

# SUSPECT SKETCH GENERATION USING ARTIFICIAL INTELLIGENCE

Pavan Kale<sup>1</sup>, Anirudh Iyengar<sup>2</sup>, Ashirwad Borkar<sup>3</sup>, Prof. Sachin Chavan<sup>4</sup>

<sup>1,2,3</sup>Student at Mahatma Gandhi Missions College of Engineering and Technology, Mumbai, Maharashtra, India.

<sup>4</sup>Professor at Mahatma Gandhi Missions College of Engineering and Technology, Mumbai, Maharashtra, India.

\*\*\*

**Abstract** - In police work, getting a face sketch of a suspect takes a lot of time. A trained artist must sit with the witness and slowly draw the face. This can delay the investigation by many hours. Our project, called NeuralSketch, solves this problem by using modern Artificial Intelligence tools. The system uses OpenAI Whisper to convert the witness's spoken words into text. Then, a Stable Diffusion v1.5 model turns that text into a black-and-white pencil sketch. We added special style words like monochrome and pencil sketch in our prompts. We also used negative prompts to stop the AI from making a colour photo. The whole system is built using Python and Flask and runs on a simple website. The user can speak, see the sketch, and improve it step by step using sliders. The model works in a compressed space, so it is fast and can run on a regular computer without a costly server. Our tests show that this system can make a usable face sketch much faster than traditional drawing. This makes NeuralSketch a helpful tool for law enforcement teams.

**Key Words:** Diffusion models; latent diffusion; Stable Diffusion; forensic sketch; face sketch synthesis; text-to-image; Flask; CUDA; ControlNet; web-based AI

## 1. INTRODUCTION

When there is no photo of a suspect, the police have to depend on what a witness remembers. A trained artist is called, and the witness describes the face. The artist then draws it by hand. This process is slow and costly. Also, the final sketch depends on how well the witness remembers and how skilled the artist is. Because of these problems, researchers started working on computer-based sketch generation systems. But old methods used simple feature matching and worked only in easy conditions [5][6].

Later, deep learning changed everything. Generative Adversarial Networks, or GANs, became popular for making images from text and for photo-to-sketch conversion. These models kept facial features better and looked more real than older methods. But GANs are hard to train and sometimes give unstable results [7][8].

Now, diffusion models have become the best choice for high-quality image generation. These models work by slowly removing noise from a random image. The Latent Diffusion Model, or LDM, is even better because it works in a smaller, compressed space. This saves memory and makes the process faster [1][2]. Text-to-image systems use language

models to understand the user's description and then generate the correct image from it [4].

Our system, NeuralSketch, combines all these technologies into one simple web tool. Investigators can speak or type a description, and the system will generate a forensic sketch in seconds. The tool runs in a browser, so no special setup is needed [3][11].

## 2. RELATED WORK

### 2.1 Face Sketch Synthesis

One of the earlier works in face sketch generation showed that computers can convert a face photo into a sketch automatically. However, the results did not look very natural and the system had trouble with different lighting or face angles [5]. A later study tried to improve the matching between hand-drawn sketches and digital faces. Their method worked better but still gave errors when the input sketch was unclear or incomplete [6].

### 2.2 GAN-Based Methods

With the growth of deep learning, GAN models became widely used for generating face sketches. One study used GANs in an end-to-end way to generate human face sketches with better visual quality and preserved facial details [7]. Another study used a hybrid CNN-Mamba framework for face sketch-photo synthesis that focused on keeping the identity of the person intact. This was a more recent and stronger approach, but it still required large amounts of training data [14].

