

# Composite Face Sketch Detection Using GAN-Based Generation and AI-Powered Identification

Sk. Arshiya Julia

Department of Information Technology  
Mahatma Gandhi Institute of Technology  
Gandipet, Hyderabad -500075,India

**Abstract** - Criminal identification through face sketches pose a significant challenge due to dependence on skilled forensic artists and the subjective nature of eyewitness-based descriptions. The proposed Composite Face Sketch Generator and Criminal Identification System addresses these limitations and concerns by combining a digital drag-and-drop face assembly interface with a Generative Adversarial Network (GAN) for realistic sketch synthesis and an AI-powered face matching module. The system enables and helps the law enforcement to rapidly generate digital composite faces without completely relying on specialist artists, and subsequently matches them against a criminal database using deep learning-based face recognition. The system employs a GAN architecture with a Generator using Dense, LeakyReLU, and Batch Normalization layers, and a Discriminator with Conv2D and Dropout layers. Losses are evaluated using Mean Squared Error (MSE) and Binary Cross Entropy (BCE). The experiment's results show us the sign of improved identification accuracy, faster suspect generation, and greater scalability compared to traditional manual methods.

**Index Terms**—Composite face sketch, criminal identification, GAN, face recognition, FaceNet, DeepFace, digital forensics, forensic sketch generation, face matching, deep learning.

## I. INTRODUCTION

The identification of criminal suspects through composite facial sketches is a fundamental component of modern law enforcement. Traditionally, forensic sketch artists recreate suspect faces from eyewitness descriptions, a process that is both time-intensive and prone to subjective interpretation errors. As documented in forensic literature, the accuracy of witness-based facial recall undergoes under severe conditions of stress, trauma, and time delay, leading to composite sketches that may poorly represent the actual suspect.

However, recent advancement and development in deep learning and computer vision offer many transformative solutions to these challenges. In particular, Generative Adversarial Networks (GANs), has truly shown its exceptional capacity for generating photorealistic face images from latent representations and conditional inputs. Similarly, AI-powered face recognition frameworks such as FaceNet and DeepFace have also achieved near-human accuracy in matching faces across diverse datasets, making them suitable for automated suspect identification.

The proposed Composite Face Sketch Generator and Criminal Identification System integrates both of these capabilities into a unified platform. It provides a drag-and-drop interface for digitally assembling composite faces, a GAN-based module for sketch synthesis and enhancement, and an AI face-matching engine for comparing generated composites against criminal databases. This system eliminates the need for trained forensic artists, accelerates the identification process, and improves the image identification accuracy by standardizing the output to a format, which is compatible with modern face recognition pipelines.

The key contributions of this research are summarized as follows:

- A digital composite face assembly interface enables investigators to create suspect sketches without specialist expertise.
- A GAN-based face generation module trained on public facial datasets for realistic sketch synthesis.
- An AI-powered face matching pipeline using FaceNet/DeepFace for accurate identification against criminal databases.
- A unified, end-to-end system that significantly reduces the investigation time and improves identification accuracy
- An evaluation and assessment of the proposed system using standard metrics including MSE, BCE, precision, recall, and match confidence scores.

## II. LITERATURE SURVEY

The convergence of digital forensics, generative modeling, and face recognition has inspired and paved the way, to a growing body of research in composite face sketch systems. The following review summarizes prior relevant works that inform the proposed system, as organized in Table I.

The early works in forensic face sketch synthesis were largely dependent on physiognomy-based feature assembly tools such as IDENTIKIT and E-FIT. Even though these systems improved upon manual drawing, they still relied heavily on subjective user input and lacked correlation with automated recognition engines. The significant emergence of deep learning introduced new paradigms for sketch-to-photo synthesis.

Jain et al. (2024) proposed CLIP4Sketch, a system that enhances sketch-to-mugshot matching by augmenting training datasets using diffusion models. This model highlights the importance of dataset diversity in improving cross-modal face matching performance. Tang et al. (2023) advanced identity-preserving face-to-photo synthesis using a hybrid CNN and GAN architecture, indicating that the combination of convolutional feature extraction with adversarial training, preserves facial identity across sketch and photograph domains.

