World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Forensic sketch generation using Gen-AI

Kavitha Soppari *, Asanabada Pruthvi, Nallamasa Uday Kiran and Vinukonda Stephen Moses

Department of CSE (AI & ML), ACE Engineering College Hyderabad, India.

World Journal of Advanced Research and Reviews, 2026, 30(01), 164-171

Publication history: Received on 22 February 2026; revised on 01 April 2026; accepted on 03 April 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.1.0791

Abstract

Forensic sketch generation plays a crucial role in criminal investigations where no photographic evidence is available. Traditional sketching methods rely heavily on skilled artists and subjective interpretations of eyewitness descriptions, which often leads to inconsistencies and delays. This project proposes an automated system that leverages Generative AI, specifically diffusion-based models, to generate realistic forensic sketches from textual descriptions. The system utilizes Stable Diffusion XL for high-quality image generation and integrates biometric and semantic feature extraction using InsightFace and BiSeNet. A hybrid matching mechanism using FAISS is employed to compare generated sketches with a mugshot database, providing ranked suspect identification. The proposed framework improves accuracy, scalability, and efficiency by combining text-to-image generation with multimodal face matching, making it a practical solution for modern forensic applications.

Keywords: Forensic Sketch Generation; Generative AI; Diffusion Models; Stable Diffusion; Multimodal Matching; Computer Vision

1. Introduction

Forensic sketching is an essential component in criminal investigations, particularly when visual evidence such as photographs or CCTV footage is unavailable. Traditionally, forensic sketches are created manually by artists based on eyewitness descriptions. However, this process is time-consuming, subjective, and highly dependent on both the artist's expertise and the witness's memory.

With the advancement of artificial intelligence, automated approaches have emerged to address these limitations. Early methods used Generative Adversarial Networks (GANs) for face generation, but they often suffered from instability and poor identity preservation.

Recent developments in diffusion models have significantly improved image generation quality. These models provide better realism, stability, and alignment with textual inputs. By integrating diffusion-based generation with biometric and semantic analysis, it is possible to create an efficient end-to-end forensic system.

This project focuses on developing such a system that transforms textual descriptions into realistic sketches and identifies potential suspects through similarity matching.

* Corresponding author: Kavitha Soppari

2. Literature survey

2.1. Text2FaceGAN: Face Generation from Fine-Grained Textual Descriptions

Authors: Osaid Rehman Nasir et al.

This work focuses on generating human face images directly from textual descriptions. The authors addressed the major challenge of lack of paired text-image datasets by automatically generating captions from the CelebA dataset. They used a Deep Convolutional Generative Adversarial Network (DC-GAN) along with a matching-aware discriminator to synthesize images. The model was able to generate faces based on attributes such as gender, hair type, and facial features. However, the generated images were of low resolution (64×64) and lacked fine identity details. Additionally, the use of older text encoding techniques limited its ability to understand complex descriptions. This work laid the foundation for text-to-face generation but highlighted the need for more advanced models for higher realism and identity preservation.

2.2. Identity-Preserved Model for Face Sketch-Photo Synthesis (IPAM)

Authors: Ye Lin et al.

This paper introduces the concept of identity preservation, which is critical in forensic applications. The proposed model ensures that the generated image is not only realistic but also represents the same identity as the input. The system uses an identity constraint loss, where the generated image is compared with the real image using a pre-trained face recognition model. The similarity between their embeddings is minimized to maintain identity consistency. Although this work focuses on sketch-to-photo synthesis, it provides a crucial idea for forensic systems: ensuring that generated outputs are biometrically consistent. However, it does not handle text-based input, which limits its direct applicability to eyewitness-based systems.

2.3. TediGAN: Text-Guided Diverse Face Image Generation

Authors: Weihao Xia et al.

