**IJARCCE** ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13334

# Frame Interpolation Using FILM

**Dhruva Kumar Shetty[1], Pramith A Naiga[2], Sidhvin P Shetty[3], Yash Karunakar Shetty[4], Dr. Maryjo M George[5]**

Student, Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodabidre, India[1,2,3,4]

Professor, Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodabidre, India[5]

**Abstract**: Frame interpolation is a computational technique used in video processing to create additional frames between existing frames, thereby enhancing the smoothness and visual quality of motion in the video.Existing techniques for frame interpolation in videos include path-based and phase-based conventional methods, convolutional neural network (CNN)-based flow-based methods, kernel-based methods utilizing convolution operations over local patches, and recent advancements such as deformable convolution-based approaches like AdaCoF. Addressing the challenges in frame interpolation is essential for developing more efficient and versatile techniques applicable across various platforms and applications. High computational costs, particularly prevalent in methods reliant on deep neural networks (DNNs), hinder deployment on resource-constrained devices or real-time applications. Complexity arises from intricate model architectures or multi-stage processes, complicating both understanding and implementation. Additionally, limited generalization restricts the practical utility of certain techniques, as they may excel on specific datasets but struggle with diverse or unseen data. Methods relying solely on pixel-wise information or local kernels may falter in accurately interpolating frames with complex motion, occlusion, or fine details. Furthermore, the large size of state-of-the-art models poses challenges for storage, training, and deployment, especially on mobile or embedded devices. Addressing these issues is paramount for advancing frame interpolation methods towards greater efficiency, practicality, and applicability across a broad spectrum of contexts and platforms. Our Compression-Driven Framework for Video Interpolation (CDFI) addresses key challenges as follows: Reduced Computational Cost, Simplicity and Efficiency, Improved Generalization, Enhanced Motion Handling, Compact Model Size.

**Keywords:** Video frame interpolation, Optical flow-based, Real-time solutions, Visual quality, Real-time applications, Stakeholders, Video processing.

## I. INTRODUCTION

Frame interpolation, the generation of intermediary frames between two input frames, plays a pivotal role in applications ranging from temporal up-sampling to the creation of slow-motion videos. Recently, a novel application has emerged within the realm of digital photography, where the prevalent practice of capturing near-duplicate photos presents an exciting opportunity. The ability to interpolate between these closely timed images offers the potential to craft captivating videos that not only reveal intricate scene dynamics but also provide a visually enhanced representation of the captured moment.

However, the challenge arises when dealing with near-duplicates, as the temporal spacing between frames can be significant, resulting in substantial scene motion. Existing frame interpolation methods, which excel in scenarios with small temporal gaps, struggle to address this unique challenge. Previous attempts to handle large motion involved training on extreme motion datasets, yet these approaches exhibit limitations when confronted with scenarios featuring smaller motion.

This research introduces a pioneering approach to frame interpolation, offering a network that adeptly handles both small and large motion. The methodology involves adapting a multi-scale feature extractor with weight sharing across scales and implementing a "scale-agnostic" bidirectional motion estimation module. By intuitively recognizing similarities between large motion at finer scales and small motion at coarser scales, this approach maximizes pixel availability for supervising large motion. The findings underscore the effectiveness of this technique in handling diverse motion scenarios, presenting a promising solution for advancing frame interpolation capabilities.

## II. LITERATURE SURVEY

In [1] This research introduces a novel module for improving video frame interpolation (VFI) by efficiently extracting both motion and appearance information in a unified manner. By rethinking the information processing in inter-frame

 

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13334

attention, the proposed module utilizes attention maps for enhancing appearance features and extracting motion information. Integrated into a hybrid CNN and Transformer architecture, this approach achieves state-of-the-art performance in VFI, demonstrating better efficiency and lower computational complexity compared to similar models. This approach requires additional modules, such as cost volume to extract motion information, which often imposes a high computational overhead.

In [2] ,This paper introduces a new method for interpolating frames in 4K videos using a technique called the bilateral transformer (Bi-Former). The process involves three main steps: predicting global motion using Bi-Former, refining the motion fields with block-wise bilateral cost volumes (BBCVs), and synthesizing an intermediate frame by warping and blending input frames. The proposed Bi-Former algorithm is shown to achieve excellent interpolation performance on 4K datasets in extensive experiments. The algorithm's requirement for significant computational resources and longer processing times, especially in the up-sampler and synthesis network, may impact its practical applicability, particularly in scenarios where real-time performance or resource efficiency is crucial.

In [3], Video frame interpolation poses challenges in research, often tackled through deep learning. Existing methods focus on relevant information for each output pixel but struggle with degrees of freedom limitations and handling real-world complex motions. To address this, we propose the Adaptive Collaboration of Flows (AdaCoF) warping module, estimating kernel weights and offset vectors for target pixels. While AdaCoF is versatile, some shortcomings include the need for fully convolutional network parameters and the absence of explicit discussion on potential failure cases or scenarios where AdaCoF might perform sub-optimally. Additionally, the proposed dual-frame adversarial loss lacks detailed examination, making it challenging to fully assess its effectiveness. Despite these limitations, experimental results showcase AdaCoF's superiority in various challenging scenarios, emphasizing its potential in video frame interpolation tasks.