Many works have explored GAN-based approaches specifically for forensic applications. The pix2pix framework, a conditional GAN architecture, has been widely applied to sketch-to-photo translation tasks, providing a strong baseline for supervised face generation. Several improvements to this framework, including cycle-consistent adversarial networks (CycleGAN), have enabled unpaired image-to-image translation, making sketch-to-photo conversion even more feasible even in the absence of exact paired training data.

In the domain of face recognition, models such as FaceNet and DeepFace have set state-of-the-art benchmarks on large-scale datasets including LFW and VGGFace2. These models generate compact facial embeddings that enable rapid similarity-based search in large databases, making them well equipped and suited for criminal identification applications.

TABLE I: Summary of Related Work on Face Sketch Generation and Recognition

Author & Year	Limitations	Existing Method	Proposed Approach	Dataset Used	Techniques	Tools	Metrics	Future Directions
Jain et al., 2024	Limited sketch diversity	Dataset augmentation	CLIP-based sketch matching	Mugshot DB	Diffusion Models, CLIP	Python, PyTorch	Rank-1 Accuracy	Real-time GAN sketch
Tang et al., 2023	Identity loss in synthesis	Basic CNN	Hybrid CNN-GAN	CUHK, VGG	CNN + GAN	TensorFlow	SSIM, FID	Video sketch synthesis
pix2pix (Isola et al.)	Requires paired data	Image-to-image GAN	Conditional GAN (pix2pix)	CUHK Face Sketch	cGAN	PyTorch	FID, SSIM	Unsupervised methods
CycleGAN (Zhu et al.)	Mode collapse issues	Unpaired translation	Cycle-consistent GAN	CUHK, CUFS	CycleGAN	Torch	KID, FID	Facial attribute control
FaceNet (Schroff et al.)	High compute cost	Triplet loss CNN	Deep face embedding	LFW, VGGFace2	Inception + Triplet	TensorFlow	TAR@FAR, Rank-1	Edge deployment

### III. PROPOSED METHODOLOGY

This section of information describes the complete architecture of the Composite Face Sketch Generator and Criminal Identification System. The system is structured as a multi-stage pipeline that progresses from interactive composite face creation to GAN-based synthesis and AI-powered matching. The system architecture is illustrated in Fig. 1.

#### A. System Overview

The proposed system consists of three primary modules: (1) the Composite Face Assembly Interface, (2) the GAN-Based Face Synthesis Module, and (3) the AI Face Recognition and Matching Module. These modules are integrated into a unified application framework built using Python with a React.js frontend and a Django/Flask backend.

#### B. Composite Face Assembly Interface

The visual interface provides investigators with an intuitive drag-and-drop environment for assembling suspect faces from a library of facial feature components. Feature categories include face shape, skin tone, eye type, nose shape, mouth style, hairstyle, and additional attributes such as facial hair and eyeglass style. Sliders allow for fine-grained control over feature parameters, enabling investigators to iteratively refine the composite based on witness feedback.

The assembled composite is represented as a structured feature vector that encodes the selected component identifiers and their associated parameter values. This set of features serves as the conditional input to the GAN synthesis module.

#### C. GAN-Based Face Synthesis

The GAN architecture consists of two adversarially trained networks: a Generator (G) and a Discriminator (D). The Generator takes a latent noise vector  $z$  combined with the composite feature vector  $c$  as input, and produces a synthesized face image:

$$G(z, c) = I_{synth}$$

##### 1) Generator Architecture

The Generator consists of a series of Dense layers followed by LeakyReLU activations and Batch Normalization layers. The final layer uses a Conv2D transposed convolution to upsample the latent representation into a face image. The architecture progressively increases spatial resolution while refining facial detail.

##### 2) Discriminator Architecture

The Discriminator classifies input images as either real (from the training dataset) or synthesized (from the Generator). It is composed of Conv2D layers with LeakyReLU activations and Dropout layers to improve robustness against overfitting during adversarial training. A sigmoid output layer produces a binary classification score.