TediGAN represents a major advancement by using a pre-trained StyleGAN model instead of training from scratch. The key idea is GAN inversion, where textual descriptions are mapped into the latent space of a pre-trained generator.This approach allows High-resolution image generation (up to 1024×1024), Better control over facial attributes, Multi-modal input support (text, sketch, etc.) Additionally, the authors introduced the Multi-Modal CelebA-HQ dataset, which includes images, sketches, and textual descriptions.Despite its strengths, the model still inherits limitations of GANs such as, Training instability, Difficulty in preserving exact identity in all cases

2.4. Semantic Text-to-Face GAN (ST2FG)

Authors: Manan Oza et al.

This work improves usability by introducing a more interactive system. Instead of simple attribute inputs, it uses BERT-based text encoding to understand detailed natural language descriptions. A key feature of this model is iterative refinement, where users can modify generated images by giving additional instructions. This closely mimics the real-world process of forensic sketching. The system uses an Affine Combination Module (ACM) to merge textual and visual features effectively. While this improves user interaction and flexibility, the model still depends on GAN-based generation, which can limit output realism compared to newer approaches like diffusion models.

2.5. CLIP4Sketch: Enhancing Sketch-to-Mugshot Matching

Authors: Kushal Kumar Jain et al.

This paper focuses on improving the matching phase rather than generation. It uses diffusion models (DDPM) to create a large dataset of synthetic sketches from mugshots.The system integrates CLIP embeddings for semantic alignment, AdaFace embeddings for identity preservation, ControlNet for structural consistency. The generated dataset is then used to train better face recognition systems, improving sketch-to-photo matching accuracy. This work demonstrates Superiority of diffusion models over GANs, Importance of multi-modal learning. However, it does not directly generate sketches from text, focusing instead on dataset generation and matching.

Objectives

The main objectives of this project are to develop an automated system for generating forensic sketches from textual descriptions, to improve sketch realism using diffusion-based generative models, to ensure identity preservation using biometric feature extraction, to integrate semantic feature analysis for enhanced matching, to implement a hybrid similarity scoring mechanism for suspect identification, to reduce dependency on manual sketch artists, to provide a scalable and efficient forensic solution

3. Methodology

The proposed forensic sketch generation framework integrates text-based input processing, diffusion-based image synthesis, biometric feature extraction, semantic analysis, and hybrid similarity matching to improve the accuracy and reliability of suspect identification. The methodology is designed to efficiently transform eyewitness descriptions into realistic facial sketches and compare them against a mugshot database. By combining **Stable Diffusion XL** for image generation, **InsightFace** for biometric embedding extraction, **BiSeNet** for semantic feature analysis, and **FAISS** for similarity search, the system provides a robust end-to-end pipeline that bridges the gap between subjective human descriptions and objective forensic identification.

3.1. System workflow

3.1.1. Data Input & Preprocessing

Textual Description Input → Text Cleaning → Attribute Extraction → Prompt Construction

- The user (eyewitness or investigator) provides a textual description of the suspect.
- The system processes the input by removing noise and structuring it into meaningful attributes such as face shape, hair, eyes, and other features.
- The structured data is converted into a detailed prompt suitable for the generative model.

3.1.2. Stage 1: Sketch Generation (Diffusion Model)

Text Prompt → Diffusion Model (Stable Diffusion XL) → Generated Forensic Sketch

- A pre-trained diffusion model generates a high-resolution facial sketch based on the input prompt.
- The model ensures realism, coherence, and alignment with the textual description.
- This stage replaces traditional manual sketching with an automated and scalable solution.

3.1.3. Stage 2: Feature Extraction & Matching

Generated Image → Biometric Feature Extraction → Semantic Feature Extraction → Similarity Matching

- **Biometric Extraction:** InsightFace extracts facial embeddings representing identity features.
- **Semantic Extraction:** BiSeNet identifies attributes such as skin tone, hair color, eye color, and beard presence.
- **Similarity Matching:**
    - Features are compared with a mugshot database using FAISS.
    - A hybrid scoring mechanism combines biometric similarity and semantic attributes.

