A DEEP LEARNING BASED TRANSFORMATIVE VIDEO RESTORATION & COMPRESSION PIPELINE

Munganda Sai Niketh

20951A04F5

A DEEP LEARNING BASED TRANSFORMATIVE VIDEO RESTORATION & COMPRESSION PIPELINE

A Project Report

submitted in partial fulfillment of the

requirements for the award of the degree of

Bachelor of Technology

in

Electronics and Communication Engineering

By

Munganda Sai Niketh

20951A04F5

Under the supervision of

Mr. Mohammad Khadir

Assistant professor

Department of Electronics and Communication Engineering

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad – 500 043, Telangana

MAY, 2024

© 2024, M Sai Niketh. All rights reserved.

DECLARATION

I certify that

a. The work contained in this report is original and has been done by us under the guidance of our supervisor(s).

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in preparing the report.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources whenever necessary.

Place: **Signature of Student:**

Date: **Roll No: 20951A04F5**

CERTIFICATE

This is to certify that the project report entitled **A DEEP LEARNING BASED TRANSFORMATIVE VIDEO RESTORATION & COMPRESSION PIPELINE** submitted **by Mr. M Sai Niketh(20951A04F5)** to the Institute of Aeronautical Engineering, Hyderabad in partial fulfillment of the requirements for the award of the Degree Bachelor of Technology in Branch **Electronics and Communication Engineering** is a bonafide record of work carried out by them under the guidance and supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor	**Head of the Department**
Mr.Mohammad Khadir	Dr. P Munaswamy
Assistant Professor	

Date:

APPROVAL SHEET

This project report entitled **A DEEP LEARNING BASED TRANSFORMATIVE VIDEO RESTORATION & COMPRESSION PIPELINE by Mr. M Sai Niketh** is approved for the award of the Degree Bachelor of Technology in **Branch Electronics and Communication Engineering.**

Examiners

Supervisor (s)

Mr. Mohammad Khadir

Assistant Professor

Principal

Dr. L V Narasimha Prasad

Date:

Place:

ACKNOWLEDGEMENT

I would like to take this opportunity to express our deepest gratitude to all those who have supported, encouraged, motivated, and extended cooperation in various ways during the implementation of our project. I am honored to acknowledge the help of all those responsible for planning the success of our project.

I would like to thank **Dr. P MUNASWAMY, Head of the Department of Electronics and Communication Engineering, Mr. Mohammad Khadir, Assistant Professor of Electronics and Communication Engineering** for their valuable advice and assistance throughout the development of this project. Without their direction, cooperation, and encouragement, this project would not have been possible.

In particular, I want to thank my friends for helping me create this project. And, I want to express my gratitude to my principal, **Dr. L V NARASIMHA PRASAD**, for always offering to help in all circumstances.

ABSTRACT

Video processing plays a pivotal role across various domains such as entertainment, surveillance, and communication. In this study, we introduce a comprehensive framework tailored for video restoration and compression, with the primary objective of enriching video quality while curbing bitrate requirements. The proposed framework comprises a sequence of restoration models, encompassing super-resolution, deblurring, denoising, and frame interpolation, followed by a compression model. Initially, the input video frames undergo compression via the libx265 codec to alleviate bitrate and storage needs. Subsequently, the compressed frames are inputted into a cascade of restoration models to rectify the diminished video quality induced by compression artifacts like noise, blur, and loss of detail. The restoration pipeline incorporates cutting-edge models such as VRT (Video Restoration Transformer), leveraging deep learning methodologies to attain superior performance in video restoration endeavors. The super-resolution model amplifies the resolution of low-resolution frames acquired post-compression, whereas the deblurring model rectifies motion blur and other spatial distortions. The denoising model further diminishes noise introduced during compression, resulting in cleaner and sharper video frames. Finally, the frame interpolation model generates intermediate frames to heighten temporal smoothness and fluidity in the video sequence. Experimental findings substantiate the efficacy of the proposed framework in enhancing video quality and mitigating compression artifacts, thereby yielding substantial enhancements in perceptual quality and fidelity. The real-time processing capabilities of the framework render it suitable for a diverse array of applications, encompassing video streaming, surveillance, and digital cinema.

