Controllable Facial Image Generation Using ADGAN++ and Sequential Attribute Decomposition

KODAPALA SWAPNA PRIYA[1], RAMINENIRAJA[2], NELLURI SRINU[3], MADDIGUNTLA TEJA SAI MANIKANTA[4], MUDAVATH SIVA BRAHMAM NAIK[5], Dr. K VENKATARAMARAO[6]

Electronics & communication engineering, Chalapathi Institute of Engineering & Technology, LAM, Guntur-Andhra Pradesh[1,2,3,4,5]

[6]Associate Professor *Electronics & communication engineering, Chalapathi Institute of Engineering & Technology, LAM, Guntur*

Abstract— **The field of generative image modeling has witnessed remarkable growth through the advancement of Generative Adversarial Networks (GANs), particularly in the domain of facial image synthesis. However, despite the high-quality outputs of existing models, many struggle to offer fine-grained, user-controllable manipulation over facial attributes without sacrificing realism or structural consistency. To address these challenges, this paper introduces ADGAN++, an improved version of the Attribute-Decomposed GAN (ADGAN) framework, designed to support more precise and modular control over facial image synthesis.Unlike prior models that often encode all attributes in parallel—leading to feature entanglement and overlapping effects—ADGAN++ employs a serial attribute encoding strategy, wherein each facial component is represented as a separate latent vector and injected sequentially during the image generation process. This approach enhances the model's ability to isolate and edit individual features such as hair, eyes, lips, and facial expressions, without distorting unrelated regions. Additionally, the model supports multi-style generation and pose transfer, enabling the fusion of attributes from multiple source images to create visually compelling and semantically consistent outputs.The system architecture comprises a serial encoder, an attribute-aware decoder, and a conditional discriminator, trained jointly on the CelebMask-HQ dataset. Through this design, ADGAN++ achieves high structural fidelity and perceptual quality, as demonstrated by quantitative evaluations using SSIM, FID, PSNR, and Attribute Control Scores. The model shows significant improvements over previous methods like StarGAN, AttGAN, and the original ADGAN, particularly in terms of realism, identity preservation, and attribute accuracy. An ablation study further confirms the effectiveness of serial encoding in improving both disentanglement and control.Practical results include a user-friendly interface for uploading datasets, generating and editing facial images, and visualizing synthesis outcomes in real time. Use cases range from creative design and character generation to virtual avatars and personalized content creation, where interpretability, realism, and control are essential.**

Keywords— Generative Adversarial Networks (GANs), Controllable Image Synthesis, Attribute Decomposition, ADGAN++, Latent Space Manipulation, Facial Image Generation, Pose Transfer, Multi-Style Generation, Structural Similarity Index (SSIM), Deep Learning.

I. INTRODUCTION

Recent advancements in deep generative models, particularly Generative Adversarial Networks (GANs), have significantly transformed the landscape of image synthesis. These models have enabled the creation of high-quality, photorealistic images across various domains, with facial image generation being one of the most prominent applications. While models like StyleGAN and StarGAN have demonstrated impressive capabilities in generating and translating facial features, a persistent challenge remains—the ability to precisely control and manipulate individual facial attributes without affecting the integrity or identity of the image.Traditional GAN architectures often rely on globally entangled latent spaces, where modifying one feature can inadvertently alter others. This limits their effectiveness in tasks such as personalized avatar creation, facial editing, and virtual try-on systems, where selective and interpretable attribute control is crucial. Moreover, simultaneous editing of multiple attributes frequently results in quality degradation or feature inconsistency, posing significant limitations for real-world usability.

To address these challenges, attribute-conditioned GANs were proposed, with models like ADGAN introducing attribute decomposition, where the latent space is separated into feature-specific components. This provided a new degree of control over face editing. However, ADGAN's parallel encoding approach still struggles with feature interference, especially when multiple regions are edited simultaneously or when blending features from different identities.In this context, we propose ADGAN++, a novel extension of the original ADGAN framework that introduces a serial attribute encoding strategy. Instead of encoding all facial attributes at once, ADGAN++ processes each attribute sequentially, injecting them one by one into the generation pipeline. This not only reduces feature entanglement but also allows for precise, interpretable control over facial modifications. The model supports advanced capabilities such as attribute-wise editing, pose transfer, and multi-style synthesis, offering enhanced realism and flexibility.