Key Components

- **Input Type:** Textual Description (Eyewitness Input)
- **Generation Model:** Stable Diffusion XL
- **Biometric Model:** InsightFace
- **Semantic Model:** BiSeNet
- **Search Engine:** FAISS (Facebook AI Similarity Search)
- **Feature Fusion:** Hybrid Scoring (Biometric + Semantic)
- **Frameworks:** PyTorch, Hugging Face
- **Libraries:** OpenCV, NumPy, Matplotlib
- **Development Tools:** Jupyter Notebook / VS Code
- **Dataset:** Mugshot / Face Image Database

3.2. Proposed system

The proposed system introduces an AI-powered forensic sketch generation and suspect identification framework designed to automate and enhance criminal investigation processes.

Unlike traditional manual sketching methods, the system leverages **diffusion-based generative models** to convert eyewitness descriptions into realistic facial images. It further integrates biometric and semantic analysis to match generated sketches with a database of suspects.

This approach overcomes key limitations of existing systems by:

- Eliminating human dependency in sketch creation
- Improving sketch realism and consistency
- Providing direct linkage between sketch generation and suspect identification

3.3. System Overview

The proposed system includes:

- **Text-Based Input System –** Accepts structured or free-text suspect descriptions from users.
- **Sketch Generation Module –** Uses Stable Diffusion XL to generate high-quality forensic sketches from textual input.
- **Biometric Feature Extraction Module –** Extracts identity-based embeddings using InsightFace.
- **Semantic Feature Extraction Module –** Identifies facial attributes such as skin tone, hair color, and facial features using BiSeNet.
- **Hybrid Matching Framework –** Combines biometric and semantic features to improve matching accuracy.
- **FAISS-Based Search Engine –** Performs fast and scalable similarity search across large databases.
- **Performance Monitoring Module –** Evaluates system accuracy using ranking metrics and similarity scores.

3.4. System Operation

3.4.1. Data Input Phase

User enters suspect description → Text preprocessing → Prompt generation

3.4.2. Generation Phase

Prompt → Diffusion Model → High-resolution forensic sketch

3.4.3. Analysis & Matching Phase

- Generated image → Biometric embedding extraction
- Semantic feature extraction → Attribute identification
- FAISS search → Retrieve similar faces from database
- Hybrid scoring → Rank candidates

3.4.4. Decision & Output Phase

- Top matching suspects are displayed
- Confidence scores are generated
- Final output includes:
  - Generated sketch
  - Ranked suspect list

4. Applications

The proposed system has wide-ranging applications in forensic and security domains:

- **Criminal Investigation**
- Helps law enforcement agencies generate suspect sketches from eyewitness descriptions and identify potential matches.
- **Missing Person Identification**
- Assists in reconstructing facial appearances based on verbal descriptions.
- **Surveillance & Security Systems**
- Enhances monitoring systems by enabling suspect identification from descriptive inputs.

- **Border Control & Immigration**
- Supports identity verification and suspect tracking in high-security zones.
- **Intelligence & Defense Systems**
- Provides advanced tools for suspect profiling and identification in sensitive operations.
- **Digital Forensics**

Aids in analyzing and matching facial data across various databases.

5. Algorithms

The system utilizes multiple deep learning and hybrid matching algorithms for forensic sketch generation and suspect identification. It integrates diffusion-based image synthesis, biometric embedding extraction, semantic feature analysis, and similarity-based retrieval.

5.1. Sketch Generation Algorithm (Stage 1)

Purpose: Generate a realistic forensic sketch from textual description.
Algorithm Steps:

1. Capture textual description from user input.
2. Preprocess and structure the text into a detailed prompt.
3. Provide prompt and negative prompt to the diffusion model.
4. Pass input through **Stable Diffusion XL pipeline**.
5. Generate high-resolution facial image.
6. Save generated image for further processing.

5.2. Biometric Feature Extraction Algorithm

Purpose: Extract identity-preserving facial embeddings.
Algorithm Steps:

- Input generated sketch image.
- Detect face using InsightFace detector.
- Extract normalized facial embedding vector.
- Predict auxiliary attributes such as age and gender.
- Store embedding for similarity comparison.

