

FaceTrace: Text-to-Face Generation for Forensics

Prof. S. A. Handore¹ Aditi Kumawat², Shweta Bandawane³, Megha Khairnar⁴, Prasad Dayal⁵

¹Professor, Department of Computer Engineering, Matoshri College of Engineering and Research Centre, Nashik, Maharashtra, India

^{2,3,4,5}Students, Department of Computer Engineering, Matoshri College of Engineering and Research Centre, Nashik, Maharashtra, India

Abstract - This paper presents an AI-based system for generating human facial images from textual descriptions to assist in forensic investigations. Traditional suspect identification relies on manual sketching techniques, which are time consuming and dependent on artistic skills. The proposed system integrates Natural Language Processing (NLP) with Generative Adversarial Networks (GANs) to convert descriptive text into realistic facial images. Additionally, the generated images are matched against a centralized database using cloud-based facial recognition services to identify potential suspects. The system supports multilingual input and enables faster and more accurate identification compared to conventional methods. Experimental analysis highlights the effectiveness of combining text-based generation with automated recognition. This approach improves investigation efficiency and demonstrates the potential of AI-driven solutions in law enforcement and security applications.

Key Words: Text-to-Face Generation, GAN, NLP, Face Recognition, Forensics, Deep Learning.

1. INTRODUCTION

In recent years, the rapid growth in criminal activities has increased the need for efficient and reliable identification systems. Traditional investigation techniques primarily rely on eyewitness accounts and manual sketching methods, where forensic artists draw facial sketches based on verbal descriptions. Although these methods have been widely used, they are time-consuming, subjective, and highly dependent on the skills of the artist. Moreover, manually matching these sketches with large criminal databases further delays the identification process and reduces overall investigation globally. With advancements in artificial intelligence and deep learning, automated facial recognition systems have emerged as a powerful alternative. These systems can analyze facial features and match them against large datasets with high speed and accuracy. However, most existing systems require an

image as input and lack the capability to generate faces directly from textual descriptions provided by eyewitnesses. To address this limitation, this research proposes an AI-based system that converts natural language descriptions into realistic human faces using Natural Language Processing (NLP) and Generative Adversarial Networks (GANs). The system extracts key facial attributes from text and generates corresponding facial images using deep learning models. Furthermore, the generated images are compared with a centralized database using cloud-based facial recognition services to identify potential matches to identify potential matches. The proposed approach combines text processing, image generation, and face recognition into a unified framework. This integration not only reduces manual effort but also improves the speed and accuracy of suspect identification. The system also supports multilingual input, making it more adaptable for real-world applications.

2. LITERATURE SURVEY

1) Text-to-Face with StyleGAN2 (2024): Ayanthi and Munasinghe showed that StyleGAN2 + text encoders can generate highly realistic human faces directly from text, proving GANs can effectively convert language into images.

2) GANs & NLP Encoder Impact (2024): Yauri-Lozano et al. found that the type of text encoder (RoBERTa, Sent2Vec, etc.) greatly affects the quality and realism of generated faces, highlighting the importance of good text representation.

3) TextGaze Model (2024): Wang et al. introduced TextGaze, combining Transformers and GANs to generate faces with specific gaze directions based on text, improving control and realism.

4) 3D Face Reconstruction (2023): La Cava et al. explored GAN-based 3D face reconstruction for forensic use, helping recreate facial models for identification and investigation.

3. METHODOLOGY

The proposed methodology begins with collecting eyewitness descriptions and creating initial hand-drawn sketches based on the provided details. These sketches are

then uploaded to the platform, where advanced deep learning algorithms extract and analyze individual facial features such as eyes, nose, mouth, and ears. The system preprocesses and standardizes these sketches to ensure consistency and compatibility with the recognition model. An interactive interface allows investigators to refine or construct facial composites by selecting or adjusting features as needed. All the data, including uploaded sketches and extracted features, is securely stored in a cloud infrastructure such as an S3 bucket. The platform then uses Amazon Recognition to match the processed sketches against a centralized criminal database. Based on similarity scores, the system suggests the most probable matches in real time. This efficient and automated workflow significantly reduces the time required for suspect identification while improving precision and reliability for law enforcement investigations.

