study are categorized into three core components: sketch generation accuracy, sketch identification using cosine similarity, and analysis of similarity distribution through a histogram.



Top 3 Matches with Scores

```
Rank 1: 023867.jpg - Cosine Similarity Score: 0.8202
Rank 2: 075113.jpg - Cosine Similarity Score: 0.7718
Rank 3: 112085.jpg - Cosine Similarity Score: 0.7669
```

Fig. 4. Top 3 Similar images and their Cosine Similarity score printed.

A. Sketch Generation Phase

The first stage involved generating a sketch based on a detailed textual prompt describing facial features such as a youthful, symmetrical face, almond-shaped brown eyes, thick arched eyebrows, a medium-sized straight nose, full lips with a Cupid's bow, and long, dark hair. Using Stable Diffusion with ControlNet, the system produced a sketch that visually aligned with the prompt. While minor stylizations and exaggerations typical of generative models were present, the overall structure and key descriptors were faithfully reproduced. The facial proportions, expression, and shape of the eyes, lips, and jawline bore a clear resemblance to the described individual, suggesting high descriptive fidelity in the sketch generation phase.

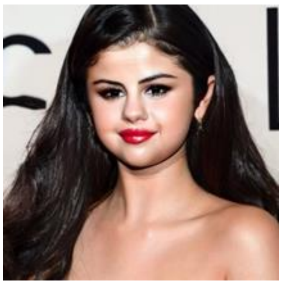


Fig. 3. Generated Image from Gan Model using text as input.

B. Sketch Identification Phase

In the second phase, the generated sketch was passed through an identification system that embedded the image using the InceptionResnetV1 model pretrained on VGGFace2. This embedding was then compared with a set of precomputed embeddings from the CelebA dataset using cosine similarity. The results yielded three top matches, with cosine similarity scores of 0.8202, 0.7718, and 0.7669 respectively. All three scores were above a strict threshold of 0.75, marking them as confident matches. Notably, the highest match (image 023867.jpg) had a similarity of 0.8202, which is considerably higher than the average, indicating a strong resemblance between the sketch and the retrieved image. Visual verification of these images further confirmed that the system accurately identified individuals with features closely matching those depicted in the sketch.

To gain deeper in insight into the behavior of the similarity scores, a histogram was plotted showing the distribution of cosine similarities between the sketch and all dataset images. The x-axis of the histogram represented the cosine similarity values, ranging from approximately -0.6 to 0.8, while the y-axis indicated the frequency of those values. The resulting distribution formed a bell-shaped curve centered around zero, indicating that the majority of dataset images had little to no similarity with the sketch. A dashed red vertical line was plotted at the Top-1 similarity score of 0.8202, clearly positioned far to the right of the main distribution. This stark separation visually confirmed that the top match was not only the best but also statistically significant in its resemblance. Only three images had scores exceeding the 0.75 threshold, reinforcing the system's high precision and low false-positive rate.

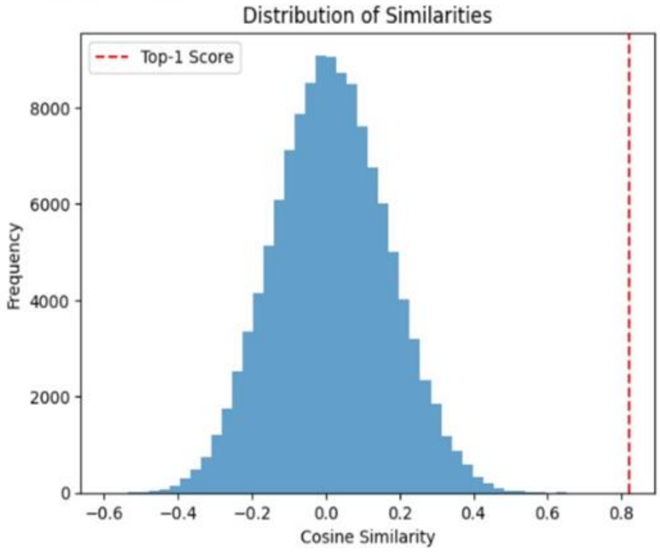


Fig. 5. Histogram: Distribution of Similarities

In summary, the results demonstrate that the proposed two-phase system performs effectively in both generating a sketch that preserves descriptive fidelity and accurately identifying the most similar real faces from a dataset. The cosine similarity scores and histogram analysis together provide quantitative and statistical validation of the system's robustness.

