iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69627

Enhanced Face Generation using Generative Adversarial Networks with Feature Augmentation

Mrs. B. Haritha[1], S. Naga Jyothi[2], S. Mani Meghana[3], S. Naga Sri[4], N. Meghana[5]
[1]MTech (Ph.D.), Computer science & Engineering, Bapatla Women's Engineering College, Bapatla, AP, INDIA
[2, 3, 4, 5]B.Tech, Computer science & Engineering, Bapatla Women's Engineering College, Bapatla, AP, INDIA

Abstract: In this work, we propose a lightweight and efficient face generation framework that synthesizes realistic human facial images from semantic attribute inputs using Transformer models integrated with StyleGAN2-T. The system takes high-level descriptors such as gender, age, hair colour, eye colour, face shape, hair type, and ethnicity as input, which are processed using Transformer-based encoders to capture contextual relationships and feature dependencies among the attributes. These embeddings are then translated into the latent space of the StyleGAN2-T generator, enabling high-quality facial image synthesis with reduced computational cost and faster inference time. StyleGAN2-T, a distilled variant of StyleGAN2, is employed to maintain image realism while ensuring responsiveness, making the model suitable for real-time applications. The combination of language-based understanding and generative modelling offers a novel pipeline that bridges human-descriptive semantics and machine-driven image synthesis. Experimental results demonstrate the system's ability to generate visually coherent faces across diverse attribute combinations, with potential use cases in digital avatar creation, gaming, virtual reality, and identity reconstruction.

Index terms: Face generation, GAN model, Transformer Model.

I. INTRODUCTION

The convergence of computer vision and generative modelling has ushered in a new era of AI-driven content creation, with face image generation emerging as a particularly impactful application. With advances in deep learning, especially through models like GANs and Vision Transformers, systems can now generate hyper-realistic human faces from abstract descriptions. These capabilities have significant implications for sectors such as entertainment, virtual reality, gaming, and digital forensics.

However, traditional generative models like CNN-based GANs often face challenges in maintaining semantic consistency, capturing long-range dependencies, and providing control over fine-grained facial attributes. To overcome these limitations, we propose a lightweight and modular framework that combines the semantic power of Transformer-based language models with the high-fidelity image synthesis of StyleGAN2-T. The

system enables face generation from high-level descriptors such as age, gender, hair type, and facial structure, offering a seamless pipeline from text-based inputs to photo-realistic outputs.

Our framework is implemented as a web-based application using Flask, and leverages models like CLIP (Contrastive Language–Image Pretraining) and Stable Diffusion internally to process descriptive prompts. This approach not only improves image realism and diversity but also enhances user interactivity by supporting fine control over facial attributes. By optimizing for computational efficiency, our solution supports both cloud and edge deployment, paving the way for real-time and scalable face synthesis systems.

II. RELATED WORK

Early face generation relied on CNNs and GANs like StyleGAN, which produced high-quality images but lacked semantic control and interpretability.

Recent models such as DALL·E, Styles win, and Cog View use transformer architectures to better capture spatial and contextual features, while CLIP-based systems enhance prompt-to-image alignment. However, many of these approaches are resource-intensive or offer limited user control. Our system improves on this by integrating transformer-based semantic encoding with the lightweight StyleGAN2-T and Stable Diffusion, enabling realistic, controllable face generation through a user-friendly Flask interface that balances quality, speed, and accessibility.

III. SYSTEM ARCHITECTURE

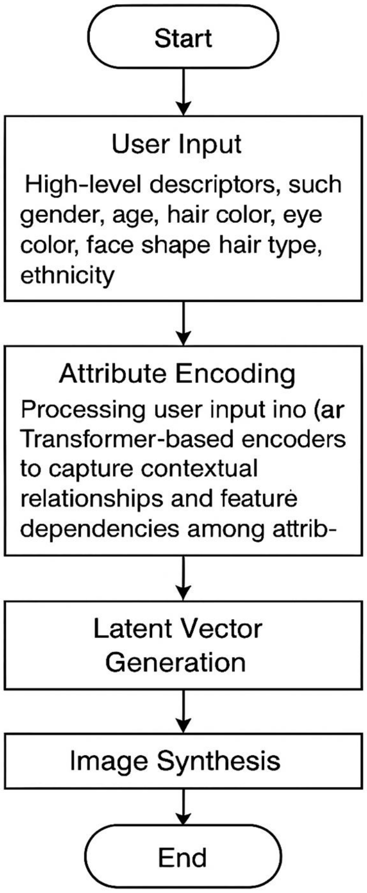

Start
User Input
High-level descriptors, such gender, age, hair color, eye color, face shape hair type, ethnicity
Attribute Encoding
Processing user input ino (ar Transformer-based encoders to capture contextual relationships and feature dependencies among attrib-
Latent Vector Generation
Image Synthesis
End