In [4], The research paper introduces a blurry video frame interpolation method that aims to reduce motion blur and simultaneously upscale frame rates. The proposed approach involves a pyramid module for synthesizing clear intermediate frames cyclically, offering adjustable spatial receptive fields and temporal scope. An inter-pyramid recurrent module is introduced to connect sequential models and exploit temporal relationships, enabling the iterative synthesis of temporally smooth results without significantly increasing the model size. Experimental results demonstrate the effectiveness of the method against state-of-the-art methods in terms of PSNR and SSIM metrics. The model's contributions are further analyzed through experiments on architecture scalability, inter-pyramid recurrent module effectiveness, ConvLSTM unit impact, and the importance of cycle consistency loss. While the paper touches on runtime costs for different scales, a more detailed discussion on the computational efficiency, especially in real-time applications, would enhance the practical applicability assessment.

## III. SCOPE AND METHODOLOGY

**Aim of the project**

The aim of the project is to create a web application using Flask and Streamlit for image interpolation using the FILM (Feature Pyramid and Image-based Learning for Motion Estimation) model. Users can upload two images, and the application will interpolate frames between them to generate a smooth transition video. The Flask application serves as the backend for uploading images and processing the interpolation, while Streamlit is intended to provide a user-friendly interface for running the interpolation process.

**Exisiting system**

Frame interpolation methods encompass a range of techniques aimed at generating intermediate frames between consecutive frames in a video sequence. Optical flow-based methods estimate motion using algorithms like Lucas-Kanade or Horn-Schunck, while deep learning-based approaches leverage convolutional neural networks (CNNs) trained on large datasets for high-quality results. Physics-based methods model scene dynamics, and hybrid approaches combine traditional and deep learning techniques to capitalize on their respective strengths. Application-specific implementations cater to specific domains like sports broadcasting or medical imaging, and real-time solutions focus on efficient, low-latency processing for live video interpolation.

**Proposed system**

Adapt a multi-scale feature extractor that shares weights, and propose a scale-agnostic bi-directional motion estimator to handle both small and large motion well, using regular training frames and to adopt a Gram matrix-based loss function to in-paint large disocclusions caused by large scene motion, leading to crisp and pleasing frames. The project aims to propose a unified, single-stage architecture, to simplify the training process and remove the reliance on additional optical flow or depth networks.

 

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13334

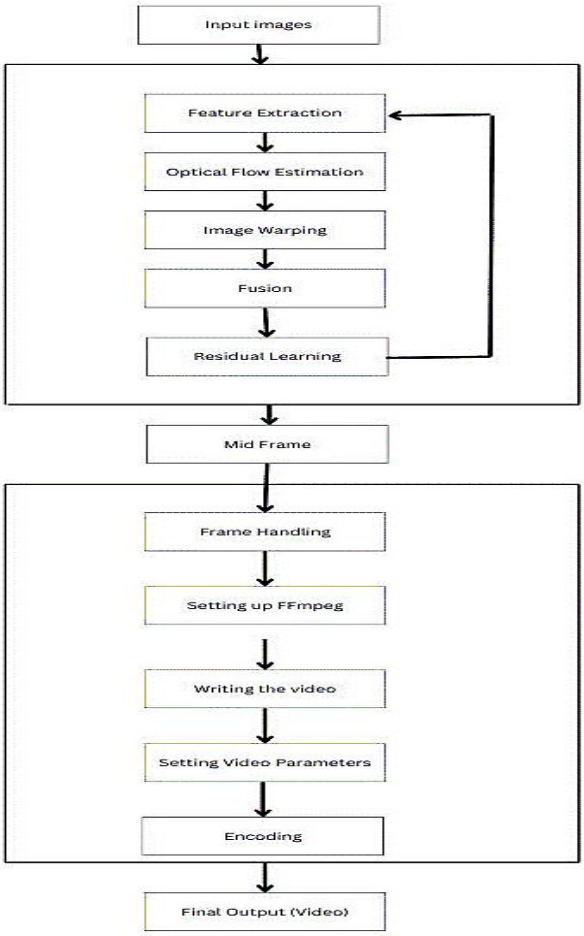


Fig 1. Proposed system

**Architecture Diagram**

An architectural explanation is a formal description and illustration of a system, organized in a manner that supports reason in relation to the structure of the system which comprises system components, the externally detectable properties of individual components, the interaction among them, and provides a plan from which products can be procured, and systems developed, that will work mutually to implement the on the whole as a system.