V. CONCLUSION AND FUTURE SCOPE

This project demonstrates an end-to-end pipeline for forensic face identification using AI, combining Stable Diffusion with ControlNet for generating realistic sketches from descriptive text prompts, and identifying matches via facial embeddings using the CelebA dataset. The system effectively transforms linguistic descriptions into visual sketches and matches them against a large dataset using FaceNet and cosine similarity, enabling a practical application for cases lacking photographic evidence. Future enhancements could include accepting voice-based descriptions, incorporating multilingual support for broader accessibility. Additionally, while the current model offers strong accuracy, it lacks optimization in terms of inference speed. This can be improved in future iterations by integrating lightweight embedding models or using model quantization techniques to enhance speed without compromising performance.

REFERENCES

[1] H. J. Jalan et al., "Suspect Face Generation System," in Proceedings of the 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), 2020.

[2] S. Patil and D. C. Shubhangi, "Forensic sketch based Face Recognition using Geometrical Face Model," in Proceedings of the 2nd International Conference for Convergence in Technology (I2CT), 2017.

[3] D. Dixit and A. Raj, "Face Sketch Maker And Criminal Identifier," International Research Journal of Modernization in Engineering Technology and Science, vol. 6, 2024.

[4] M. Jha et al., "Creation of Face sketch Aiding in Forensic Investigation based on Textual Description," in Proceedings of the Fourth International Conference on Inventive Systems and Control (ICISC 2020), 2020.

[5] H. Wu et al., "Towards Criminal Sketching with Generative Adversarial Network," in Proceedings of the 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), 2020.

[6] S. Reed et al., "Generative Adversarial Text to Image Synthesis," in Proceedings of the 33rd International Conference on Machine Learning, 2016.

[7] T. Xu et al., "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2027,

pp. 1316-1324.

[8] S. Ouyang, T. Hospedales, Y.-Z. Song and X. Li, ―Cross-Modal Face Matching: Beyond Viewed Sketches," Lecture Notes in Computer Science, vol. 9004, 2018.

[9] A. Thakare et al., "Implementation of Digital Forensics Face Sketch Recognition using Fusion based Deep Learning Convolutional Neural Network," International Research Journal of Modernization in Engineering Technology and Science, vol. 5, 2023.

[10] C. Galea and R. A. Farrugia, "Forensic Face Photo-Sketch Recognition Using a Deep Learning-Based Architecture," IEEE Signal Processing Letters, vol. 24, 2017.

[11] I. Zhao et al., "Generating Photographic Faces From the Sketch Guided by Attribute Using GAN," IEEE Access, vol. 7, 2019.

[12] A. Kumar et al., "Realistic face generation using a textual description," in Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021), 2021.

[13] H. Zhang et al., "StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks," International Research Journal of Modernization in Engineering Technology and Science, vol. 6, 2018.

[14] D. M. Ayanthi and S. Munasinghe, "Text-To-Face Generation with Stylegan2," in Proceedings of the International Conference on Communication System, Computing and IT Applications (CSCITA), 2022, pp. 49-64.

[15] S. Ramzan et al., "Text-to-Image Generation Using Deep Learning," MDPI, Eng. Proc, vol. 4, 2022.

[16] J. E. Martis et al., "Text-to-Sketch Synthesis via Adversarial Network," Tech Science Press, vol. 76, CMC, 2023.

[17] S. Klum et al., "Sketch Based Face Recognition: Forensic vs. Composite Sketches," in Proceedings of the International Conference on Biometrics (ICB), 2013.

[18] A. V. Vidhyap, "Forensic Sketch to Real Image using DCGAN," Procedia Computer Science, vol. 8, 2023.

[19] S. N. Bushr and K. U. Maheswar, "Crime Investigation using DCGAN by Forensic Sketch-to-Face Transformation (STF)," in Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021), 2021.

[20] D. M. Mohammed et al., "Forensic Facial Reconstruction from Sketch in Crime Investigation," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 15, 2024.

[21] Z. Liu et al., "Deep Learning Face Attributes in the Wild," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3730–3738.

[22] T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

[23] A. Adhikari, D. Jariwala, A. Nigade, V. Kulkarni, S. Karande, and J. Patil, “Exploring GANs for Image Synthesis and Recognition in Forensic Contexts,” in Proceedings of the 2025 12th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, Apr. 2025, pp. 1887–1892.