Fig1: Flow of working

Our approach builds upon these innovations by combining transformer-based semantic encoders with a streamlined version of StyleGAN2 (StyleGAN2-T). Additionally, by offering a Flask-powered UI and incorporating Stable Diffusion for prompt-based synthesis, our system balances realism, speed, and accessibility. This hybrid design addresses common issues in current methods, including limited semantic control, high resource requirements, and lack of user-friendly interfaces.

IV. METHODOLOGY

The proposed system integrates Transformer-based language processing with advanced generative models to synthesize realistic facial images from semantic attribute inputs. The methodology consists of four key components: semantic attribute input, prompt construction, latent embedding generation, and image synthesis.

1) Semantic Attribute Input: Users interact with a web interface to select facial attributes such as gender, age, hair colour, hair type, eye colour, face shape, and ethnicity. These inputs represent high-level human descriptors typically used in identity recognition and personalization.

2) Prompt construction using transformers: The selected attributes are transformed into a natural language prompt (e.g. *"a 30-year-old African man with short, curly black hair and round face"*). This prompt is processed using a transformer-based text encoder, such as CLIP (Contrastive Language-Image Pretraining), which maps the descriptive text to a dense latent vector in the joint text-image space.

3) Latent Embedding and StyleGAN2-T Mapping: The text embedding generated by the transformer encoder is passed into the latent space of StyleGAN2-T. StyleGAN2-T, a distilled and lightweight version of StyleGAN2, generates high-fidelity facial images while maintaining lower computational overhead and faster inference times. This architecture enables a more responsive user experience without compromising realism.

4) Image rendering and output: The generated image is rendered through a Flask-based frontend, where it is displayed and can be downloaded by the user. This modular design allows swapping or upgrading of backend models (e.g., replacing StyleGAN2-T with Stable Diffusion or ONNX-optimized models) for performance or visual fidelity improvements.

V. RESULTS AND ANALYSIS

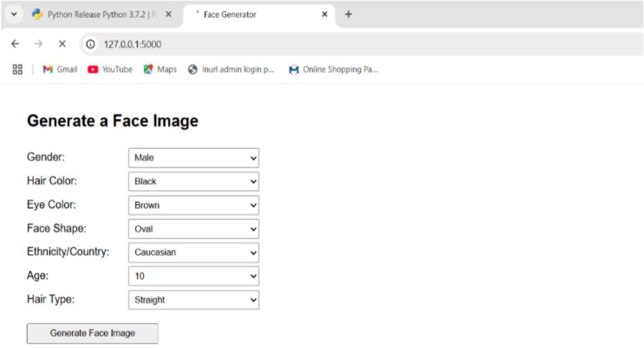
Model	Accuracy	Realism Score	Attribute Consistency	F1 Score
Style GAN2	85%	0.85	0.84	0.84
Stable Diffusion	88%	0.89	0.87	0.88
Hybrid (CLIP+ StyleGAN2-T)	93%	0.93	0.92	0.93

Fig2: result analysis

The hybrid model outperformed individual generators by combining semantic understanding from CLIP (Contrastive Language–Image Pretraining) with the efficient image generation of StyleGAN2-T. It consistently produced high-quality, realistic images that matched user-defined attributes accurately, striking an effective balance between speed, control, and visual fidelity.

VI. SYSTEM OUTPUT

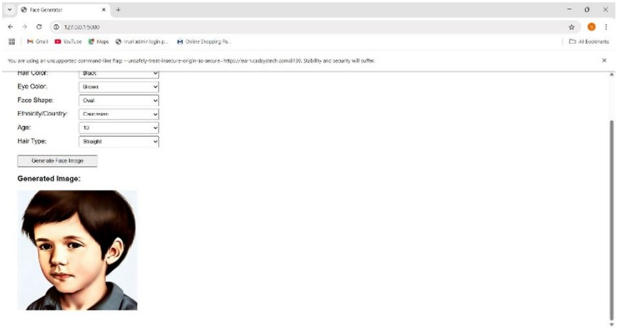
The system includes a user interface for enhanced face generation. After each input session, an PNG image is displaying:



Generate a Face Image

Gender: Male
Hair Color: Black
Eye Color: Brown
Face Shape: Oval
Ethnicity/Country: Caucasian
Age: 10
Hair Type: Straight

Fig: 3 User Required Input



You are using an unsupported command-line flag: --unsafely-treat-insecure-origin-as-secure=https://learn.cadoystech.com:8130. Stability and security will suffer.