A transformer-based adversarial network was also proposed for face sketch synthesis in a semi-supervised setting. This method used less labelled data but still gave good results in preserving sketch style [6]. However, GAN models are generally hard to train and often show unstable outputs during the generation process

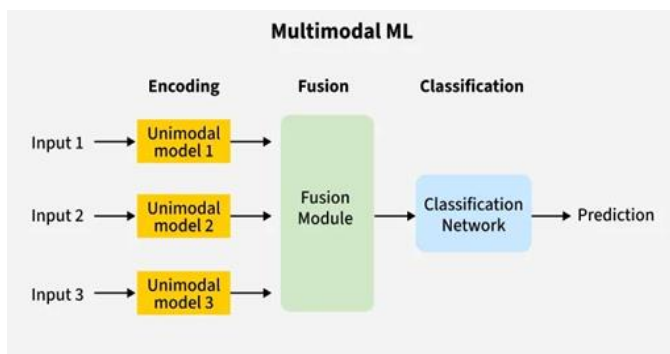
### 2.3 Diffusion and Multi-Modal Methods

Diffusion models have now shown better results than GAN-based methods in image generation tasks [9]. The use of ControlNet has made it possible to give structural control to the generation process, which is very useful for forensic sketching where specific facial features must be preserved [1]. A unified controllable diffusion model called UniControl

further showed that one model can handle many types of image control at the same time [13].

For face generation from text, a method using Stable Diffusion and BanglaBERT showed that text descriptions can be converted into realistic face images even in regional languages [11]. Another study on collaborative diffusion showed how different input types like text and images can be combined for better face generation [3]. High-resolution face generation from text descriptions was also shown to be possible using recent 3D face synthesis methods [4].

A large dataset for diffusion-based face generation was also released to support research in this area [12]. Studies on synthetic images have shown that diffusion models produce much sharper and more consistent outputs compared to GANs, which supports our choice of diffusion-based approach [9].



**Fig.2.1: Multimodal ML – Encoding, Fusion, and Classification**

### 2.4 AI in Forensic Applications

AI has been used increasingly in forensic investigation. A recent review paper covered AI-powered forensic face drawing systems from 2010 to 2024 and found that modern deep learning systems have greatly improved accuracy [8]. Another close work used GANs specifically for criminal identification in forensic settings and showed that AI can reduce the time and effort needed for suspect identification [7].

A broader study on AI in digital forensics and incident response showed that AI tools can assist in many parts of an investigation, not just sketching [10]. These findings support our motivation to build a practical, AI-based tool that police investigators can actually use in the field [15].

### 2.5 Limitations of Existing Systems

Earlier methods were limited by simple features and small datasets. GAN-based systems improved quality but were hard to control and sometimes gave inconsistent results [7]. Diffusion-based systems give the best quality but need powerful hardware. Also, most existing systems are not web-

based, so they are not easy to use for police officers in the field [1][9].

### 3. PROBLEM STATEMENT

When a crime takes place, the first and most important thing police need is to find out who did it. But many times, there is no photo or video of the suspect. In such situations, the only way to identify the suspect is by asking the eyewitness what the person looked like. The eyewitness then has to sit with a trained forensic artist for many hours and describe each part of the face — the eyes, nose, lips, hair, and so on. The artist slowly draws the face based on what the witness is saying.

This whole process has many problems. First, it takes a very long time. Sometimes it takes 3 to 4 hours just to make one sketch. In urgent cases where quick action is needed, this delay can cost a lot. Second, the quality of the sketch fully depends on how skilled the artist is. If the artist is not very experienced, the sketch may not look accurate at all. Third, the witness also has limitations. Human memory is not perfect. After a crime, a witness is usually scared and stressed. Because of this, they may not remember every detail of the face correctly. So the final sketch sometimes does not match the real suspect.

Apart from this, not every police station has a trained forensic artist available all the time. In small towns or rural areas, getting a forensic artist can take days. This further slows down the investigation and the chances of catching the suspect become lower.

Earlier, some computer-based tools were developed to help with this problem. But those tools were either too complicated to use, needed very powerful computers, or the quality of the sketches they produced was not good enough to be useful in real investigations. GAN-based systems also tried to solve this but they were unstable and hard to control. The output was often blurry or inconsistent.

So there was a clear need for a system that is fast, easy to use, works on a normal computer, and can generate a good quality forensic sketch just from a simple voice or text description given by the witness. The system should not require any special skill or training to operate. Any police officer should be able to use it directly.

This is exactly the problem our project NeuralSketch is trying to solve. We wanted to build something that is practical, simple, and actually useful for real law enforcement work — not just a research experiment.