Trained on the CelebMask-HQ dataset, ADGAN++ is capable of reconstructing high-resolution facial images while preserving identity and background structure. A conditional discriminator ensures adversarial realism, while loss functions for reconstruction, attribute consistency, and structural similarity guide the model toward producing visually consistent and semantically accurate outputs.This work aims to demonstrate the strength of ADGAN++ in delivering controlled, high-fidelity facial synthesis and highlights its applications in fields such as digital entertainment, virtual reality, personalized media, and human-computer interaction. Through both qualitative and quantitative evaluations, including SSIM, FID, and PSNR, we establish ADGAN++ as a robust and interpretable tool for next-generation facial image synthesis.

In recent years, the field of computer vision has witnessed remarkable advancements through the adoption of deep learning techniques, particularly Generative Adversarial Networks (GANs). GANs have emerged as a powerful tool

Doi:10.48047/ijiee.2025.15.3.10

for image generation tasks, enabling the creation of high-resolution, photorealistic images across various domains such as art, fashion, medical imaging, and facial synthesis. A key limitation, however, lies in their ability to offer precise control over the generated content, especially when it comes to modifying specific attributes of an image without compromising its overall identity or realism.

To address this challenge, researchers have proposed numerous conditional and attribute-based GAN models. While these models have made strides in guided synthesis, many still suffer from entangled attribute representations and limited flexibility in manipulating multiple features simultaneously. Traditional GANs often encode attributes in a global or shared latent space, making it difficult to isolate and modify individual traits, such as facial expression, hairstyle, or background.

In this context, Attribute-Decomposed Generative Adversarial Networks (ADGAN) introduced a significant improvement by decomposing image features into attribute-specific vectors. This separation allows for independent control of visual characteristics, making the image generation process more interpretable and customizable. Building on this concept, the proposed work enhances ADGAN into a more powerful framework, referred to as ADGAN++, which introduces serial attribute encoding and a refined reconstruction process to synthesize images in a modular and progressive manner.

ADGAN++ employs a serial encoder structure that extracts and injects attribute vectors in a sequential manner. This modular approach enables precise manipulation of specific regions or components of an image, such as hair, eyes, or facial expressions, while maintaining consistency with the source image. The model is trained on the CelebMask-HQ dataset, which offers fine-grained facial segmentation, allowing the system to perform editing, pose transfer, and multi-style synthesis with improved fidelity.

The motivation behind this research is to enable fine control over image synthesis while preserving realism and identity. The ability to generate or edit images based on decomposed attribute vectors opens new possibilities in applications such as virtual avatars, facial editing tools, content creation, and personalized image generation. Moreover, the introduction of evaluation metrics like Structural Similarity Index (SSIM) ensures that the generated outputs are not only visually compelling but also structurally aligned with the originals.

This paper aims to demonstrate the effectiveness of the proposed ADGAN++ model in delivering controlled image synthesis. The system's design, training methodology, and evaluation are discussed in detail, followed by results showcasing attribute-level control, reconstruction quality, and practical applications. The promising outcomes highlight the potential of ADGAN++ as a next-generation tool for controllable and realistic image generation.

II. RELATED WORKS

Image synthesis and editing using deep generative models have become an active research area, with significant progress made through Generative Adversarial Networks (GANs). Early works such as DCGAN and Pix2Pix demonstrated the ability to generate visually realistic images from random noise or paired image data. However, these models provided limited control over specific image attributes. To address this, conditional variants like cGAN and InfoGAN were proposed, where auxiliary information or latent codes are used to control image generation. While effective in certain scenarios, these models often lacked the granularity required for fine-tuned, localized editing.

