AI based Frame Interpolation Using OneVPL

[1]Madhumathi R, [2]Chandra Kumar N, [3]Vigneshwaran S, [4]Vignesh P,
Department. of Computer Science and Engineering,
Sri Ramakrishna Engineering College,
Coimbatore, Tamil Nadu, India
Email: chandrakumar.2101031@srec.ac.in

[5]Anbumani P
Visual Collaboration Technologies,
Bangalore, India
Email: anbumani@vcollab.com

2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-7024-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCCNT61001.2024.10725543

***Abstract*— The increasing popularity of multimedia content, particularly videos, has highlighted the significance of compression methods like HEVC and H.264, which effectively reduce file sizes without compromising quality. However, in the encoding and decoding process, hardware malfunctions or other issues may lead to frame loss, affecting frame rates and overall quality. To mitigate this problem, interpolation techniques are utilized to insert frames and maintain smooth playback. This method employs Artificial Intelligence (AI), specifically a trained Frame Interpolation for Large Motion (FILM) model, to generate intermediate frames by analyzing pairs of input images. This technique ensures seamless motion and enhances video quality despite frame loss during encoding and decoding. This work develops a pipeline using the FILM model and hardware acceleration to generate smooth videos. Testing demonstrates a 4x improvement in visual quality over raw input, setting a new standard for video interpolation and processing in multimedia applications.**

Keywords—Video Frame Interpolation, Artificial Intelligence, FILM Model, OneVPL, Frame Rate Enhancement, FFMPEG Integration.

I. Introduction

In today's digital landscape, multimedia content, especially videos, has become ubiquitous, experiencing a remarkable surge in popularity. To efficiently store and transmit these videos, compression formats like HEVC and h264 are extensively utilized, effectively reducing file sizes while preserving quality. However, the encoding and decoding process involved in these compression formats may lead to the loss of frames due to hardware issues or other factors, resulting in reduced frame rates (FPS) and diminished video quality. To address this challenge, interpolation techniques come into play, inserting additional frames within the video sequence to compensate for the lost ones. This work employs advanced artificial intelligence methodologies to generate these intermediate frames by analysing pairs of input images. Specifically, we leverage a Frame Interpolation for Large Motion model, meticulously trained to produce high-quality intermediate frames.

Furthermore, the performance of this method is optimized by harnessing the power of FFMPEG, configured for GPU processing. This configuration accelerates the generation of videos from the interpolated frames, ensuring optimized performance and enhancing the overall viewing experience. Through this innovative approach, this technique aim to alleviate the impact of lost frames and enhance the visual quality of digital videos.

II. Literature Survey

This literature review delves into the latest advancements in video processing and interpolation. The algorithm creates slow-motion videos from similar photos with significant scene motion [1]. It employs a scale-agnostic feature extractor and motion estimator, optimizing with a Gram matrix loss for clear frames. A streamlined single-network approach eliminates the need for external networks, excelling in large motion benchmarks and performing well in various datasets.

Ibaba et al explored advancements in video compression to balance computational efficiency with high-definition quality [2]. It concludes that artificial intelligence techniques offer the greatest promise for shaping future industry standards in compression as demand for high-resolution videos increases. Zhang et al used GANs to generate videos of any length, ensuring smooth transitions with alias-free operations and pre-existing knowledge [3]. By integrating the Temporal Shift Module (TSM), it improves dynamic consistency. A novel B-Spline-based motion representation guarantees temporal smoothness for infinite-length videos, while low-rank temporal modulation reduces repetitive content in longer sequences.

RIFE introduced a Real-time Intermediate Flow Estimation algorithm for Video Frame Interpolation (VFI). Unlike traditional methods, it employs IFNet, a neural network that directly predicts intermediate flows in a coarse-to-fine manner, improving speed and avoiding artifacts [4]. Through tailored distillation, RIFE achieves top-tier performance on public benchmarks, operating 4-27 times faster than Super Slomo and DAIN with superior outcomes.