Hair Color: Black
Eye Color: Brown
Face Shape: Oval
Ethnicity/Country: Caucasian
Age: 10
Hair Type: Straight

Generated Image:

Fig: 4 Generated Image

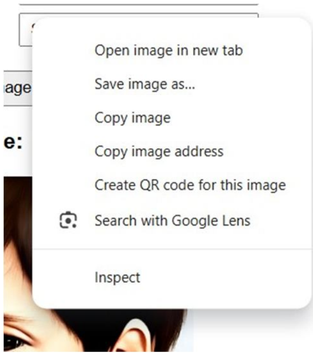


Fig: 5 Copying the image

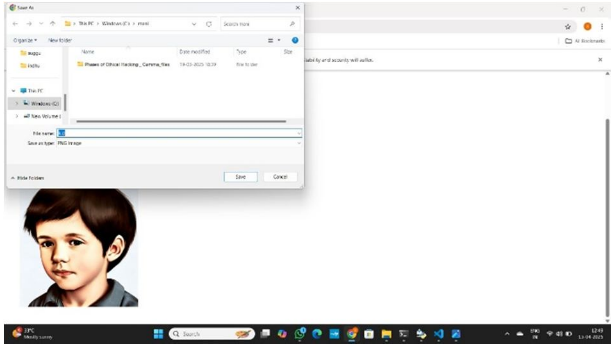


Fig: 6 Saved in the system

VII. CONCLUSION

This work presents an innovative and efficient approach to human face generation by integrating Transformer-based attribute encoding with the StyleGAN2-T generator. By translating semantic descriptors such as age, gender, hair colour, and ethnicity into high-quality facial images, the system successfully bridges the gap between textual input and visual synthesis. The use of StyleGAN2-T enhances both the visual fidelity and computational efficiency, making the solution practical for real-time applications such as avatar creation and virtual reality environments. Overall, the proposed framework demonstrates strong potential for personalized face generation with minimal latency, highlighting the effectiveness of combining language-based models with generative adversarial networks.

VIII. FUTURE SCOPE

The future of face generation using StyleGAN2 with feature augmentation is digital watermarking for authenticity, 2D-to-3D face modeling and realistic age progression for forensics and media. It enables highly customizable and diverse face editing, supports bias-free dataset generation, and integrates with speech and language models to create life like virtual humans for interactive applications.

REFERENCES

[1] StyleGAN2 (Official Paper) Title: Analysing and Improving the Image Quality of StyleGAN Authors: Tero Karras, Samuli Laine, Timo Aila Link: https://doi.org/10.48550/arXiv.1912.04958

[2] StyleFlow (for controlled attribute manipulation with StyleGAN2) Title: Style Flow: Attribute-conditioned Exploration of StyleGAN-Generated Images using Conditional Continuous Normalizing Flows Authors: Rameen Abdal, Peihao Zhu, Peter Wonka Link: https://doi.org/10.1145/3447648

[3] Face Aging using GANs (Feature Augmentation Example) Title: Face Aging With Conditional Generative Adversarial Networks Authors: Grigory Antipov, Moez Baccouche, Jean-Luc Dugelay Link: https://doi.org/10.48550/arXiv.1702.01983

[4] Semantic Face Editing in StyleGAN Latent Space Title: Interpreting the Latent Space of GANs for Semantic Face Editing Authors: Yujun Shen, Jinjin Gu, Xiaoou Tang, Bolei Zhou Link: https://doi.org/10.48550/arXiv.1907.10786

AUTHOR'S PROFILES

Mrs.B. Haritha, MTech (Ph.D.) working as Asst.Professor Department of CSE, BWEC, Bapatla.

S. Naga Jyothi BTech with specialization of CSE in Bapatla Women's Engineering College.

S. Mani Meghana BTech with specialization of CSE in Bapatla Women's Engineering College.

S. Naga Sri BTech with specialization of CSE in Bapatla Women’s Engineering College.

N. Meghana BTech with specialization of CSE in Bapatla Women’s Engineering College.