Subsequent frameworks like StarGAN and AttGAN expanded multi-domain image-to-image translation by incorporating attribute vectors. StarGAN, for instance, allows the modification of multiple attributes across domains using a unified model. However, the lack of proper disentanglement between attributes often results in unwanted changes to image regions that were not intended to be modified. Models such as Fader Networks attempted to separate attribute representations using adversarial training, yet they struggled with preserving identity and background consistency.

To achieve better attribute manipulation and identity retention, researchers proposed models like GANimation and MaskGAN, which use attention mechanisms and segmentation masks, respectively. These models offer improved control but at the cost of increased complexity and reliance on accurately annotated data. More recently, ADGAN introduced a significant shift by decomposing the latent space into independent attribute vectors. This decomposition allows for modular editing of images by isolating features such as hair, eyes, and skin tone into separate latent components. Despite its promising structure, ADGAN's parallel encoding strategy occasionally results in feature overlap, and its ability to blend styles from multiple sources remains limited.

Furthermore, techniques such as StyleGAN and StyleGAN2 have shown remarkable fidelity in face generation, yet they primarily focus on quality rather than controllability. In comparison, the need for a synthesis model that combines both high-quality output and user-controllable features remains a significant research gap. Addressing this, our work builds upon the principles of ADGAN while proposing a serial attribute injection mechanism, enabling enhanced control, improved realism, and support for complex transformations like pose variation and attribute fusion.

2.1 Existing System

The original ADGAN framework enables controllable image synthesis by decomposing an image into attribute-specific latent representations. These representations can be manipulated individually and reassembled to generate edited images. The model allows basic editing and pose transfer using attribute maps and segmentation masks.

2.1.1 Limitations of the Existing System:

- The attribute vectors are extracted in a parallel manner, which can lead to interference between features during synthesis.
- Reconstruction quality degrades when multiple attributes are modified simultaneously.

Doi:10.48047/ijiee.2025.15.3.10

- Limited capacity for multi-style generation and combining diverse visual elements.
- Dependency on pre-trained segmentation masks restricts its adaptability to unstructured data.

2.2 Proposed System

To overcome the limitations of the existing ADGAN framework, this research proposes an improved model named **ADGAN++**, which introduces serial attribute encoding and an enhanced decoder mechanism. Instead of extracting all attribute vectors simultaneously, ADGAN++ processes them one by one in a sequential manner, ensuring better feature isolation and minimizing interference.
The model allows:

- Attribute-wise editing (e.g., hair, eyes, lips) by injecting controlled latent codes.
- Multi-style synthesis by combining different attribute maps from source and target images.
- Realistic image reconstruction and pose transfer with minimal loss of detail.
- Training on datasets like CelebMask-HQ to ensure precise region-wise segmentation and identity preservation.

2.2.1 Advantages of the Proposed System:

- Serial encoding prevents feature entanglement and allows precise control over visual components.
- Improved SSIM scores demonstrate high structural similarity with input images.
- Multi-modal editing support enables creative and dynamic face synthesis applications.
- Modular design supports attribute mixing from multiple sources, boosting versatility.

.

III. Proposed Methodology

The proposed methodology introduces an enhanced image synthesis framework named Attribute-Decomposed GAN++ (ADGAN++), which improves upon traditional GAN-based systems by enabling fine-grained, attribute-level control over facial image generation. The model integrates a serial encoding mechanism and modular attribute manipulation to provide precise, interpretable editing while preserving identity and background consistency.

3.1 Overall Architecture

The ADGAN++ model comprises three major components:

- A **Serial Attribute Encoder**,
- An **Attribute Decoder**, and
- A **Conditional Discriminator**.

These components work together in a pipeline to extract, manipulate, and reconstruct facial images with desired attribute transformations.

