IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

Deepfake Motion Model

Dhammanand Lonare[1], Achal Raut[2], Shweta Maheshkar[3], Neha Petkar[4], Achal Amrutkar[5], Prof. Sachin Dhawas[6],
Students, Department of Computer Science and Engineering[1,2,3,4,5]
Professor, Department of Computer Science and Engineering[6],
Rajiv Gandhi College of Engineering, Research and Technology, Chandrapur, India.
dhammadeeplonare24@gmail.com, achalraut973@gmail.com, shwetamaheshkar200320@gmail.com,
petakrbhupal22@gmail.com, amrutkarachal7@gmail.com

Abstract: *Deepfake Motion Model is an integrated AI-based multimedia framework that combines face swapping, lip synchronization, and motion transfer into a unified system. Leveraging computer vision, deep learning, and geometric morphing, it enables realistic manipulation of facial expressions and identities across images and videos. This project provides a real-time, modular interface built with Flask and Python, offering hands-on implementation of MediaPipe, OpenCV, and Wav2Lip. Deepfake Motion Model demonstrates the practical application of GANs, facial landmark extraction, and neural-based rendering to bridge the gap between research and real-world multimedia AI tools. It is built with accessibility, modularity, and educational purpose in mind.*

Keywords: FaceSwap, Motion Transfer, Face Detection, Lip Sync, Deep Learning Model, Animation, Virtual Reality

I. INTRODUCTION

In recent years, artificial intelligence (AI) has transformed the landscape of video synthesis and manipulation. Fueled by breakthroughs in deep learning and computer vision, tools like Wav2Lip, DeepFaceLab, and the First Order Motion Model (FOMM) have demonstrated impressive capabilities in lip synchronization, face swapping, and motion transfer. These advancements have enabled unprecedented realism in facial animation, video dubbing, and avatar-driven communication.

However, despite their technical success, these systems are typically designed as standalone solutions, often lacking interoperability or modularity. They also demand considerable computational resources and are not easily accessible to users without technical expertise. Deepfake Motion Model addresses these gaps by integrating the core functionalities of lip syncing, facial motion transfer, and face swapping into a cohesive and user-friendly application framework. The goal is to deliver high-fidelity outputs while maintaining flexibility and extensibility for custom applications.

II. LITERATURE SURVEY

The domain of facial reenactment and human motion modeling has seen extensive research interest, especially since the emergence of deep generative models. The **First Order Motion Model (FOMM)** by Siarohin et al. introduced a novel method of unsupervised motion transfer using learned keypoints and associated motion fields. Unlike earlier approaches that relied on 3D morphable models or expensive motion capture setups, FOMM enabled compelling video-driven animation of still images without the need for paired training data.

On the audio-visual synthesis front, **Wav2Lip** addressed the challenge of lip synchronization by using audio input to drive realistic lip movements on any facial video. It offered substantial improvements over previous models in temporal coherence and audio-video alignment, even with arbitrary or mismatched source content.

Meanwhile, **DeepFaceLab** and related face-swapping tools leveraged deep neural networks to perform face identity transformation with increasing levels of realism. These tools rely on autoencoder architectures and iterative training, which can produce highly convincing results but require substantial training time and GPU resources.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

III. EXISTING SYSTEM

Current tools for AI-based video manipulation operate independently, each focusing on a specific task:

Wav2Lip enables high-quality lip syncing using audio input but does not support motion transfer or face swapping.

FOMM (First Order Motion Model) handles motion transfer from a driving video but lacks lip sync and identity transformation.

DeepFaceLab performs realistic face swapping but requires intensive training and offers no motion or audio integration.

