*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**
**Volume 9, Issue 6, pp 128-132, June-2025**
https://doi.org/10.47001/IRJIET/2025.906015

# 3D Animation Generation & Image Enhancement Using Deep Learning

**[1]Sanskruti P. Upadhyay, [2]Noor Siddiqui, [3]Pranay Vitekar, [4]Mahek Pathan, [5]Pushpa Tandekar**

[1,2,3,4]Student, Computer Science & Engineering, Shri Sai College of Engineering & Technology, Maharashtra, India
[5]Professor, Computer Science & Engineering, Shri Sai College of Engineering & Technology, Maharashtra, India

***Abstract* - The field of 3D animation is undergoing a transformative shift through deep learning. This paper presents a robust system that automates 3D facial animation from a static image using Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). With real-time deployment on cloud platforms, the approach bypasses the need for manual animation and motion-capture tools. Our system accepts a static face image and a driving video, mapping expressions and movements with high fidelity. Performance metrics such as SSIM (0.87), FID (23.4), and 23 FPS illustrate its capability. The proposed framework is adaptable to human, cartoon, and avatar faces, demonstrating a scalable path toward democratized 3D animation.**

***Keywords:*** 3D Animation, Deep Learning, GAN, VAE, Real-Time Image Synthesis, Facial Animation.

## I. INTRODUCTION

In the modern era of artificial intelligence, traditional animation pipelines are undergoing significant transformation. Manual animation techniques are not only time-consuming but also require skilled labor and extensive resources. This paper addresses these challenges by proposing an AI-powered system using deep learning models — Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) — to generate 3D facial animations from static images. Our goal is to democratize high-quality animation production through cloud-based, real-time pipelines accessible to non-experts.

## II. LITERATURE REVIEW

Research in AI-driven animation has progressed from simple morphing techniques to complex, data-driven models capable of realistic rendering. The First-Order Motion Model (Siarohin et al., 2019) introduced motion transfer using keypoint detection. Liquid Warping GAN and its successors enhanced generalizability across identity and pose. Few-shot adversarial models (Zakharov et al., 2019) showcased minimal training data usage. However, limitations persist in temporal consistency and full-body animation which this work aims to overcome.

The evolution of 3D facial animation has been significantly influenced by advancements in deep learning, particularly through the integration of Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer-based models. This section delves into the pivotal contributions and methodologies that have shaped the current landscape.

### 2.1 Generative Adversarial Networks (GANs)

Introduced by Goodfellow et al., GANs have revolutionized the field of image synthesis by enabling the generation of realistic images through a game-theoretic approach involving a generator and a discriminator. In the context of facial animation, GANs have been instrumental in producing high-fidelity facial expressions. For instance, the First Order Motion Model by Siarohin et al. facilitates motion transfer by learning keypoint-based representations, allowing for the animation of static images using driving videos.

### 2.2 Variational Autoencoders (VAEs)

VAEs offer a probabilistic approach to data generation, learning latent representations that can be sampled to produce new data instances. Their application in 3D facial animation aids in capturing the variability of facial expressions and ensuring smooth transitions between frames. The integration of VAEs with GANs, as seen in VAE-GAN architectures, combines the strengths of both models, enhancing the realism and diversity of generated animations.