Jiang et al introduced a convolutional neural network for end-to-end multi-frame video interpolation, combining motion interpretation and occlusion reasoning to improve quality across various lengths [5]. Zeng et al discussed the development of a video code based on FFmpeg, encompassing codec design, encoding, decoding processes, and performance evaluation. It discusses the integration of the code within FFmpeg, highlighting its contribution to video processing. Through rigorous testing and evaluation, the paper demonstrates the effectiveness of the codec implementation, providing insights into improving video compression and processing techniques [6].

Joy and Kounte explained an overview of traditional and recent trends in video processing, covering topics such as compression, enhancement, analysis, and emerging technologies [7]. It provide insights into the evolution of video processing techniques and their potential future directions. Kechiche introduced a real-time image and video processing method and architecture. It outlines techniques for efficient processing and their implementation in real-time systems. The study provides insights into practical applications and performance considerations in image and video processing [8]. Liu discussed video classification technology based on deep learning [9]. It explores methods and architectures for effective classification, highlighting advancements and applications across various domains.

Nidhi and Aggarwal reviewed recent advancements in assessment methods, focusing on techniques for evaluating and enhancing video quality [10]. The review provides insights into the state-of-the-art approaches and their applications in assessing the quality of video content. Lee introduced a video quality model considering compression, resolution, and frame parameters, accounting for space-time regularities [11]. It evaluates the impact of these factors on perceived quality, informing improvements in video processing techniques. Zhang et al explored methods for dynamically adjusting video resolution to optimize streaming quality, contributing to improved user satisfaction in video streaming applications [12]. It focuses on predicting the quality of user-generated videos efficiently.

Tu et al explored methods to accurately predict video quality, enabling better user experience in video-sharing platforms and multimedia applications [13]. The authors in [14] analyses texture variation to detect frame rate up-conversion artifacts in video processing. Bao et al presented MEMC-Net, a neural network that integrates Motion Estimation and Motion Compensation (MEMC) for video interpolation and enhancement [15].

III. EXISTING SYSTEM

Conventional approaches to frame interpolation typically rely on methods that assess the motion between successive frames and create additional frames based on this motion analysis. Below are some commonly used traditional techniques:

Motion Vector Interpolation: This technique gauges the motion vectors between consecutive frames through methods such as block matching or optical flow estimation. Subsequently, intermediate frames are synthesized by distorting and merging pixels from the original frames according to the motion vectors.

Optical Flow: Algorithms for optical flow ascertain the movement of pixels between successive frames by scrutinizing the intensity patterns. These algorithms produce a comprehensive flow field that delineates the motion of each pixel. Intermediate frames can then be crafted by distorting pixels from the original frames in line with the optical flow vectors.

Temporal Filtering: Temporal filtering methods employ linear or non-linear filters to interpolate pixels temporally. For instance, linear interpolation computes the pixel values for the intermediate frame as a weighted mean of the pixel values in the adjacent frames.

Spatio-Temporal Methods: These approaches amalgamate spatial and temporal information for interpolation. They might utilize methods such as block-based matching followed by spatial interpolation within the blocks.

Phase-Based Interpolation: This method operates in the frequency domain by decomposing input frames into distinct frequency bands using techniques like the Discrete Wavelet Transform (DWT) or Discrete Fourier Transform (DFT) Interpolation is then executed in the frequency domain before transformation back to the spatial domain.

Keyframe Interpolation: When the motion proves too intricate to accurately estimate between every frame, keyframes can be selected at intervals, and intermediate frames are generated by interpolating between these keyframes.

All methods encounter significant challenges with occlusions and disocclusions, where objects become covered or uncovered between frames. Handling non-rigid or complex motions, such as those in water or crowds, remains particularly difficult. Artifacts, including ghosting, blurring, and edge discontinuities, are common across these techniques. Additionally, many methods are computationally intensive, necessitating substantial processing power and making real-time applications challenging without optimization.

IV. PROPOSED SYSTEM

The objective of this work is to develop an AI-based pipeline that leverages the FILM model and OneVPL hardware acceleration to generate high-quality intermediate frames. This approach aims to enhance video quality, mitigate the impact of frame loss during encoding and decoding, and ensure smooth playback in multimedia applications. This proposed FILM model efficiently generates high-quality intermediate frames using scale-agnostic feature extraction and bidirectional flow estimation, utilizing OneVPL and NVENC for seamless, enhanced video playback.