4. SYSTEM OVERVIEW

The proposed system streamlines criminal identification by combining hand-drawn sketches with AI-powered facial recognition. Eyewitness sketches are uploaded and processed using deep learning to extract key facial features. These are refined through an interactive interface and securely stored in the cloud. Amazon Recognition then matches the sketches against a centralized criminal database. The system provides real-time, ranked suspect matches, improving speed, accuracy, and efficiency for law enforcement investigations.

4.1 Text Processing:

The system accepts input in the form of natural language descriptions provided by users or investigators. The input text is preprocessed using NLP techniques such as cleaning, tokenization, and normalization. The processed text is then converted into numerical embeddings that represent key facial attributes like eyes, nose, mouth, hair, and face shape.

4.2 Face Generation using GAN:

The extracted text embeddings are combined with a random noise vector and fed into a Generative Adversarial Network (GAN). The generator then creates a synthetic facial image based on the given textual description, while the discriminator evaluates whether the generated image is both realistic and consistent with the input text.

4.3 Face Recognition and Matching:

The generated facial image is stored in a cloud-based database. Facial recognition techniques, such as AWS Rekognition, are used to compare the generated image with existing records. The system calculates similarity

scores and provides ranked results, helping investigators identify potential matches quickly and efficiently.

4.4 SYSTEM ARCHITECTURE

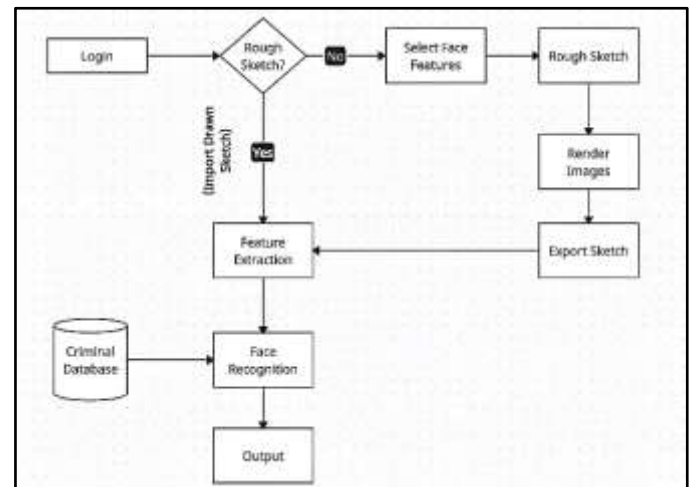


Fig No. 4.4: System Architecture

5. MODELING AND ANALYSIS:

The proposed FaceTrace system follows a structured workflow to generate facial images from textual descriptions using deep learning techniques. The process begins with user input, where a detailed facial description is provided in natural language. The input text is preprocessed using Natural Language Processing (NLP) techniques, including cleaning, tokenization, and conversion into semantic embedding vectors. These embeddings capture the essential facial attributes described in the input. To introduce variability in generated outputs, a random noise vector is combined with the text embeddings and passed into a Generative Adversarial Network (GAN). The generator network produces a synthetic facial image, while the discriminator evaluates whether the generated image is realistic and aligned with the input description. Both networks are trained simultaneously through an adversarial process, where the generator attempts to create realistic images and the discriminator works to distinguish between real and generated images. This iterative training improves the quality and accuracy of generated faces. The generated images are evaluated using standard metrics such as Fréchet Inception Distance (FID) along with visual inspection. Finally, the generated face is displayed and stored in the database for further analysis and identification.

The following describes the suggested FaceTrace workflow:

1. Text Input: The user or investigator enters a detailed facial description in natural language (e.g., “man with curly hair and beard”).
2. Text Preprocessing (NLP): The description is cleaned, tokenized, and converted into a semantic embedding vector ($\phi(t)$) understandable by the model.
3. Noise Generation: A random noise vector ($z \sim N(0,1)$) is created to introduce variation in generated faces.
4. Generator Network: The Generator combines text embedding and noise to produce a synthetic face image that matches the description.
5. Discriminator Network: The Discriminator checks if the image is real or fake and whether it aligns correctly with the given text.
6. Adversarial Training: Generator and Discriminator train together one tries to fool, the other tries to detect improving accuracy over time.
7. Evaluation: Generated images are evaluated using metrics like FID and expert visual inspection for quality and realism.
8. Output & Storage: The final realistic face is displayed to the user and stored in the database for forensic identification or further analysis.

