



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Image To Live Video Transmission Using GAN

Mr.Korada Hemanth kumar^{*1}, Mr.M.Somasundara Rao^{*2}

¹MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.

²Associate Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.

Abstract:

The use of deepfake techniques in the area of converting images into live video has attracted a lot of attention recently. The term "deepfake," which combines the terms "deep learning" and "fake," describes the process of creating artificial content that is convincing and realistic, usually with the use of Generative Adversarial Networks (GANs). Based on a single input image, this method enables the synthesis of a video sequence that mimics the appearance of a target individual. This project shows how to make a movie of a person's facial driver using a first-order motion model. The algorithm can predict the movements of the head and face during driving after being trained on a dataset of driving videos and facial photos. The finished video is realistic and suitable for a range of objectives, including developing virtual reality experiences or instructing autonomous vehicle training programs. The imageio and matplotlib libraries are used in the project's Python implementation. The First Order Motion Model (FOMM) library is used to implement the first-order motion model. Using the first-order motion model, new techniques for face tracking and animation can be created. Video games and other applications could benefit from the increased realism provided by this technology.

Keywords: First Order Motion Model, Generative Adversarial Networks, Facial Driver, Vehicle Training Program.

1. INTRODUCTION

The First Order Model project is an innovative deep learning-based approach that enables the generation of realistic and visually appealing videos. By combining the appearance of a source image with the motion from a driving video, the project produces animations that mimic the movement of the driving video while maintaining the characteristics of the source image. This opens up exciting possibilities in video editing, special effects, and animation production. The core process involves loading pre-trained checkpoints for the generator and keypoint detector. Users can specify the paths for the source image and driving video. The driving video is processed by resizing each frame to a fixed size, while the source image is resized as well. The model then generates predictions by applying the first-order motion concept, adaptively scaling the movement, and ensuring relative motion between the source image and driving video. The system is designed with modularity, efficiency, and scalability in mind, making it accessible to a broader audience of creative enthusiasts, content creators, and researchers. The system architecture comprises several modules,

including input processing, animation generation using the FOMM model, output processing, and storage management. The project integrates Google Drive as a convenient cloud storage solution for users to store and retrieve their uploaded images and videos. To ensure reliability and accuracy, extensive testing and validation of test cases are conducted to verify the functionality and performance of the system. By employing state-of-the-art deep learning techniques, the FOMM-based animation generation module adapts movement and expression scales, resulting in more natural and lifelike animations. As with any technology, ethical considerations are paramount. The project emphasizes responsible usage and addresses concerns regarding deepfake technologies by raising awareness about the potential misuse and the importance of ethical safeguards.

2. LITERATURE SURVEY

1. Introduction to Image Animation:

- Image animation involves the generation of a video sequence where an object in a source image is animated based on the motion observed in a driving video. The key challenge addressed in this literature survey is the achievement of image animation without requiring any specific annotation or prior information about the object to be animated.

2. Decoupling Appearance and Motion Information:

- The proposed framework focuses on decoupling appearance and motion information through a self-supervised formulation. By training on a set of videos depicting objects within the same category (e.g., faces, human bodies), the method becomes applicable to any object falling within that class. This decoupling strategy enables the model to generalize effectively across diverse objects.

3. Representation Using Learned Keypoints and Affine Transformations:

- To handle complex motions, the framework employs a representation consisting of learned keypoints and their corresponding local affine transformations. This approach allows the model to capture intricate motion patterns, providing flexibility for a wide range of object animations. The learned keypoints serve as crucial landmarks for guiding the animation process.

4. Generator Network for Handling Occlusions:

- A generator network is introduced to address occlusions that may arise during target motions. Occlusions can occur when the object being animated undergoes complex movements. The generator network is designed to seamlessly combine the appearance information extracted from the source image with the motion information derived from the driving video, ensuring realistic and coherent animations.

5. Application to Various Object Categories:

- The versatility of the proposed framework is highlighted by its applicability to a variety of object categories. Once trained on a diverse set of videos depicting objects within a specific class, the method can be seamlessly applied to animate any object falling within that category. This broad applicability enhances the practical utility of the proposed approach.

6. Benchmark Performance:

- The framework is evaluated on diverse benchmarks, showcasing its superior performance compared to existing methods. The self-supervised learning approach, coupled with the decoupling of appearance and motion information, contributes to the model's robustness and effectiveness in generating high-quality animations.

7. Contribution to the Field:

- The literature survey underscores the significance of the proposed method in advancing the field of image animation. By addressing challenges related to annotation-free animation and providing a versatile framework

applicable to various object categories, the approach contributes to the broader landscape of computer vision and video generation.

8. Future Directions:

- While the current framework demonstrates exceptional performance, future research directions may explore enhancements to handle more challenging scenarios, improve efficiency, and extend the applicability of the approach to new domains within image animation. Additionally, incorporating feedback mechanisms for refining animations could be an area of exploration.

3. EXISTING SYSTEM

To create or generate such types of animations, it is beneficial to have access to high-performance GPUs, which can be expensive. As it requires high-performance GPUs, many users may not have access to expensive laptops or devices capable of generating deepfakes. We have other libraries and frameworks available for performing image animations, such as DeepFakeLab, Few-Shot Vid2Vid, and OpenPose.