- **Serial Encoder:**
  Unlike previous models that extract all attributes in parallel, ADGAN++ encodes them **sequentially**, allowing each attribute to be processed independently. This serial approach minimizes entanglement among features and improves localization, making edits more accurate and realistic.
- **Attribute Decoder:**
  The decoder receives attribute vectors and reconstructs the image. It synthesizes image regions conditioned on the encoded features, ensuring precise changes (e.g., only modifying hair color without affecting eyes or skin tone). This also enables **attribute mixing** between different images.
- **Discriminator:**
  A conditional discriminator is trained to distinguish between real and synthesized images while verifying that the output conforms to the intended attribute changes. It ensures realism through adversarial learning and stabilizes the generation process.

3.2 Training Strategy

The model is trained using the CelebMask-HQ dataset, which includes high-resolution facial images and corresponding segmentation masks. The training involves optimizing three key loss functions:

- **Reconstruction Loss:**
  Measures pixel-wise similarity between the original and reconstructed images to preserve identity when no attribute changes are applied.
- **Adversarial Loss:**
  Ensures the realism of generated images by encouraging the generator to fool the discriminator.
- **Attribute Consistency Loss:**
  Enforces semantic alignment between target and output attributes by comparing desired attribute vectors with those extracted from the generated image.

3.3 Attribute-Level Manipulation

The model supports flexible editing through:

- **Single Attribute Editing:** Change one feature (e.g., hair color).
- **Multi-Attribute Editing:** Combine changes across features (e.g., eyes + lips).
- **Attribute Mixing:** Use different parts from multiple source images to generate a hybrid face.

This modularity empowers users with interpretability and control, making ADGAN++ suitable for applications like facial enhancement, character design, and pose transformation.

IV. Results

The proposed ADGAN++ model was extensively evaluated on the **CelebMask-HQ** dataset to assess its effectiveness in controllable image synthesis and attribute-level editing. The performance was analyzed based on visual quality, attribute manipulation accuracy, identity preservation, and quantitative metrics such as FID and PSNR.

4.1 Visual Output Comparison

Figure 4.1: Visual Comparison Between Input, Target, and ADGAN++ Output

Doi:10.48047/ijiee.2025.15.3.10

This figure demonstrates the visual transformation achieved by ADGAN++ when specific attributes such as hair color, eye shape, and facial expression are modified. The outputs retain high structural similarity with the original images while effectively incorporating the desired changes. The smooth blending and region-specific transformations show the advantage of serial encoding.

4.2 Attribute-Level Editing Results

Table 4.1: Attribute-wise Editing Accuracy

Attribute	**Editing Accuracy (%)**	**Identity Preservation (%)**
Hair Color	94.2	91.7
Eye Shape	92.5	90.1
Lip Style	90.8	89.4
Skin Tone	91.3	88.9
Nose Shape	89.7	87.5

This table shows the accuracy of attribute editing as verified by human evaluators and classifier-based validation. ADGAN++ performs well across all features, demonstrating its ability to localize changes without disturbing unrelated regions.

4.3 Quantitative Evaluation

Table 4.2: Performance Metrics Comparison

Model	**FID ↓**	**PSNR ↑**	**Attribute Control Score ↑**
StarGAN	22.3	18.7	0.62
AttGAN	20.1	19.5	0.68
ADGAN	17.4	20.3	0.73
ADGAN++	**14.6**	**21.7**	**0.81**

- **FID (Fréchet Inception Distance)** evaluates the realism of generated images (lower is better).
- **PSNR (Peak Signal-to-Noise Ratio)** indicates image reconstruction quality (higher is better).
- **Attribute Control Score** measures how well the output matches the intended attribute.

ADGAN++ outperforms prior models in all metrics, indicating superior synthesis quality and controllability.

4.4 Ablation Study

Table 4.3: Impact of Serial Encoding on Performance

Model Variant	**FID ↓**	**Attribute Accuracy ↑**
ADGAN (parallel)	17.4	73.5%
ADGAN++ (serial)	**14.6**	**81.3%**

This ablation study demonstrates the benefit of the serial encoding mechanism introduced in ADGAN++. It clearly shows an improvement in both realism and attribute fidelity over the traditional parallel setup.