The FILM model takes in two frames then it undergoes a series of convolutional paths to generate an intermediate frame. FILM model with its scale-agnostic feature extraction and bidirectional flow estimation, generates an accurate representation of the middle frame from the provided two frames. FILM model has a dynamic to adapt to high-quality image frames hence it can generate intermediate images for high-quality images. Fig.1 illustrates the working of FILM model.

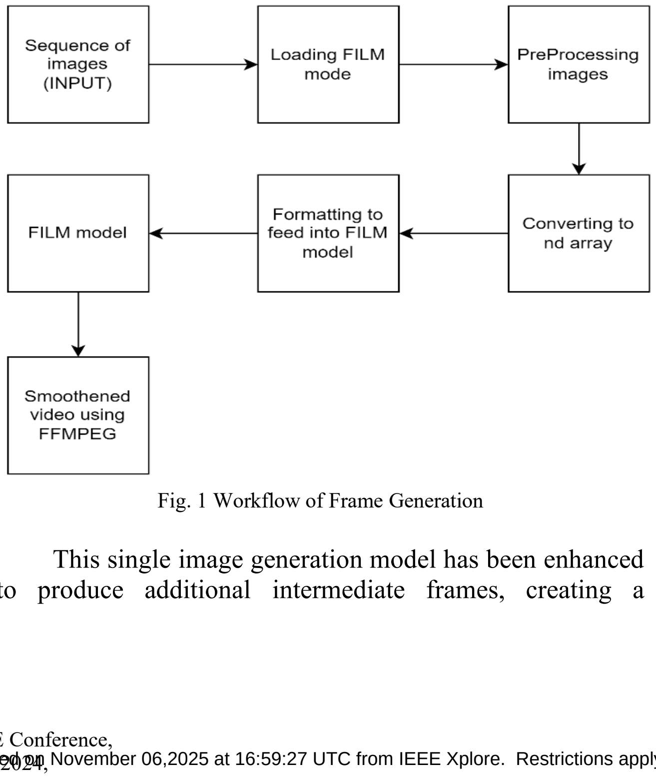

Fig. 1 Workflow of Frame Generation

This single image generation model has been enhanced to produce additional intermediate frames, creating a continuous sequence of images. These interpolated and original images are then seamlessly stitched together using FFMPEG, which is configured with hardware accelerators OneVPL for Intel hardware and NVENC for Nvidia hardware. This process ensures smooth video playback and improved visual fidelity in multimedia applications.

The workflow involves storing sequences of images in an input buffer of fixed size. Each image in the buffer is preprocessed and stored as an nd-array. Two consecutive frames from the buffer are fed into the FILM model, which generates an intermediate frame. Both the original and interpolated frames are stored in the output buffer.

After generating intermediate frames for all input images, the output buffer is used to store the generated images in the I/O system. FFMPEG then utilizes these frames to create a video with a higher frame rate, effectively doubling the frame rate by incorporating the interpolated images.

To achieve efficient parallelism in frame interpolation algorithms, we utilized the OneVPL hardware acceleration driver programming mode. This involved compiling OneVPL for FFMPEG and integrating OneVPL's hardware acceleration into the FFMPEG framework. As a result, FFMPEG can seamlessly leverage OneVPL's capabilities for enhanced video processing.

By employing this approach, we ensured that the video generation process is both efficient and high-quality, making full use of hardware acceleration for optimal performance.