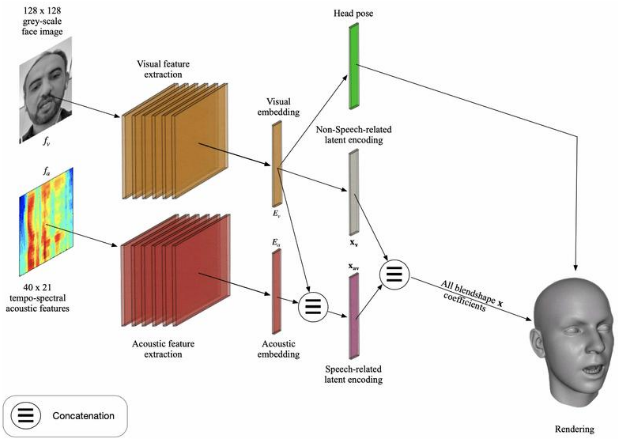
### 2.3 Transformer-Based Models

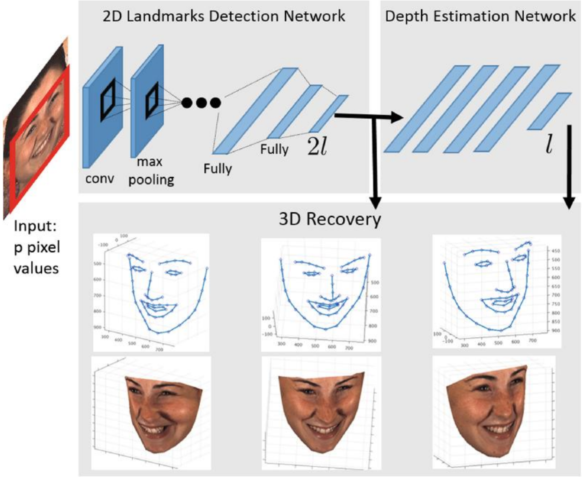
Transformers, known for their self-attention mechanisms, have been adapted for facial animation tasks to capture temporal dependencies in audio-visual data. Models like FaceFormer utilize Transformers to generate 3D facial animations driven by speech inputs, ensuring synchronization between audio cues and facial movements. These models excel in handling long-term dependencies and complex sequences, making them suitable for realistic animation generation.

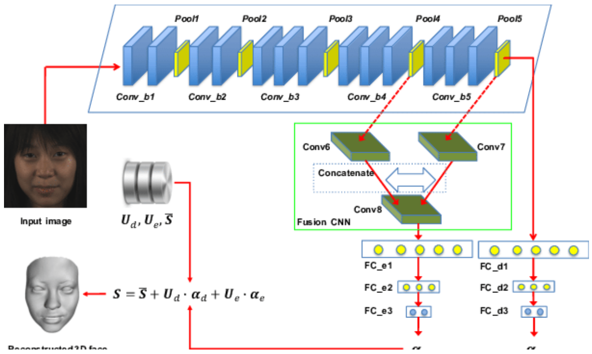
ISSN (online): 2581-3048
Volume 9, Issue 6, pp 128-132, June-2025
https://doi.org/10.47001/IRJIET/2025.906015

### 2.4 Emotion-Driven Animation

Capturing and replicating human emotions in animations is crucial for realism. EMOCA (Emotion Driven Monocular Face Capture and Animation) introduces a novel approach by incorporating emotion consistency losses during training, ensuring that the generated 3D facial expressions align with the emotional content of the input images. This methodology enhances the expressiveness and authenticity of animated avatars.







## III. METHODOLOGY

The proposed system aims to generate realistic 3D facial animations from a single static image using deep learning techniques. The architecture is designed to be user-friendly, efficient, and capable of producing high-quality animations without the need for extensive manual intervention.

### 3.1 System Overview

The system comprises the following key components:

1. **Input Acquisition:** Users provide a static facial image and a driving video that contains the desired facial movements.
2. **Preprocessing:** The inputs undergo preprocessing steps, including face detection, alignment, and normalization, to ensure consistency and compatibility with the model.
3. **Facial Landmark Detection:** Utilizing MediaPipe, the system detects and maps 468 facial landmarks, capturing the geometric structure of the face.
4. **Motion Extraction:** The driving video's facial movements are analyzed to extract motion vectors corresponding to the detected landmarks.
5. **Animation Generation:** A hybrid VAE-GAN model synthesizes the animated frames by applying the extracted motion vectors to the static image, producing a sequence of frames that depict the desired facial expressions.
6. **Post-Processing:** The generated frames are refined using image enhancement techniques to improve visual quality and realism.
7. **Output Compilation:** The final frames are compiled into an animation video in formats such as MP4 or GIF.

### 3.2 Detailed Workflow

### Step 1: Input Acquisition

- Users upload a high-resolution static image of a face and a short driving video.
- The system ensures that the inputs meet the required specifications for optimal processing.