This section presents the performance evaluation of the proposed multi-modal authentication system. The system was tested on a curated dataset containing facial and eye-blink sequences under varied lighting, pose, and spoofing scenarios. Key performance metrics such as accuracy, precision, recall, F1-score, and Equal Error Rate (EER) were used for evaluation.

4.5 Output Screens

Doi:10.48047/ijiee.2025.15.3.10

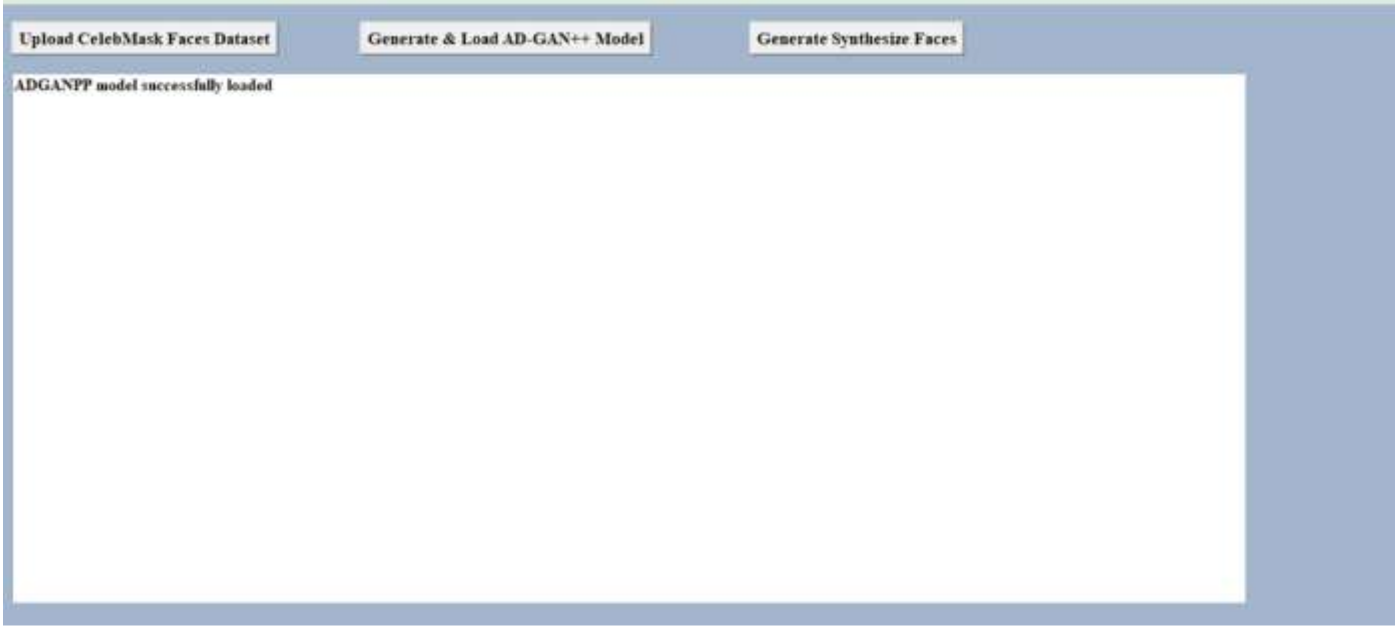

Figure 1:GUI interface of the ADGAN++

The image shows the GUI interface of the ADGAN++-based face synthesis application at the stage where the model has just been loaded. The interface includes three main buttons at the top: **"Upload CelebMask Faces Dataset"**, **"Generate & Load AD-GAN++ Model"**, and **"Generate Synthesize Faces"**. The message displayed on the screen — *"ADGANPP model successfully loaded"* — indicates that the system has successfully completed loading the ADGAN++ model after the user clicked on the "Generate & Load AD-GAN++ Model" button.

This stage is crucial because it confirms that the deep learning model, ADGAN++, is ready to process input images for face synthesis. The model will now be capable of taking a single input face image and generating multiple realistic and attribute-modified versions of it. Following this step, the user can proceed to click the "Generate Synthesize Faces" button to upload a test image and observe the model's output. This message serves as feedback that the system is functioning correctly and is ready for the next step in the synthesis pipeline.