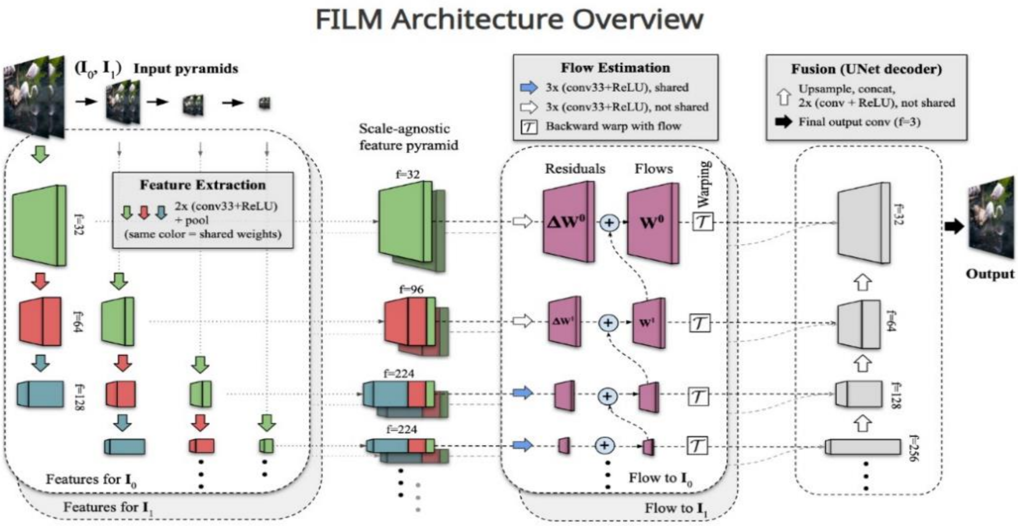


Fig 2. Architecture Diagram

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13334

## IV. CONCLUSION

In summary, the project aims to advance the field of video frame interpolation by integrating various methodologies, including optical flow-based, deep learning-based, physics-based, hybrid approaches, application-specific implementations, and real-time solutions. By harnessing these diverse techniques, the project seeks to enhance the quality and efficiency of generating intermediate frames in video sequences. Through the development of user-friendly interfaces and the incorporation of cutting-edge technologies, such as deep learning, the project endeavors to make frame interpolation more accessible and effective for researchers, filmmakers, broadcasters, and other stakeholders. Overall, this endeavor represents a significant step forward in video processing, promising to improve visual quality, enable real-time applications, and support various domains requiring frame interpolation.

## REFERENCES

[1] Extracting Motion and Appearance via Inter-Frame Attention for Efficient Video Frame Interpolation:Guozhen Zhang1 Yuhan Zhu1 Haonan Wang1 Youxin Chen3 Gangshan Wu1 Limin Wang1, 2, *1State Key Laboratory for Novel Software Technology, Nanjing University, China2Shanghai AI Lab, China 3Samsung Electronics (China) R&D Centre, China
[2] BiFormer: Learning Bilateral Motion Estimation via Bilateral Transformer for 4K Video Frame Interpolation Junheum Park Korea University jhpark@mcl.korea.ac.kr Jintae Kim Korea University jtkim@mcl.korea.ac.kr Chang-Su Kim* Korea University changsukim@korea.ac.kr
[3] AdaCoF: Adaptive Collaboration of Flows for Video Frame Interpolation HyeongminLee1 TaeohKim1 Tae-youngChung1 DaehyunPak1 YuseokBan2 SangyounLee1∗
[4] Blurry Video Frame Interpolation Wang Shen Wenbo Bao Guangtao Zhai∗ Li Chen Xiongkuo Min Zhiyong Gao Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
[5] Time Lens: Event-based Video Frame Interpolation Stepan Tulyakov∗∗,1 Daniel Gehrig∗,2 Stamatios Georgoulis1 Julius Erbach1 Mathias Gehrig2 Yuanyou Li1 Davide Scaramuzza2 1Huawei Technologies, Zurich Research Center2Dept.ofInformatics,Univ.ofZurichandDept.ofNeuroinformatics,Univ.ofZurichandETHZurich
[6] Deep Video Frame Interpolation Using Cyclic Frame Generation Yu-Lun Liu,1,2,3 Yi-Tung Liao,1,2 Yen-Yu Lin,1 Yung-Yu Chuang1,2
[7] Video Frame Interpolation via Deformable Separable Convolution Xianhang Cheng,1 Zhenzhong Chen1∗ 1School of Remote Sensing and Information Engineering, Wuhan University, China
[8] Phase Net for Video Frame Interpolation SimoneMeyer1,2 AbdelazizDjelouah2 BrianMcWilliams2 AlexanderSorkine-Hornung2∗ MarkusGross1,2 ChristopherSchroers2
[9] Phase-BasedFrameInterpolationforVideo SimoneMeyer1 OliverWang2 HenningZimmer2 MaxGrosse2 AlexanderSorkine-Hornung2 1ETHZurich 2DisneyResearchZurich
[10] Frame interpolation in video stream using optical flow methods S Lyasheva1, R Rakhmankulov1 and M Shleymovich1

 