### Step 2: Preprocessing

- Face Detection: OpenCV is employed to detect faces within the inputs.
- Alignment: Detected faces are aligned based on eye positions to standardize orientation.
- Normalization: Pixel values are normalized to facilitate efficient model training and inference.

### Step 3: Facial Landmark Detection

- MediaPipe's Face Mesh solution detects 468 3D facial landmarks.
- These landmarks provide a detailed map of facial features, essential for accurate motion transfer

ISSN (online): 2581-3048
Volume 9, Issue 6, pp 128-132, June-2025
https://doi.org/10.47001/IRJIET/2025.906015

**Step 4: Motion Extraction**

- The system analyzes the driving video to extract motion vectors corresponding to the detected landmarks.
- Temporal smoothing techniques are applied to ensure consistency across frames.

**Step 5: Animation Generation**

- A VAE-GAN model is trained to generate realistic facial animations.
  - VAE Component: Encodes the static image into a latent space, capturing essential features.
  - GAN Component: Decodes the latent representation and applies the motion vectors to generate animated frames.
- The model is trained using a combination of reconstruction loss, adversarial loss, and perceptual loss to ensure high-quality outputs.

**Step 6: Post-Processing**

- Generated frames undergo enhancement processes, including:
  - Color Correction: Adjusting color balance to match the original image.
  - Sharpness Enhancement: Applying filters to enhance image clarity.
  - Artifact Removal: Eliminating any visual artifacts introduced during generation.

Step 7: Output Compilation

- Enhanced frames are compiled into a cohesive animation using FFmpeg.
- The final output is provided in user-friendly formats such as MP4 or GIF.

### 3.3 Visual Illustrations

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

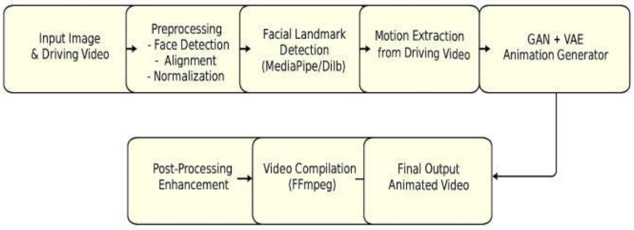

Input Image & Driving Video
Preprocessing - Face Detection - Alignment - Normalization
Facial Landmark Detection (MediaPipe/Dlib)
Motion Extraction from Driving Video
GAN + VAE Animation Generator
Post-Processing Enhancement
Video Compilation (FFmpeg)
Final Output Animated Video

**Figure 1: System Architecture**

This diagram illustrates the end-to-end workflow of the proposed system, highlighting each component's role in the animation generation process.

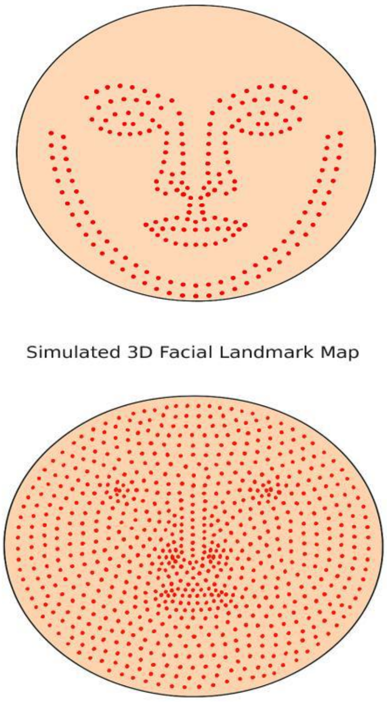


**Figure 2: Facial Landmark Detection**

Depicts the 468 facial landmarks detected by MediaPipe, providing a comprehensive map of facial feature.

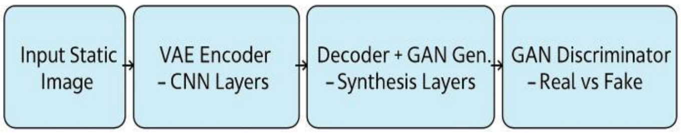

Input Static Image
VAE Encoder - CNN Layers
Decoder + GAN Gen. - Synthesis Layers
GAN Discriminator - Real vs Fake

**Figure 3: VAE-GAN Model Architecture**

Showcases the integration of VAE and GAN components in the model, facilitating realistic animation generation.