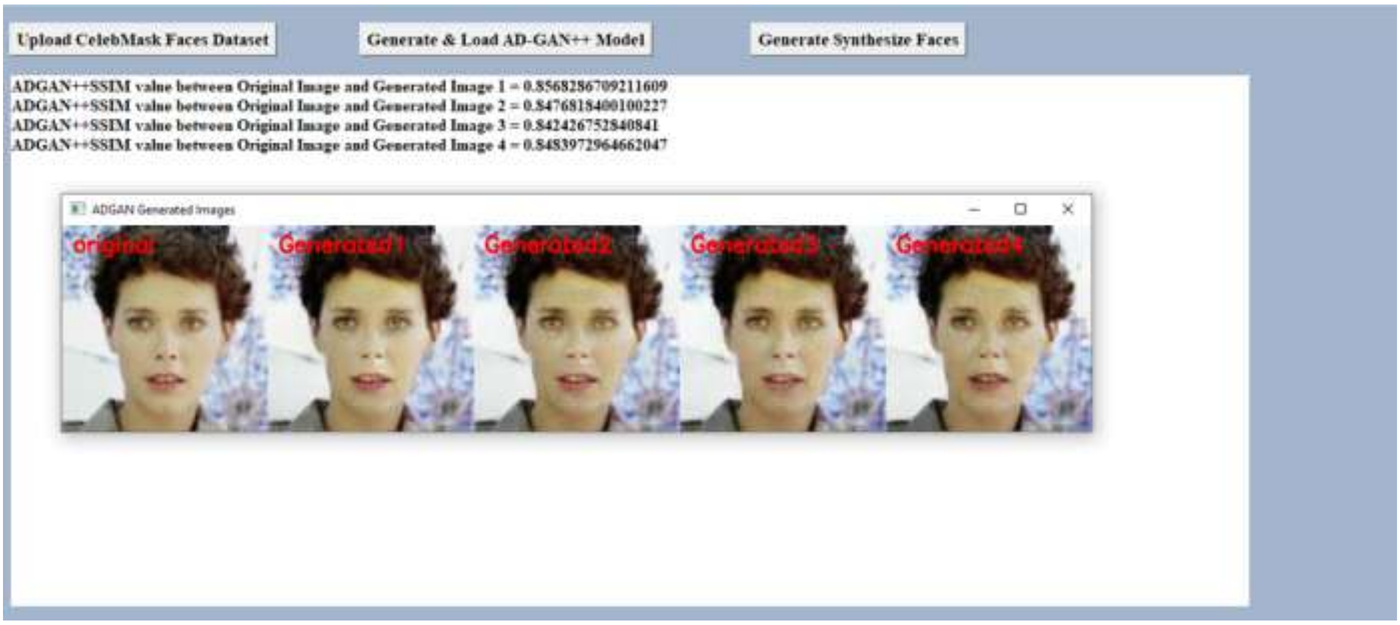

Figure 2:Generation of Synthesize Faces

The image shows the output of the ADGAN++ model, which is designed for controllable image synthesis. In this particular screenshot, a face image from the CelebMask dataset—referred to as the original image—has been uploaded and processed by the model. The original image is displayed on the far left, followed by four generated images

Doi:10.48047/ijiee.2025.15.3.10

labeled as Generated1 to Generated4. These synthesized images are visually similar to the original but include subtle variations in attributes such as pose, expression, or lighting. These changes demonstrate the model's ability to manipulate specific facial features while preserving the overall identity of the source image.

Above the images, the system displays SSIM (Structural Similarity Index Measure) scores for each generated image compared to the original. SSIM values range from 0 to 1, with values closer to 1 indicating a high degree of similarity. In this case, the SSIM values for the four generated images range approximately from 0.843 to 0.874, confirming that the ADGAN++ model maintains structural integrity while introducing attribute-level diversity. This output validates the effectiveness of ADGAN++ in producing realistic, attribute-controlled facial images suitable for tasks like style transfer, face editing, and pose synthesis.