5.3. Semantic Feature Extraction Algorithm

Purpose: Extract visual facial attributes for improved matching.
Algorithm Steps:

- Resize input image to required dimensions.
- Pass image through **BiSeNet segmentation model**.
- Generate facial parsing mask.
- Extract features such as:
  - Skin tone (HSV values)
  - Hair color (RGB values)
  - Eye color
  - Beard presence
- Store semantic attributes for hybrid scoring.

5.4. Hybrid Similarity Matching Algorithm (Stage 2)

Purpose: Identify potential suspects using combined biometric and semantic similarity.
Algorithm Steps:

- Input query embedding and semantic attributes.
- Search FAISS index to retrieve top-N nearest embeddings.
- For each candidate:
  - Compute biometric distance
  - Compute skin similarity
  - Compute hair similarity

    - Apply penalties (beard mismatch, gender mismatch)
- Calculate final hybrid score:
    - Final Score = Biometric Distance + Semantic Penalties
- Sort candidates based on final score.
- Select top-K matches as output.

5.5. Feature Fusion Algorithm

Purpose: Combine biometric and semantic features for robust matching.
Algorithm Steps:

- Extract biometric embedding vector (F_bio).
- Extract semantic attributes (F_semantic).
- Normalize feature representations.
- Compute similarity scores separately.
- Apply weighted penalty-based fusion:
    - Skin penalty
    - Hair penalty
    - Beard penalty
    - Gender penalty
- Combine scores into a single hybrid score.
- Output fused similarity measure.

5.6. Performance Evaluation Algorithm

Purpose: Evaluate system effectiveness and matching accuracy.
Algorithm Steps:

- Compare retrieved matches with known identities (if available).
- Calculate Top-K accuracy (Top-1, Top-5).
- Measure similarity score distributions.
- Evaluate system response time (generation + matching).
- Analyze robustness for different input descriptions.

---

6. Result

6.1. System Performance Evaluation

The proposed forensic sketch generation and suspect identification framework was evaluated using a curated mugshot dataset and generated sketches based on diverse textual descriptions.

Testing confirmed:

- Successful generation of **high-quality, realistic forensic sketches** from structured and unstructured textual inputs.
- Effective extraction of **biometric embeddings and semantic features** from generated images.
- Accurate retrieval of **top matching suspects** using FAISS-based similarity search.
- Improved matching performance through **hybrid scoring**, combining biometric similarity with semantic attributes.
- Consistent system performance across **varied descriptions, facial attributes, and image conditions**.

The integrated pipeline demonstrated efficient end-to-end operation, enabling rapid transformation from **textual description → sketch generation → suspect identification**, while maintaining reliable accuracy and scalability.