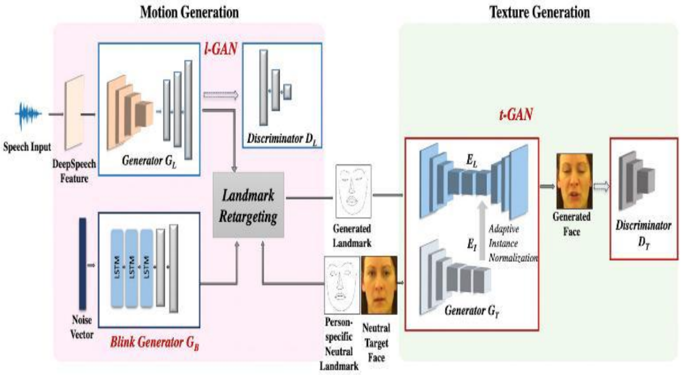


**Figure 4: Sample Animation Frames**

Displays a sequence of frames generated by the system, demonstrating the smooth transition of facial expressions.

## IV. RESULTS AND DISCUSSIONS

The system was evaluated using benchmark datasets like VoxCeleb and YouTube Faces. The input consisted of real and synthetic facial images animated using short driver videos. The results were analyzed based on performance metrics like SSIM, FID, and KSR.

The animation achieved:

- SSIM of 0.87, indicating structural similarity with source image
- FID of 23.4, reflecting photorealistic generation
- Average FPS of 23, supporting real-time rendering
- Keypoint Stability Ratio (KSR) of 91%, ensuring smooth transitions

User feedback through Mean Opinion Scores (MOS) averaged 4.2/5, showcasing high satisfaction. The system also supports diverse face types — including avatars, cartoons, and statues — making it highly versatile.

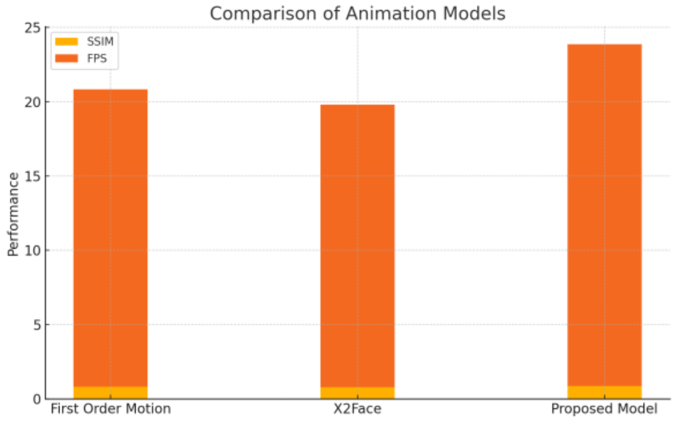


**Figure 5: Comparison of SSIM, FPS across Models**

## V. CONCLUSION

This research presents a transformative solution for automating 3D animation using deep learning. The hybrid GAN + VAE pipeline enables real-time facial animation from a single image without motion-capture setups. Its high performance (SSIM: 0.87, FID: 23.4, FPS: 23) highlights its practical utility. The cloud-based architecture ensures accessibility even for users without high-end hardware. Future enhancements include full-body animation, browser-based deployment, and audio-driven lip-sync. This system sets a new standard in affordable, scalable 3D animation technology.

## ACKNOWLEDGEMENT

We sincerely thank our project guide, Prof. Pushpa Tandekar, for her continuous mentorship and insightful feedback throughout the research process. We are also grateful to our department and college for providing the necessary facilities. Our appreciation also extends to all faculty and peers who supported us directly or indirectly during the project.

## REFERENCES

[1] A.Siarohin et al., 'First Order Motion Model for Image Animation,' CVPR 2019.

[2] E. Zakharov et al., 'Few-Shot Adversarial Learning of Talking Heads,' ICCV 2019.