### 3) Loss Functions

Training of the GAN is governed by two complementary loss functions. The Mean Squared Error (MSE) measures the pixel-level reconstruction fidelity between the synthesized and target face images:

$$L_{MSE} = (1/N) \sum ||I_{real} - I_{synth}||^2$$

The Binary Cross Entropy (BCE) loss governs the adversarial training of the Discriminator:

$$L_{BCE} = -[y \log(D(x)) + (1 - y) \log(1 - D(x))]$$

Training progresses in a dual-loop fashion, alternating between optimizing the Generator to minimize reconstruction loss and fool the Discriminator, and optimizing the Discriminator to correctly classify real versus synthesized images. Training progress is monitored via Epochs vs. Loss graphs.

### D. Data Collection and Preprocessing

Human facial images are sourced from publicly available datasets including LFW (Labeled Faces in the Wild), CUHK Face Sketch Database, and the VGGFace2 dataset. Preprocessing ensures all images are normalized to a standard resolution, converted to the appropriate color space (grayscale or RGB as required), and augmented using techniques including random horizontal flipping, rotation, and contrast jitter to increase training diversity.

The dataset is partitioned into training, validation, and test subsets at a ratio of 80:10:10 for unbiased evaluation.

### E. AI Face Matching Module

Once a composite face sketch is synthesized by the GAN, it is passed to the face recognition and matching module. This module generates a compact facial embedding vector  $e$  using a pre-trained deep learning model (FaceNet or DeepFace):

$$e = FaceNet(I_{synth})$$

The embedding is compared against a pre-indexed criminal database using cosine similarity. For each database entry  $e_i$ , the similarity score is computed as:

$$sim(e, e_i) = (e \cdot e_i) / (||e|| ||e_i||)$$

Matches with similarity scores exceeding a predefined threshold  $\theta$  are returned as candidate identifications, ranked in descending order of confidence:

$$Match = \{ e_i : sim(e, e_i) \geq \theta \}$$

TABLE II: Acronyms and Symbols Used in the Proposed System

Term / Symbol	Description
G	Generator network in the GAN architecture
D	Discriminator network in the GAN architecture
z	Latent noise vector (random input to Generator)
c	Composite feature vector (conditional input)
$I_{synth}$	Synthesized face image output from Generator
$I_{real}$	Real face image from training dataset
$L_{MSE}$	Mean Squared Error loss for reconstruction
$L_{BCE}$	Binary Cross Entropy loss for adversarial training
$e$	Facial embedding vector from FaceNet/DeepFace
$sim(e, e_i)$	Cosine similarity between query and database embedding
$\theta$	Confidence threshold for valid face matches
FaceNet	Deep face embedding model by Schroff et al.
DeepFace	Facebook AI face recognition framework
Dlib	C++ library for facial landmark detection
pix2pix	Conditional GAN for image-to-image translation
MSE	Mean Squared Error - pixel-level reconstruction metric

BCE	Binary Cross Entropy - adversarial classification loss
FID	Frechet Inception Distance - GAN image quality metric
SSIM	Structural Similarity Index Measure
mAP	Mean Average Precision for recognition ranking
Rank-1	Top-1 identification accuracy in face matching
LFW	Labeled Faces in the Wild benchmark dataset
CUHK	Chinese University of Hong Kong Face Sketch Dataset
VGGFace2	Large-scale face recognition dataset

#### IV. EXPERIMENTAL SETUP AND EVALUATION

This proposed system was derived and implemented using the Python programming language within a deep learning environment. The development hardware comprised a computer with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA GPU (RTX 3060 or equivalent) to support efficient GAN training and face embedding generation.

The software stack included React.js (frontend interface), Django/Flask (backend server), OpenCV and NumPy (image preprocessing), TensorFlow/PyTorch (GAN training and model deployment), FaceNet/DeepFace/Dlib (face recognition pipeline), and MySQL/PostgreSQL (criminal database management). The development and experimentation were conducted using Jupyter Notebook and Visual Studio Code.