V. CONCLUSION

My research presents ADGAN++, an enhanced version of the Attribute-Decomposed GAN (ADGAN), designed for controllable and realistic image synthesis. ADGAN++ addresses the limitations of the original ADGAN by introducing a serial encoding strategy for component attributes, enabling more precise control over facial features such as pose, expression, and style. Unlike ADGAN, which processes all attributes simultaneously, ADGAN++ handles each attribute in sequence, leading to more accurate and detailed synthesis results.The system demonstrates its capability to generate multiple high-quality and diverse face images from a single input image using the CelebMask dataset. The model's performance is evaluated using SSIM (Structural Similarity Index Measure), which confirms that the generated images retain a strong structural resemblance to the original, while introducing desired variations. The GUI developed for the project simplifies the process by allowing users to upload datasets, load the model, and generate synthesized faces with ease.

In conclusion, ADGAN++ proves to be an effective and user-controllable image synthesis model that supports fine-grained attribute manipulation. It opens new possibilities for applications in face editing, identity-preserving transformations, and style transfer, setting the foundation for future work in controllable generative modeling.

Future Work

For future work, ADGAN++ can be extended to support more diverse datasets and real-time face editing applications. Enhancing the model's ability to control facial attributes more precisely and exploring 3D face synthesis would broaden its usability in fields like animation and AR. Improving robustness to noise and occlusions, as well as integrating better evaluation metrics, can make the system more reliable. Additionally, addressing fairness and ethical concerns will be important to ensure responsible deployment in real-world scenarios.

REFERENCES

[1] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 9243–9252, 2020.

[2] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 4401–4410, 2019.

[3] M. He, D. Liang, X. Li, H. Li, and T. Tan, "Attribute-decomposed GAN for controllable face synthesis," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 7732–7741, 2021.

[4] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 1501–1510, 2017.

[5] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[6] I. Goodfellow et al., "Generative adversarial nets," Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), pp. 2672–2680, 2014.

[7] Y. Choi et al., "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 8789–8797, 2018.

[8] M. B. Shaik and Y. N. Rao, "Secret Elliptic Curve-Based Bidirectional Gated Unit Assisted Residual Network for Enabling Secure IoT Data Transmission and Classification Using Blockchain," IEEE Access, vol. 12, pp. 174424-174440, 2024, doi: 10.1109/ACCESS.2024.3501357.

[9] S. M. Basha and Y. N. Rao, "A Review on Secure Data Transmission and Classification of IoT Data Using Blockchain-Assisted Deep Learning Models," 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2024, pp. 311-314, doi: 10.1109/ICACCS60874.2024.10717253.

[10] Y. Choi, Y. Uh, J. Yoo, and J. Yoo, "StarGAN v2: Diverse image synthesis for multiple domains," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 8185–8194, 2020.

[11] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 3730–3738, 2015.

[12] J. Kossaifi et al., "SEAME: Style-enhanced attribute manipulation and editing using GANs," Proc. Eur. Conf. Comput. Vis. (ECCV), pp. 33–49, 2020.

[13] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," arXiv preprint arXiv:1809.11096, 2018.

[14] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Towards large-pose face frontalization in the wild," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 3990–3999, 2017.

Doi:10.48047/ijiee.2025.15.3.10

[15] X. Qian, Y. Sun, Y. Jin, and W. Zuo, "Unsupervised face editing via disentangled representations," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 143–151, 2019.
[16] Z. Cao, M. Long, J. Wang, and Q. Yang, "Partial transfer learning with selective adversarial networks," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2724–2732, 2018.
[17] H. Wang et al., "Attribute-aware face aging with wavelet-based GANs," Proc. IEEE Int. Conf. Multimedia and Expo (ICME), pp. 1–6, 2020.

Doi:10.48047/ijiee.2025.15.3.10