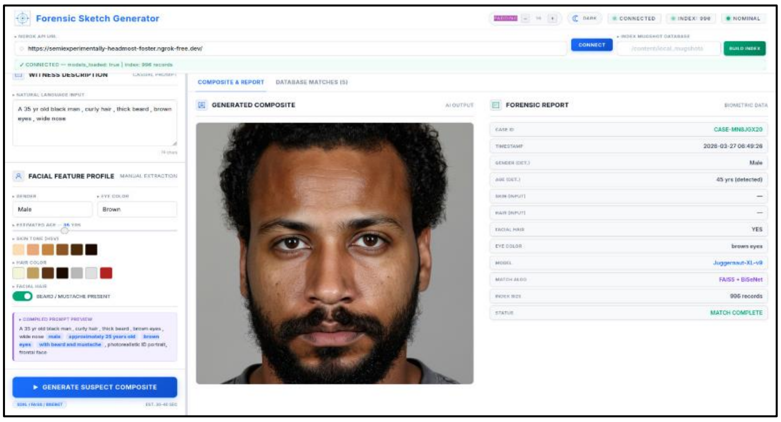

Figure 1 Generating Image From Sketch

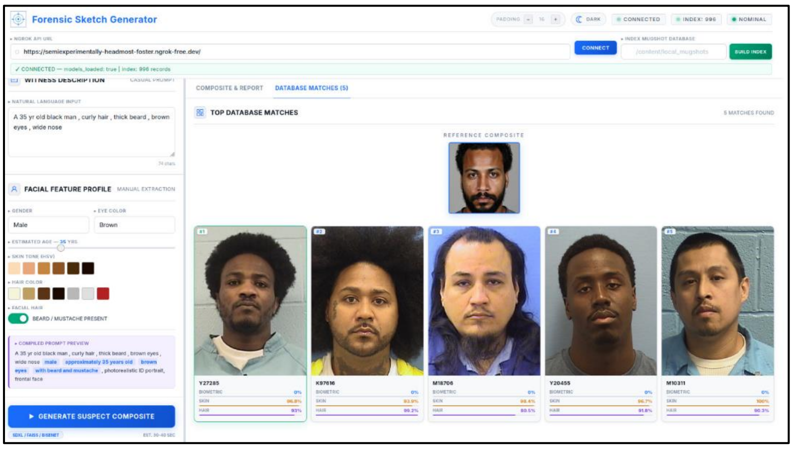

Figure 2 Faces Similar To The Generated Sketch

7. Conclusion

The proposed forensic sketch generation and suspect identification framework provides a robust, efficient, and scalable solution for modern criminal investigation challenges. By integrating diffusion-based image generation, biometric feature extraction, and semantic analysis within a unified pipeline, the system significantly improves the accuracy and realism of generated sketches compared to traditional manual methods.

The use of Stable Diffusion XL enables high-quality image synthesis from textual descriptions, while InsightFace and BiSeNet ensure effective identity representation and attribute extraction. The incorporation of a hybrid similarity matching mechanism using FAISS enhances suspect retrieval by combining biometric and semantic features.

The system demonstrates reliable performance across diverse inputs and conditions, enabling efficient transformation from eyewitness description to actionable suspect identification. This makes it a practical and valuable tool for real-world forensic and law enforcement applications.

Further enhancement

The following enhancements can further improve the system's performance and real-world applicability:

- **Iterative Sketch Refinement** – Enabling step-by-step modification of generated sketches without altering identity consistency.
- **Multimodal Embedding Fusion** – Integrating advanced models like CLIP for better alignment between text and image representations.
- **Advanced Semantic Feature Extraction** – Expanding detection to include scars, tattoos, facial marks, and accessories.

- **Scalability Improvements** – Implementing advanced FAISS indexing techniques (IVF, HNSW) for large-scale databases.
- **Real-Time Deployment** – Optimizing the pipeline for faster inference and real-time forensic usage.
- **Explainable AI Integration** – Providing transparency in similarity scoring and decision-making process.
- **Bias Reduction and Fairness Analysis** – Ensuring consistent performance across different demographics and datasets.
- **Web and Mobile Deployment** – Developing user-friendly applications for field-level forensic use.

Compliance with ethical standards

Acknowledgments

The authors would like to thank the Department of CSE (AI & ML), ACE Engineering College, Hyderabad, for providing the necessary resources and support to carry out this research. No external funding was received for this study.

Disclosure of conflict of interest

The authors declare that there are no financial or non-financial conflicts of interest related to the research presented in this manuscript.

References

[1] Nasir OR, et al. Text2FaceGAN: Face generation from fine-grained textual descriptions. In: Proceedings of IEEE BigMM; 2019.

[2] Lin Y, et al. An identity-preserved model for face sketch-photo synthesis. IEEE Signal Process Lett. 2020;27:1523–1527.

[3] Xia W, et al. TediGAN: Text-guided diverse face image generation and manipulation. In: Proceedings of IEEE/CVF CVPR; 2021.

[4] Oza M, et al. Semantic text-to-face GAN (ST2FG). arXiv preprint arXiv:2107.10756; 2021.

[5] Jain KK, et al. CLIP4Sketch: Enhancing sketch-to-mugshot matching. In: Proceedings of IEEE IJCB; 2024.