##### A. Datasets

Three primary datasets were used for training and evaluation: (1) the LFW dataset for face recognition benchmarking, (2) the CUHK Face Sketch Database for sketch-to-photo translation training, and (3) the VGGFace2 dataset for large-scale face embedding pre-training. All datasets were preprocessed, augmented, and partitioned at an 80:10:10 ratio for training, validation, and testing.

##### B. Performance Analysis

The performance of the GAN synthesis module was evaluated using image quality metrics including Frechet Inception Distance (FID), which measures the statistical distance between distributions of real and generated images, and the Structural Similarity Index Measure (SSIM), which assesses perceptual similarity. Additionally, qualitative evaluation by human judges assessed realism, diversity, and resemblance of generated composites to target faces.

The face matching module was evaluated using Rank-1 identification accuracy (the proportion of queries for which the correct match appears as the top-ranked result), Mean Average Precision (mAP), and match confidence scores. Results demonstrated that integrating the GAN synthesis module with the face matching pipeline significantly improved identification accuracy compared to raw sketch-based matching baselines.

TABLE III: Performance Comparison of Proposed System vs. Existing Methods

Method	FID (↓)	SSIM (↑)	Rank-1 Acc. (%)	mAP (%)
Manual Sketch + ML	N/A	0.41	52.3	47.8
pix2pix GAN	62.4	0.67	71.2	65.4
CycleGAN	54.1	0.71	74.8	68.9
Proposed (GAN + FaceNet)	41.7	0.79	83.5	78.2

#### V. CONCLUSION AND FUTURE SCOPE

The proposed Composite Face Sketch Generator and Criminal Identification System presents a significant advancement in digital forensics by eliminating the dependence on skilled forensic sketch artists and integrating AI-powered face recognition. The system provides the investigators with a user-friendly drag-and-drop interface for composite face creation, a GAN module for generating photorealistic face images, and a deep learning-based matching engine that compares synthesized faces against criminal databases with high accuracy.

Experimental results confirm that the proposed system achieves a Rank-1 identification accuracy of 83.5% and an mAP of 78.2%, outperforming traditional manual sketch approaches and prior GAN-based baselines. The standardized digital output format significantly improves compatibility with modern face recognition pipelines, enabling scalable deployment across large law enforcement databases.

Future research directions include: the development of hybrid enhancement techniques combining CLAHE-based preprocessing with GAN-based synthesis for improved sketch quality under adverse conditions; integration of real-time video sketch generation capabilities for dynamic surveillance scenarios;

expansion of the facial feature library to improve diversity and inclusivity; and deployment of the system within edge computing environments such as autonomous surveillance units and body-worn cameras for field deployment.

#### REFERENCES

1. Kushal Kumar Jain, Steve Grosz, Anoop M. Namboodiri, and Anil K. Jain, "CLIP4Sketch: Enhancing Sketch to Mugshot Matching through Dataset Augmentation using Diffusion Models," arXiv, 2024. <https://arxiv.org/abs/2408.01233>
2. Duoxun Tang, Xin Liu, Kunpeng Wang, Weichen Guo, Jingyuan Zhang, Ye Lin, and Haibo Pu, "Toward Identity Preserving in Face-to-Photo Synthesis with a Hybrid CNN-GAN Framework," 2023.
3. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc. IEEE CVPR, 2017, pp. 1125-1134.
4. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. IEEE ICCV, 2017, pp. 2223-2232.
5. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in Proc. IEEE CVPR, 2015, pp. 815-823.
6. O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in Proc. British Machine Vision Conference, 2015.
7. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Vedaldi, "VGGFace2: A Dataset for Recognising Faces across Pose and Age," in Proc. IEEE FG, 2018, pp. 67-74.
8. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," UMass Amherst Technical Report, 2007.
9. T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in Proc. IEEE CVPR, 2019, pp. 4401-4410.
10. B. T. Dasu, M. V. Reddy, K. V. Kumar, P. Chithaluru, N. Ahmed, and D. S. Abd Elminaam, "A self-attention driven multi-scale object detection framework for adverse weather in smart cities," Scientific Reports, vol. 16, p. 1992, 2026, doi: 10.1038/s41598-025-31660-4.