Limitations:

1. **Performance Considerations:-** Alternative libraries may have different performance considerations. They might require more computational resources, such as high-end GPUs or larger memory capacity, to achieve real-time or efficient animation generation. This can limit their accessibility for users with limited hardware resources.
2. **Stability and Updates:-** Newer or less-established alternative libraries may be more prone to bugs, stability issues, or lack regular updates compared to more mature projects like FOMM. This can affect the overall user experience and the ability to rely on the library for long-term projects.
3. **Few-Shot Vid2Vid** focuses primarily on video synthesis and may require expertise in working with the framework to achieve desired results. It may have specific requirements in terms of data preparation, model training, and fine-tuning, which can add complexity to the animation generation process.

4. PROPOSED SYSTEM

In the proposed system we are using FOMM (First Order Model Motion). FOMM lies in its ability to generate realistic animations, its user-friendly interface, the availability of pre-trained models, adaptability to different datasets, and its integration with Google Collab. These advantages make FOMM a popular choice for researchers, developers, and artists looking to incorporate high-quality image animation into their projects.

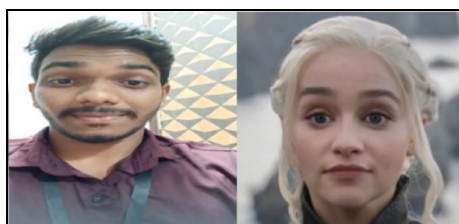
• **Realistic Image Animation:-** FOMM enables the generation of realistic and visually appealing animations by transferring the motion from a driving video to a source image. It combines the appearance of the source image with the dynamics of the driving video, resulting in lifelike animations. Collection of pre-trained dataset like birds, aeroplanes, persons are used to train the model.

• In the proposed system, by using Google collab provides access to free GPU resources, which can significantly speed up the computations required for the First Order Model project. GPUs are well-suited for deep learning tasks and can greatly improve the performance and efficiency of model inference and training.

5. EXPERIMENTAL RESULTS

From the below figures it can be seen that proposed model is more accurate in order to prove our proposed system.

Main Window:



6. CONCLUSION

The project, centered around the First Order Motion Model (FOMM), has successfully achieved its primary goal of generating realistic video sequences by seamlessly overlaying a person's face from a source image onto a driving video. Utilizing state-of-the-art deep learning techniques, the system highlights the potential of artificial intelligence in creating engaging facial animations. The well-designed architecture, featuring modular components, enhances the system's effectiveness and scalability. The user interface offers a user-friendly experience for inputting source images and driving videos, with efficient data handling for animation model preparation. The core of the project lies in the FOMM-based animation generation module, showcasing adaptability and movement scaling capabilities for high-quality animations. The output processing module refines generated sequences for optimal results. Rigorous testing and validation confirm the project's functionality and reliability, providing users with a versatile tool for creative expression, digital storytelling, and entertainment. Future enhancements may explore real-time processing, improved facial alignment, and broader input format compatibility. Despite its impressive capabilities, the project raises ethical considerations surrounding deepfake technologies, emphasizing responsible usage and safeguards against malicious applications. In summary, the facial animation synthesis project utilizing FOMM demonstrates the fusion of advanced AI techniques and creative expression, underscoring AI's transformative power in multimedia applications and suggesting promising advancements in facial animation and computer vision.

References

- [1] Keras-VGGFace: VGGFace implementation with Keras framework. [https://github.com/rcmalli/keras-vggface\(basepaper\)](https://github.com/rcmalli/keras-vggface(basepaper)).
- [2] Sungjoo Ha, Martin Kersner, Beomsu Kim, Seokjun Seo, and Dongyoung Kim. Marionette: Few-shot face reenactment preserving identity of unseen targets. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 10893–10900, 2020.
- [3] Aliaksandr Siarohin, Stephane Lathuiliere, Sergey Tulyakov, Elisa Ricci, Nicu Sebe: First Order Motion Model for Image Animation. arXiv:2003.00196v3 [cs.CV] 1 Oct 2020
- [4] Javier Galbally and Sebastien Marcel. Face anti-spoofing based on general image quality assessment. In The 22nd International Conference on Pattern Recognition, pages 1173–1178. IEEE, 2014.

- [5] Aliaksandr Siarohin, Stéphane Lathuilière, S. Tulyakov, E. Ricci, N. Sebe: Animating Arbitrary Objects via Deep Motion Transfer. 30 Aug 2019.
- [6] Ayush Tewari, Michael Zollhoefer, Florian Bernard, Pablo Garrido, Hyeonwoo Kim, Patrick Perez, and Christian Theobalt. High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42 (2):357–370, 2018.
- [7] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros: Everybody Dance Now, 27 February 2020
- [8] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. Generative adversarial networks for image and video synthesis: Algorithms and applications. Proceedings of the IEEE, 109(5):839–862, 2021.
- [9] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, V. Lempitsky: Few-Shot Adversarial Learning of Realistic Neural Talking Head Models, 27 February 2020.