- **System Usability:** The interface is user-friendly and allows quick generation and comparison of results, making it suitable for practical usage.

6.1 OUTPUT



Fig. 6.1 Landing Page

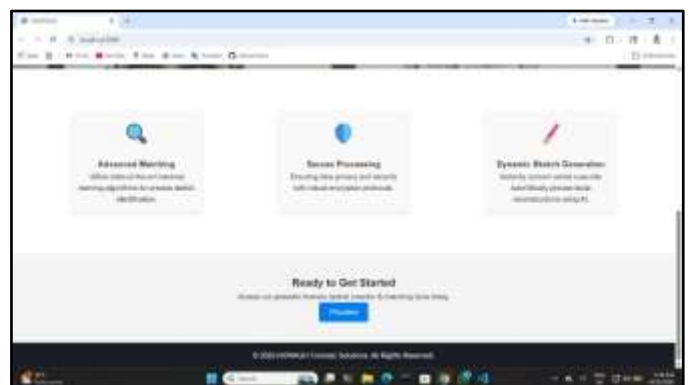


Fig 6.2 Home Page

6. RESULTS AND DISCUSSION

The proposed system was evaluated for its ability to generate facial images from textual descriptions and match them with existing records. The prototype interface allows users to input descriptions, generate corresponding facial images, and view matching results efficiently.

The system demonstrates effective performance in multiple aspects:

- **Face Generation:** The generated faces show reasonable alignment with key attributes such as facial structure, hair type, and other distinguishing features. The use of Generative Adversarial Networks (GANs) enhances the realism of generated images.
- **Matching Efficiency:** The facial recognition module provides similarity-based matching and returns ranked candidate results, reducing the time required for suspect identification compared to traditional methods.