1. **Model Loading and Image Loading**: Load the FILM model's weights into TensorFlow via TensorFlow Hub for inference. Load a series of input images using Input-Output (IO) packages, ensuring continuous loading of image pairs for interpolation.
2. **Image Processing**: Convert the loaded images into nd-arrays for processing with the model. Transform each image into an nd-array tensor to serve as input for the model. Create a nd-array for the time dimension to specify the degree of intermediateness for each image, with a value of "0.5" representing the perfect intermediate frame.
3. **Model Input Construction**: Construct an input structure in the format {time, x0, x1}, where "time" denotes the time nd-array, "x0" represents the first input image, and "x1" represents the second input image.
4. **Model Inference**: Feed the constructed input structure into the FILM model to generate an intermediate frame as output for the provided pair of input images.
5. **Intermediate Frame Storage**: Store the generated intermediate frames as image files in the specified directory.
6. **Video Creation**: Utilize FFmpeg to create a video from the series of images, including those generated by the model. Configure FFmpeg to support GPU acceleration for faster processing. Produce a video with a frame rate twice that of the original video or the video that could be generated with the provided images.
7. **Result Analysis**: Verify that the resulting video comprises additional interpolated frames, resulting in smoother playback without any lag.

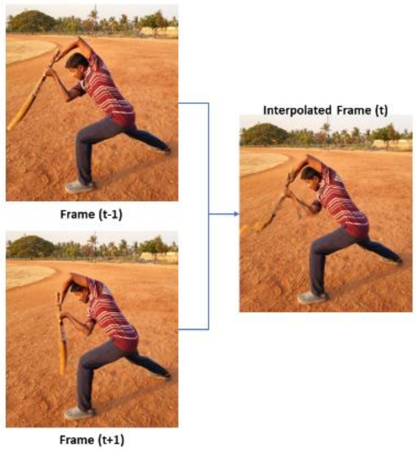

Fig. 2. Single Frame Interpolated using FILM model

The FILM model takes in two images/Frames and generates an intermediate image for those two frames. This is shown in fig. 2, where it generates high-quality intermediate images.

In fig.3 Buffer-based Frame Generation Model called FILM enhances image generation by dynamically adjusting parameters based on contextual information stored in a buffer, resulting in more coherent and realistic frames.

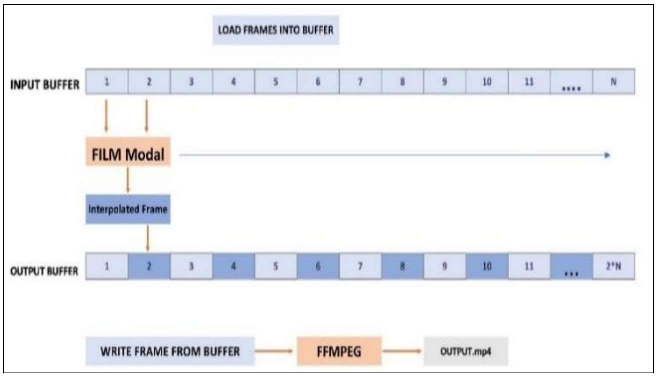

Fig. 3 Buffer-based Frame Generation Model (FILM)

Fig.4 depicts the overall flow diagram of the proposed system that uses a web app to interface with the backend (FILM) that takes sequences of images as input and generates a smoothened video as output.

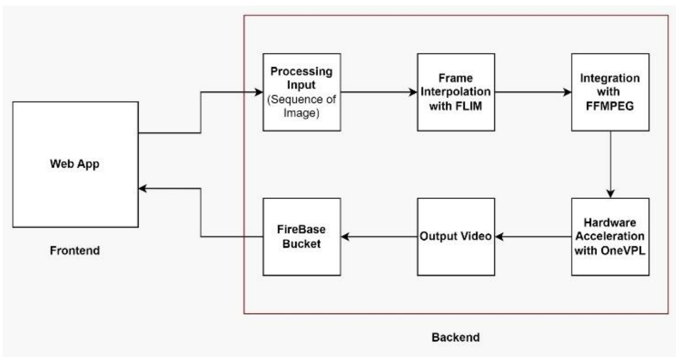

Fig. 4. System Flow Diagram

After successful generation of video, it is uploaded to a firebase bucket, providing a centralized storage solution. UI fetches the video from this location, ensuring seamless integration between processing and presentation. This approach enables efficient storage and retrieval of videos, enhancing user experience by delivering smooth playback directly within the application interface, without the need for additional storage or processing on the client side.

V. RESULTS

Fig.5 illustrates the difference in frame rates achieved. This depicts that there's a reduction of video duration in the smoothened video(output), since the FPS is increased with interpolated frames, maintaining constant FPS throughout the generated video.