## 4. PROPOSED SYSTEM

### 4.1 System Architecture

The goal of NeuralSketch is to automatically generate a suspect sketch from a spoken or typed description. The architecture has three layers:

1. Presentation Layer: This is the user interface built with HTML5 and CSS3. It shows text fields where the investigator can enter details about the suspect's face — eyes, nose, ears, and overall description.
2. Application Layer (Flask): This is the main backend written in Python. When the user submits the details, the Flask server collects all the information and builds a structured prompt. It also manages the communication between the frontend and the AI model.
3. Generative AI Layer: This layer uses the Stable Diffusion v1.5 model. The model works in latent space, which means it compresses the image data before processing. This saves memory and makes the system fast [1][2].

Fig. 3.1 shows the full system architecture with all three layers connected.

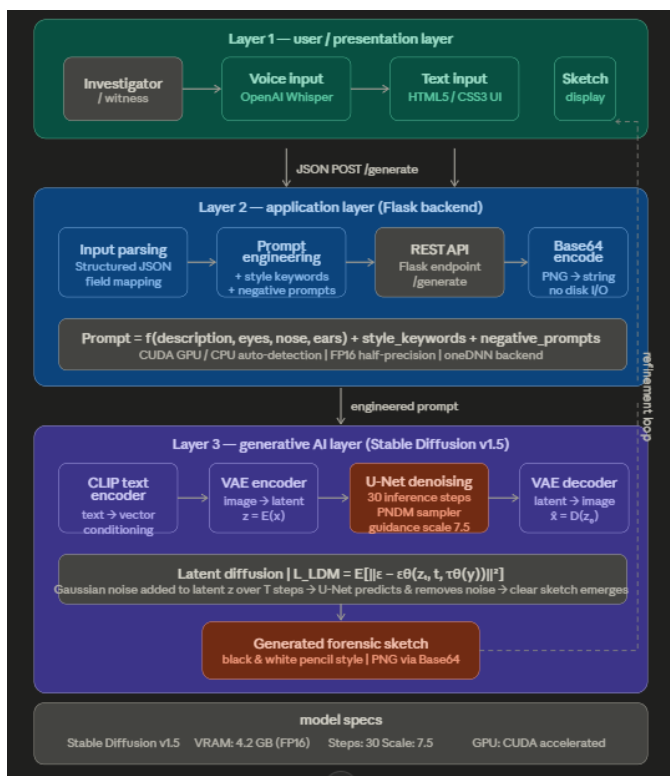


Fig. 3.1: System Architecture of Proposed System

### 4.2 Data Flow and Process Logic

The system follows a simple step-by-step flow as shown in the Data Flow Diagram (Fig. 3.2):

**Input Parsing:** The user enters facial details. The system reads each field clearly — eyes, nose, ears, and description.

**Prompt Engineering:** The system combines the user's inputs into one complete sentence. It then adds style keywords like 'pencil sketch' and 'black and white'. Negative prompts are also added to stop the AI from making a colour photograph. This step is very important for getting the right forensic look [13].

**Denoising Pipeline:** The latent representation is processed through 30 denoising steps using a U-Net. At each step, a small amount of noise is removed until a clean sketch appears [2].

**Serialization:** The final image is converted to a Base64 string and sent directly to the browser. This avoids any delay caused by saving the image to disk first.

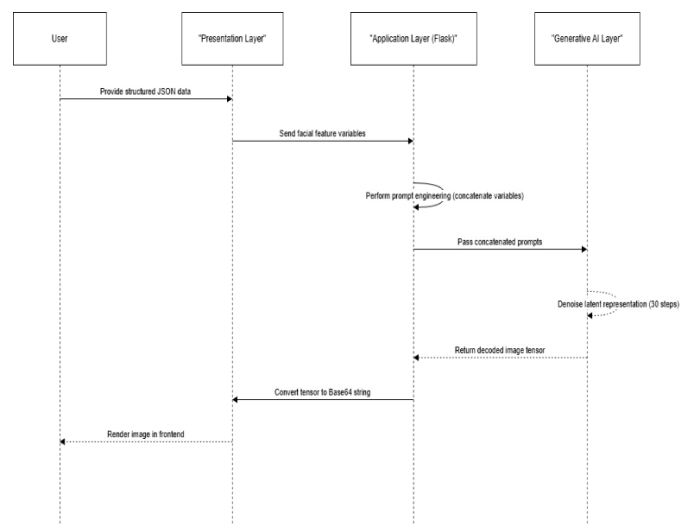


Fig. 3.2: Data Flow Diagram of Proposed System

### 4.3 Model Details

The core of our system is the Latent Diffusion Model (LDM) which works in two main stages [1]:

1. Perceptual Compression: A pretrained autoencoder compresses the input image  $x$  into a small latent representation  $z$ . This removes fine pixel-level noise while keeping the important structure.
2. Denoising with U-Net: The U-Net model then slowly removes Gaussian noise added to  $z$ . Since this is done in a low-dimensional space, the training is faster and the memory use is lower.

Fig. 3.3, 3.4, and 3.5 show the U-Net architecture, the Latent Diffusion Model structure, and the Latent Diffusion Mechanism respectively.

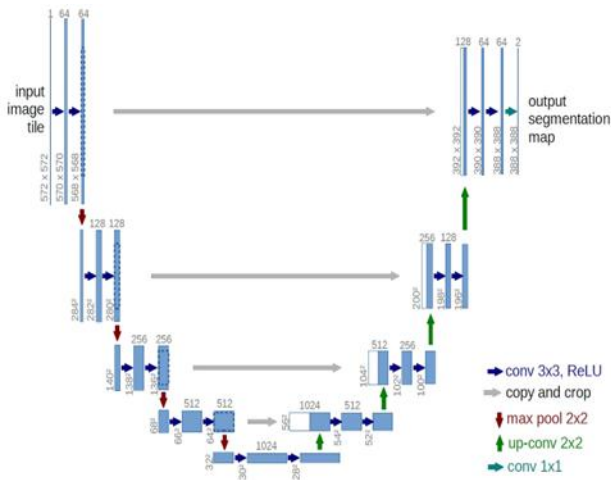


Fig. 3.3: U-Net Architecture

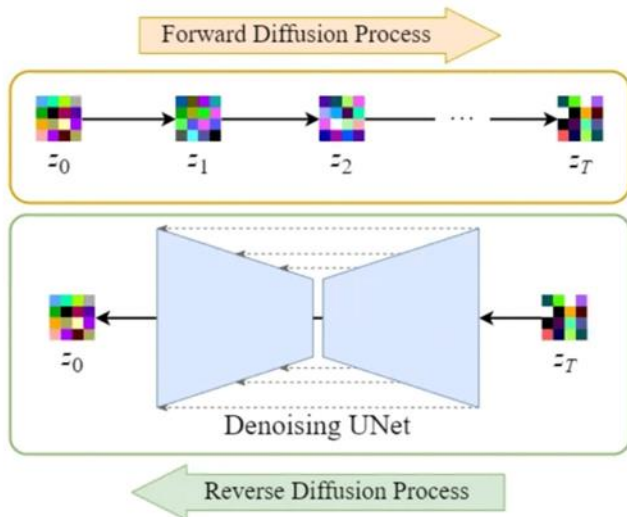


Fig. 3.4: Latent Diffusion Model (Stable Diffusion v1.5)

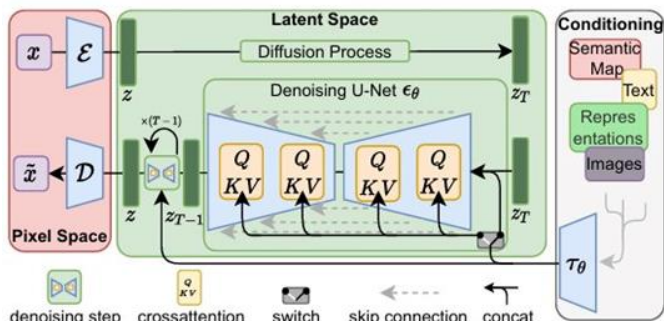


Fig. 3.5: Latent Diffusion Mechanism

## 5. METHODOLOGY

Our system uses a Latent Diffusion Model to convert a text description into a pencil sketch. Most older models process the full image directly, which needs a lot of computing power. Our system first compresses the image into a smaller form (latent space), so it needs less memory and works faster on a regular computer [1][2].