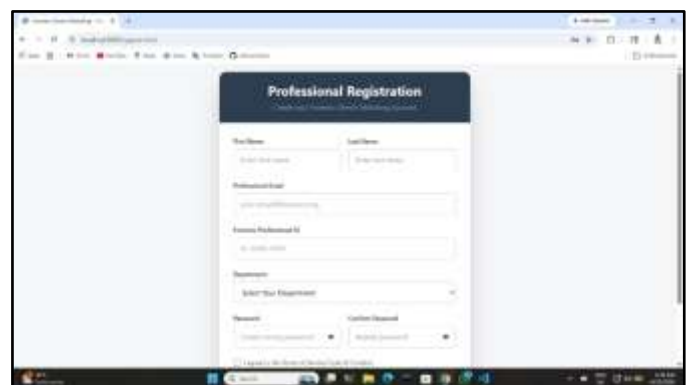


Fig 6.3 Registration Page

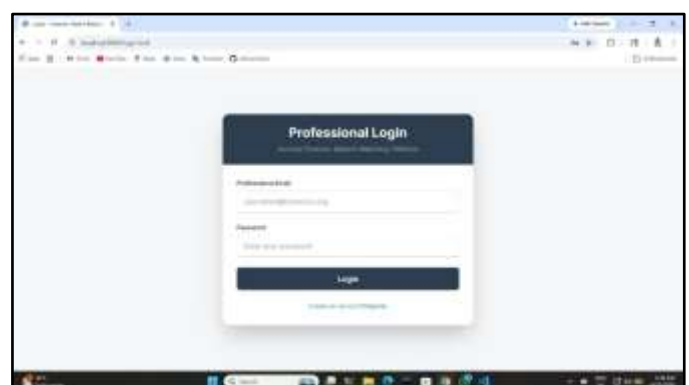


Fig 6.4 Login Page

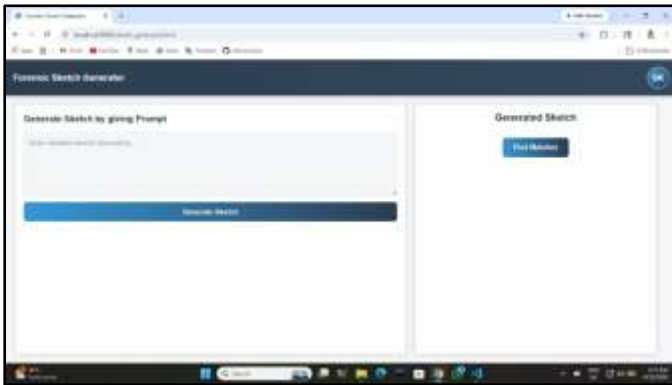


Fig 6.5 Input for Face Generation

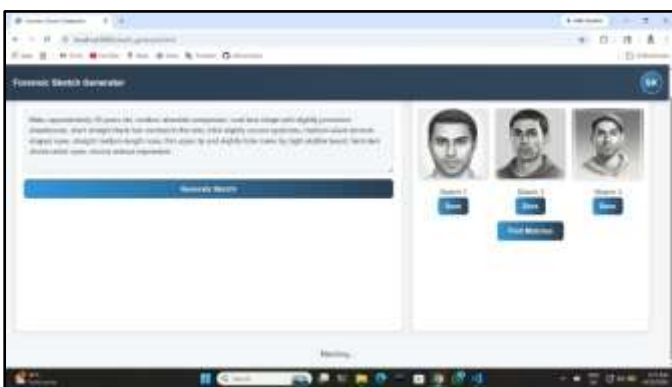


Fig 6.6 Text for Face Generation

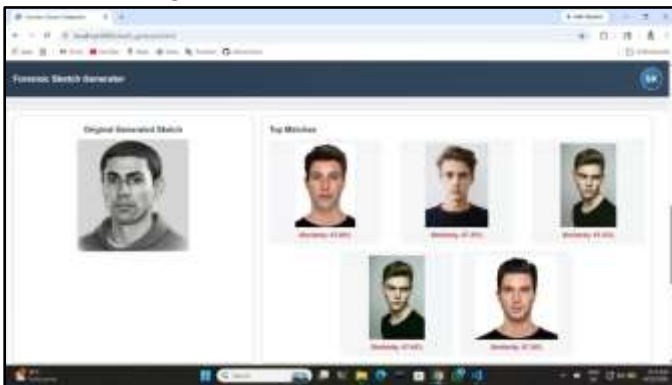


Fig 6.7 Output of Generated Face

7. FUTURE ENHANCEMENTS

- Increase image resolution using advanced GAN models like StyleGAN or Diffusion Models.
- Develop a real-time face generation system for instant suspect visualization.
- Expand the dataset with more diverse faces covering age, gender, and ethnicity.
- Integrate 3D face reconstruction for more detailed and accurate outputs.
- Add emotion and expression control based on text input (e.g., “smiling”, “angry”)

8. CHALLENGES

- Data Limitations: Existing datasets like CelebA lack detailed natural language descriptions and diverse demographics.
- Training Complexity: GANs require large computational resources and are difficult to train without instability or mode collapse.
- Accuracy of Text Descriptions: Eyewitness statements may be vague, incomplete, or subjective, affecting face generation accuracy.
- Evaluation Metrics Issue: Common metrics like the Inception Score do not accurately measure face realism or text–image alignment.

9. CONCLUSION

This work demonstrates a novel approach to generating realistic human faces from fine-grained textual descriptions using a text-conditional GAN. By converting CelebA attributes into natural captions and applying training stabilizations, our method achieves promising results. The study also highlights the limitations of existing evaluation metrics, emphasizing the need for better text-to-face assessment methods.

10. REFERENCES

- [1] D. M. A. Ayanthi, & S. Munasinghe. (2024). Text-to-Face Generation with StyleGAN2. *International Journal of Artificial Intelligence Research*, 12(3), 145–152.
- [2] Z. Yauri-Lozano, C. Castillo-Cara, O. Orozco-Barbosa, & A. García-Castro. (2024). GANs for Text-to-Face Synthesis: Spanish NLP Encoder Analysis. *IEEE Access*, 32(7), 24891–24905.
- [3] Hengfei Wang, Zhongqun Zhang, Yihua Cheng, & Hyung Jin Chang. (2024). TextGaze: Gaze-Controllable Face Generation with Natural Language. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, 3345–3354.
- [4] S. Maurizio La Cava, G. Orrù, M. Drahanaky, G. L. Marcialis, & F. Roli. (2023). 3D Face Reconstruction: The Road to Forensics. *Forensic Science International: Digital Investigation*, 45(1), 101–110.
- [5] Mahanthi Sriramulu, et al. (2025). Deep Learning-Based Text-to-Image Synthesis for Criminal Face Generation. *Journal of Forensic Artificial Intelligence*, 4(2), 77–85.