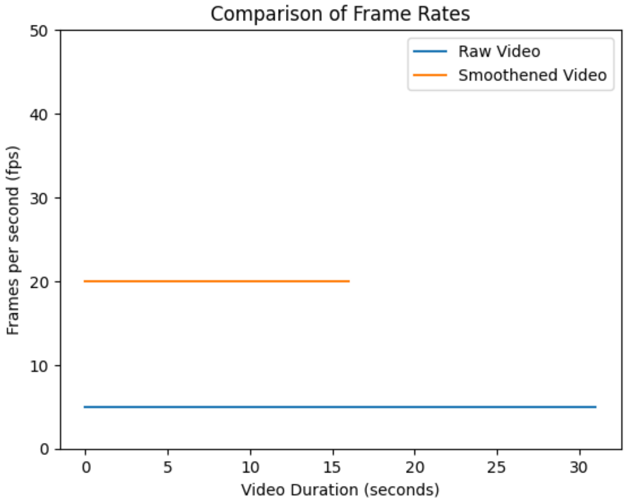

Fig. 5. Comparison of raw input with generated output

Fig.6 shows the user interface to upload images, and the displays interpolated video from backend.

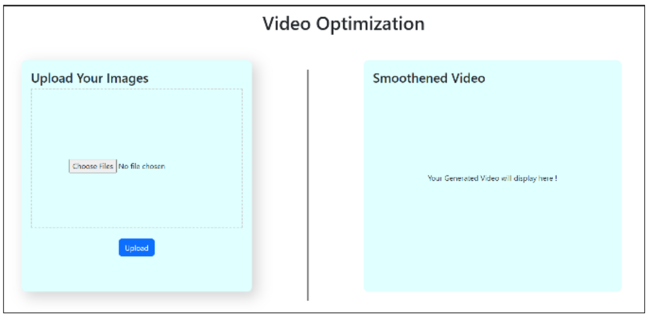

Fig.6. User-Interface for Generating Videos

Fig.7 shows the media properties of the raw video which is of 5 FPS.

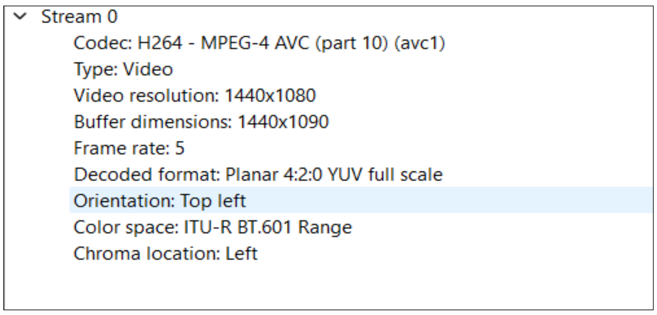

Fig. 7. Media properties of Video Generated from Original Images

Fig.8 describes the media properties of the video that is generated from using interpolation where the original 150 frames were interpolated to form about 300 frames. Then again 300 frames were interpolated to form about 600 frames, a 30-second video, which is a 20 FPS video.

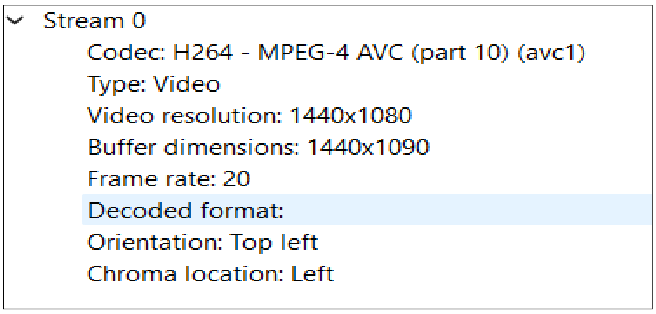

Fig. 8. Media Properties of Video Generated using the FILM Model

The FILM model with OneVPL hardware acceleration generated high-quality interpolated images that maintain the flow of the video. With this technique, we achieved a 4X improvement in frame rate where 5 FPS video is interpolated to form 20 FPS video.