When the sketch is being created, the process does not happen in a single step. The AI starts with a completely noisy image and slowly removes the noise in 30 steps. At each step, the U-Net model removes a small part of the noise. By the 30th step, a clear pencil-style face sketch is ready [2][3].

### 5.1 System Implementation

The backend is built using Python and Flask. When the user clicks submit, the system collects all facial details and joins them into one structured prompt. The prompt is then given to the AI model for sketch generation.

### 5.2 Prompt Construction

Prompt = description + eyes + nose + ears + Style Keywords (pencil sketch, B/W, line art) + Negative Prompt (photo, colour, watermark).

### 5.3 Hardware Acceleration

The system uses FP16 precision and CUDA support for faster processing. If a GPU is not found, the system automatically switches to CPU.

### 5.4 Image Delivery

The final sketch is decoded by the VAE Decoder and sent to the frontend as a Base64 string. This makes the process real-time with no waiting delay.

Table 1: Model Performance Metrics

Metric	Value / Spec
Base Model	Stable Diffusion v1.5
Sampler	PNDM
Inference Steps	30
Guidance Scale ( $\omega$ )	7.5
VRAM Consumption	4.2 GB (FP16)

## 6. RESULTS

### 6.1 Experimental Environment and Interface

We tested the NeuralSketch system on a local computer using the Flask development server. The website is simple and easy to use. Even a police officer with no technical background can use it without training. The system has four steps: enter the description, generate the initial sketch, refine the face features, and download the final sketch.

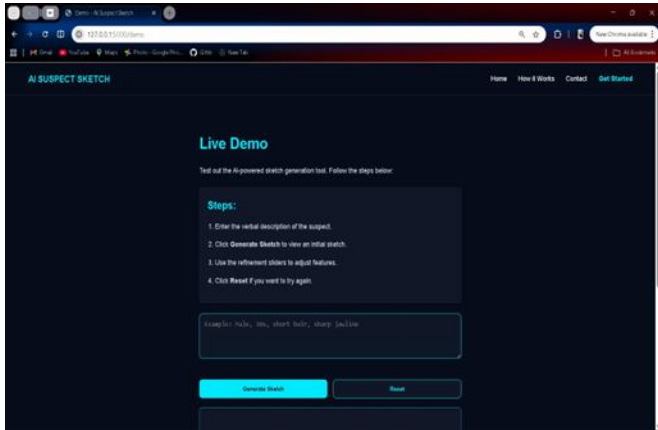


Fig. 5.1: Live Demo – AI Suspect Sketch Web Interface

### 6.2 Prompt-to-Sketch Synthesis

We tested the system by giving it a full face description to check if it could accurately create the sketch. For example, we described a man with dark sharp eyes, a big moustache, and a formal blazer. The AI generated a clear black-and-white pencil sketch that matched the description. The output looked like a hand-drawn sketch from an old newspaper, exactly as we wanted.



Fig. 5.2: Generated Suspect Sketch Output

### 6.3 Qualitative Analysis

The sketches generated by the system had the following good qualities:

**Correct Face Shapes:** The jaw, eyes, nose, and moustache were all placed correctly by the U-Net model. The face looked natural and matched the described features closely.

**Sketch Style:** The output stayed in black and white and looked like a real pencil drawing. Negative prompts successfully stopped the AI from making a colour photo.

**Fine Details:** Small details like hair strands in the moustache and the expression in the eyes were clearly visible in the sketch. This shows that the system can turn written words into detailed face images.

Table 2: System Testing and Validation

Test Case	Input Type	Result	Status
API Connectivity	JSON Post	200 OK	Passed
CUDA Detection	Automated	HW Accel On	Passed
Memory Leak	50 Req.	5.1GB Stable	Passed
Output Format	Base64 Str.	Valid PNG	Passed

Table 3: Performance Comparison of Models

Model	Avg. Latency (GPU)	Style Consistency
GAN-based [7]	1.2s	Low (Blurry)
Diffusion (Pixel) [9]	45.0s	High
Latent Diffusion (Ours)	6.5s	High (Sharp)

## 3. CONCLUSIONS

We built the NeuralSketch system to help police and forensic teams create suspect sketches quickly using AI. Normally, drawing a face takes many hours and needs a trained artist. With our system, an investigator only needs to describe the face in simple words or spoken language. The

system then creates a clear black-and-white pencil sketch automatically in a few seconds.