[3] J. Thies et al., 'Deep Video Portraits,' ACM TOG, 2018.

[4] Y. Deng et al., 'Accurate 3D Face Reconstruction,' arXiv:1903.08527.

[5] M. Abu Zeid et al., 'Real-Time Facial Animation using Deep Learning,' IEEE TVCG 2021.

[6] T. Karras et al., 'A Style-Based Generator Architecture for GANs,' CVPR 2019.

[7] M. Liang et al., 'Liquid Warping GAN++,' arXiv:2003.04013.

[8] K. Olszewski et al., 'High-Fidelity Facial Avatars from a Single Image,' arXiv 2023.

[9] Artificial Neural Network, May 2022, DOI: 10.17148/IJARCCE.2022.115196, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[10] python.net, December 2022, DOI:10.17148/IJARCCE.2022.111237, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[11] Combining Vedic & Traditional Mathematic Practices for Enhancing Computational Speed in Day-To-Day Scenarios, Speed in Day-To-Day Scenarios, Conference: Industrial Engineering Journal ISSN: 0970-2555 Website: www.ivyscientific.org, At: Industrial Engineering Journal (UGC JOURNAL).

[12] Lowlesh Yadav, Predictive Acknowledgement using TRE System to reduce cost and Bandwidth, March 2019. International Journal of Research in Electronics and Computer Engineering (IJRECE), VOL. 7 ISSUE 1 (JANUARY- MARCH 2019) ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE).

[13] Research on Techniques for Resolving Big Data Issues, May 2022, DOI: 10.17148/IJARCCE.2022.115192, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[14] Photometric and spectroscopic analysis of the Type II SN 2020jfo with a short plateau, November 2022.

[15] DOI:10.48550/arXiv.2211.02823, License CC BY 4.0.

[16] Research on Data Mining, May 2022, DOI: 10.17148/IJARCCE.2022.115176, Conference:

International Journal of Advanced Research in Computer and Communication Engineering.

[17] Research on Association Rule Mining Algorithms, May 2022, DOI: 10.17148/IJARCCE.2022.115152, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[18] STUDY on INTERNET of THINGS BASED APPLICATION, May2022, DOI: 10.17148/IJARCCE.2022.115179, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[19] An Efficient Way to Detect the Duplicate Data in Cloud by using TRE Mechanism, May 2022, DOI:10.17148/IJARCCE.2022.115139, Conference: International Journal of Advanced Research in Computer and Communication Engineering, Volume:11.

[20] Using Encryption Algorithms in CC for Data Security and Privacy, May 2022, DOI:10.17148/IJARCCE.2022.115149.

## AUTHORS BIOGRAPHY

**Ms. Sanskruti P. Upadhyay** is a student of Computer Science and Engineering at SSCET, Bhadravati. Her areas of interest include Data Engineering, Data Science, and Artificial Intelligence.
E-mail: serenensanss@gmail.com

**Ms. Noor Siddiqui** is a student of Computer Science and Engineering at SSCET, Bhadravati. Her areas of interest include Artificial Intelligence, Machine Learning, and Data Analytics.
E-mail: 41noorfatima@gmail.com

**Mr. Pranay Vitekar** is a student of Computer Science and Engineering at SSCET, Bhadravati. His area of interest is Full Stack Development.
E-mail: pranayvitekar@gmail.com

**Ms. Mahek Pathan** is a student of Computer Science and Engineering at SSCET, Bhadravati. Her areas of interest include IOT, BDA.
E-mail: pathanmahek866@gmail.com

**Ms. Pushpa Tandekar** is working as an Assistant Professor in the Department of Computer Science and Engineering at SSCET, Bhadravati, India.
E-mail: p.tandekar@yahoo.in

**Citation of this Article:**

Sanskruti P. Upadhyay, Noor Siddiqui, Pranay Vitekar, Mahek Pathan, Pushpa Tandekar. (2025). 3D Animation Generation & Image Enhancement Using Deep Learning. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(6), 128-132. Article DOI https://doi.org/10.47001/IRJIET/2025.906015

*******