VI. CONCLUSION AND FUTURE ENHANCEMENT

This work implements a pipeline using the FILM model and hardware acceleration to generate smooth videos with improved visual quality. By interpolating intermediate frames and optimizing for hardware acceleration, we achieve enhanced video continuity and fidelity. Testing revealed our generated smooth video to be four times better than the raw input. This demonstrates the effectiveness of this approach in producing high-quality videos for multimedia applications, providing seamless playback and setting a new standard for video interpolation and processing.

A potential future enhancement for the project could involve integrating advanced motion estimation algorithms into FILM. These algorithms, like optical flow estimation or deep learning-based motion prediction, would improve frame interpolation accuracy, especially in scenes with complex motion. Additionally, incorporating adaptive frame rate control mechanisms would dynamically adjust frame rates based on scene complexity, optimizing performance and resource usage. These enhancements would elevate FILM's capabilities, ensuring smoother motion and higher-quality video output across various scenarios

REFERENCES

[1] F. Reda, J. Kontkanen, E. Tabellion, D. Sun, C. Pantofaru, and B. Curless, "FILM: Frame Interpolation for Large Motion", European Conference on Computer Vision, Springer, pp. 250-266, 2022.

[2] A. Ibaba, S. Adeshina, and A. M. Aibinu, "A Review of Video Compression Optimization Techniques", IEEE International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), pp. 1-5, 2021.

[3] Q. Zhang, C. Yang, Y. Shen, Y. Xu, and B. Zhou, "Towards Smooth Video Composition", International Conference on Learning Representation (ICLR), 2022.

[4] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou, "RIFE: Real-Time Intermediate Flow Estimation for Video Frame Interpolation", Journal of Computer Vision and Pattern Recognition, pp. 1-22, 2022.

[5] H. Jiang, D. Sun, V. Jampani, M. H. Yang, E. Learned-Miller, and J. Kautz, "Super SloMo: High-Quality Estimation of Multiple

Intermediate Frames for Video Interpolation", IEEE International Conference on Computer Vision and Pattern Recognition, pp. 9000-9008, 2018.

[6] H. Zeng, Z. Zhang, and L. Shi, "Research and Implementation of Video Codec Based on FFmpeg", IEEE International Conference on Network and Information Systems for Computers (ICNISC), pp. 184-188, 2016.

[7] H. K. Joy, and M. R. Kounte, "An Overview of Traditional and Recent Trends in Video Processing", IEEE International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 848-851, 2019.

[8] L. Kechiche, L. Touil, and B. Ouni, "Real-time image and video processing: Method and architecture", 2nd IEEE International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 194-199, 2016.

[9] M. Liu, "Video Classification Technology Based on Deep Learning", IEEE International Conference on Information Science, Parallel and Distributed Systems (ISPDS), pp. 154-157, 2020.

[10] Nidhi, and N. Aggarwal, "A review on Video Quality Assessment", Recent Advances in Engineering and Computational Sciences (RAECS), IEEE Access, pp. 1-6, 2014.

[11] D. Y. Lee, J. Kim, H. Ko, and A. C. Bovik, "Video Quality Model of Compression, Resolution, and Frame Based on Space-Time Regularities", IEEE Transactions on Image Processing, vol. 31, pp. 3644-3656, 2022.

[12] Y. Zhang et al, "Improving Quality of Experience by Adaptive Video Streaming with Super-Resolution", IEEE Conference on Computer Communications, pp. 1957-1966, 2020.

[13] Z. Tu, C. J. Chen, Y. Wang, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "Efficient User-Generated Video Quality Prediction", IEEE Picture Coding Symposium (PCS), pp. 1-5, 2021.

[14] G. Yang, M. Xia, L. Li, R. Li, and X. Sun, "Detecting video frame rate up-conversion based on frame-level analysis of average texture variation", Multimedia Tools and Applications, vol. 76, pp. 8399-8421, 2017.

[15] W. Bao, W. S. Lai, C. Ma, X. Zhang, Z. Gao, and M. H. Yang, "MEMC-Net: Motion Estimation and Motion Compensation Driven Neural Network for Video Interpolation and Enhancement", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, pp. 933-948, 2021.