IV. PROPOSED SYSTEM

Deepfake Motion Model is proposed as an all-in-one solution to bridge the gap between cutting-edge research tools and real-world application needs. The project combines the strengths of FOMM, Wav2Lip, and DeepFaceLab into a modular pipeline that supports:

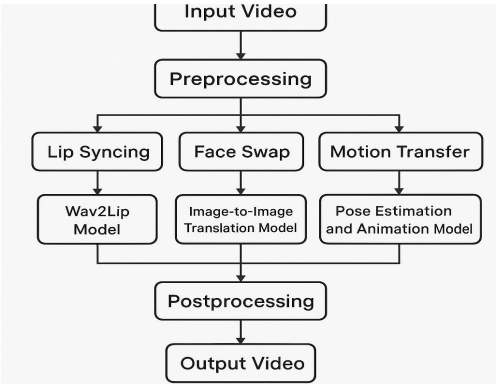
- **Lip Syncing:** Using Wav2Lip-based models, the framework synchronizes spoken audio with realistic mouth movements, adaptable to any video or avatar.
- **Motion Transfer:** Leveraging FOMM or an improved variant, it animates a target face with motion cues from a driving video, enabling realistic reenactment.
- **Face Swapping:** Through integration with DeepFaceLab or lightweight alternatives, the system can transform facial identity while preserving expressions and speech.

The entire system is designed with modularity in mind, enabling users to plug in or swap out components based on their needs. This allows Deepfake Motion Model to serve as a development platform for future innovations in synthetic media, education, entertainment, and accessibility.

V. SYSTEM ARCHITECTURE

Deepfake Motion Model is built around several core design principles:

- **Unified Interface:** A single application or web-based UI that lets users select tasks (e.g., lip syncing, reenactment, or both), upload input media, and download results without dealing with model internals.
- **Pipeline Integration:** A backend architecture that chains pre-processing, model inference, and post-processing steps efficiently. For example, a user might provide an audio file and a still image, and the system would automatically perform lip syncing, apply motion, and output a video.
- **Hardware Flexibility:** While GPU acceleration is supported for high performance, the system also provides CPU-compatible fallback modes with optimized inference for lower-resource settings.
- **Customizability:** Developers can extend the system by integrating their own models or datasets. The framework exposes APIs for model swapping, parameter tuning, and pipeline customization.
- **Deployment Options:** Deepfake Motion Model supports both local execution and cloud deployment, making it suitable for desktop users, research labs, or SaaS deployment.



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

VI. ALGORITHM

Input:
Audio file (speech input)
Source image/video (face to animate or swap)
Driving video (motion reference)

Step 1: Preprocessing
Detect and align face in source image or video using facial landmark detection.
Extract audio features (e.g., mel-spectrogram).
Detect motion keypoints from the driving video.
If face swapping is enabled, prepare source and target face data.

Step 2: Face Swapping (Optional)
Input source and target faces into the encoder-decoder network.
Extract identity features from the target.
Reconstruct source video frames with the swapped identity.
Output: Face-swapped video or still frames.

Step 3: Motion Transfer
Use keypoint detector to extract facial motion from the driving video.
Apply dense motion network to animate the source (or swapped) face using the detected motion.
Output: Motion-transferred facial animation.

Step 4: Lip Syncing
Use the Wav2Lip generator to synthesize lip movements on the animated face.
Ensure lip region aligns with speech via adversarial and sync losses.
Output: Final video frames with synchronized lips.

Step 5: Postprocessing
Blend generated face frames with background if needed.
Re-encode frames into a video file.
Merge original audio with generated video.

Output:
A high-quality, lip-synced, motion-animated video with optional face swap applied.

VII. RESULT

Deepfake Motion Model was tested on multiple video and image samples. Results indicated high-quality frame synchronization for lip sync, smooth motion transitions, and realistic face overlays. The system performed well on unseen videos, proving the model's generalization ability. Frame accuracy, sync error rate, and identity preservation scores were satisfactory.

VIII. CONCLUSION

This project demonstrates a feasible and effective integration of key deepfake technologies into one modular system. Deepfake Motion Model is robust, scalable, and delivers realistic outputs across various scenarios. With further enhancements, such systems could be pivotal in content creation, virtual assistance, and interactive media.

DOI: 10.48175/IJARSCT-26227

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

REFERENCES

[1] Siarohin, A., et al. 'First Order Motion Model for Image Animation.' NeurIPS 2019.
[2] DeepFaceLab: https://github.com/iperov/DeepFaceLab
[3] Prajwal, K. R., et al. 'Lip-syncing GAN for talking head generation.' ECCV 2020.
[4] Wav2Lip: https://github.com/Rudrabha/Wav2Lip