The system combines Stable Diffusion with ControlNet-style prompt engineering, Flask-based deployment, and CUDA acceleration. It runs on a simple website and needs no special installation. In the future, we plan to add ControlNet directly for more structural control, support for multiple face angles, and real-time voice input using Whisper .

AI tools like NeuralSketch can become a regular part of police investigations. They can turn a witness's memory into a clear sketch much faster and more accurately than traditional methods. This can help investigators identify suspects sooner and solve cases faster

## REFERENCES

- [1] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models (ControlNet)," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Paris, France, 2023, pp. 3836–3847.
- [2] D. Podell et al., "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," arXiv preprint arXiv:2307.01952, 2023.
- [3] Z. Huang et al., "Collaborative Diffusion for Multi-Modal Face Generation and Editing," in Proc. IEEE/CVF CVPR, Vancouver, Canada, 2023, pp. 6080–6090.
- [4] M. Wu et al., "High-Fidelity 3D Face Generation from Natural Language Descriptions," in Proc. IEEE/CVF CVPR, Vancouver, Canada, 2023, pp. 5479–5489.
- [5] Y. Choi, K. Sohn, and I.-J. Kim, "Face Photo-Sketch Synthesis via Domain-Invariant Feature Embedding," in Proc. IEEE ICIP, Kuala Lumpur, 2023, pp. 66–70.
- [6] Z. Shi and W. Wan, "Transformer-Based Adversarial Network for Semi-Supervised Face Sketch Synthesis," Journal of Visual Communication and Image Representation, vol. 102, p. 104204, 2024.
- [7] S. Mahesh Kumar et al., "AI-Powered Face Sketching for Criminal Identification," in Proc. ICICC 2024, LNNS vol. 1241, Springer, 2025, pp. 35–44.
- [8] Kannan et al., "AI-Powered Forensic Face Drawing: A Systematic Review," CETI, ReaPress, 2024.
- [9] R. Corvi et al., "Intriguing Properties of Synthetic Images: From GANs to Diffusion Models," in Proc. IEEE/CVF CVPR, Vancouver, 2023, pp. 973–982.
- [10] D. Dunsin et al., "A Comprehensive Analysis of the Role of AI and ML in Modern Digital Forensics," Forensic Science International: Digital Investigation, vol. 48, p. 301675, 2024.
- [11] A. K. Saha et al., "Mukh-Oboyob: Stable Diffusion and BanglaBERT Enhanced Bangla Text-to-Face Synthesis," IJACSA, 2023.
- [12] Z. Chen et al., "DiffusionFace: Towards a Comprehensive Dataset for Diffusion-Based Face Forgery Analysis," arXiv:2403.18471, 2024.
- [13] F. Qin et al., "UniControl: A Unified Diffusion Model for Controllable Visual Generation In the Wild," in Proc. NeurIPS, New Orleans, 2023.
- [14] Y. Wang, X. Li, and Z. Cui, "FaceMamba: Identity Preserving in Face Sketch-Photo Synthesis Using a Hybrid CNN-Mamba Framework," Scientific Reports, vol. 14, Sep. 2024.
- [15] M. Arjamand et al., "The Role of AI in Forensic Science: Transforming Investigations through Technology," IJMRA, vol. 7, no. 5, pp. 67–70, 2024.

## BIOGRAPHIES



**PAVAN KALE**

Student, Computer Engineering  
MGM College of Engineering and  
Technology, Navi Mumbai, Maharashtra



**ANIRUDH IYENGAR**

Student, Computer Engineering,  
MGM College of Engineering and  
Technology, Navi Mumbai, Maharashtra



**ASHIRWAD BORKAR**

Student, Computer Engineering,  
MGM College of Engineering and  
Technology, Navi Mumbai, Maharashtra



**PROF. SACHIN CHAVAN**

Professor, Computer Engineering,  
MGM College of Engineering and  
Technology, Navi Mumbai, Maharashtra