Keywords: Video processing, Video restoration, Video compression, Super resolution, Deblurring, Denoising, Frame interpolation, Deep learning, Compression artifacts.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LQ	Low quality
LR	Low resolution
HR	High resolution
SR	Super-resolution
VRT	Video restoration transformer
TMSA	Temporal mutual self-attention
PW	Parallel warping
GT	Ground truth
PSNR	Peak signal-to-noise ratio
SSIM	Structural Similarity Index measure
MS-SSIM	Multiscale structural similarity index measure
PSNR_Y	Y (Luminance channel)
SSIM_Y	Y (Luminance channel)
CNN	Convolutional neural network
RNN	Recurrent neural network
FPS	Frames per second
GPU	Graphics processing unit
FPGA	Field-programmable gate array
QR	Query matrix reference frame
KR	Key matrix reference frame
VR	Value matrix reference frame
A	Attention map computed using softmax function
XR	Feature matrix reference frame
PQ	Projection matrix for the query

PK	Projection matrix for the key
PV	Projection matrix for the value
SoftMax	Compute attention weights
MA	Mutual self-attention
Z	Output MLP layer
X	Input feature matrix
W1, W2	Weight matrix
b1, b2	Bias vector
RELU	Rectified linear unit activation function

CHAPTER 1

INTRODUCTION

In recent times, advancements in deep learning have brought about a transformative shift in the realm of image and video restoration. These advancements are geared towards improving the visual fidelity of images and videos by tackling issues such as noise, blur, and low resolution. The capability to rejuvenate deteriorated visual content holds significant promise across various domains, including surveillance, entertainment, healthcare, and remote sensing. Within this document, we provide an overview of our project concerning video restoration through the utilization of deep learning techniques, with a particular focus on the impact of compression on video processing.

Video compression stands as a crucial element within contemporary video processing workflows, facilitating efficient storage and transmission of video content. Nonetheless, compression algorithms often introduce artifacts and degrade quality, particularly at higher compression ratios. Common artifacts resulting from compression include blockiness, ringing, and loss of fine details, all of which can substantially impair the perceptual quality of compressed videos.

The degradation stemming from compression artifacts presents a significant hurdle for subsequent video processing endeavors, such as restoration. When implementing restoration techniques on compressed videos, it is imperative to consider the interaction between compression artifacts and restoration algorithms, along with their impact on the final output quality. For instance, artifacts like blockiness and ringing can disrupt motion estimation and compensation in tasks like deblurring and frame interpolation, resulting in less-than-optimal outcomes.

Conversely, restoration techniques can effectively alleviate the effects of compression artifacts and heighten the visual fidelity of compressed videos. Through the utilization of deep learning-based methodologies, it becomes feasible to learn intricate mappings between compressed and uncompressed video frames, thereby enabling precise artifact removal and detail enhancement. Furthermore, restoration techniques can exploit temporal redundancy within video sequences to bolster the overall perceptual quality and fidelity of compressed videos.

Our project revolves around a comprehensive pipeline for video restoration, encompassing several pivotal components: super-resolution, deblurring, denoising, and frame interpolation. Each component assumes a critical role in restoring various aspects of video quality, ultimately culminating in visually pleasing and high-quality output. By integrating these restoration techniques into a cohesive framework, our objective is to tackle the challenges posed by compression artifacts and restore the original visual content of compressed videos.

To summarize, compression artifacts stemming from video compression algorithms can diminish the visual quality of compressed videos and present hurdles for subsequent video processing tasks. Nevertheless, deep learning-driven restoration techniques hold promise in mitigating these artifacts and improving the perceptual quality of compressed videos. In the forthcoming sections, we will delve into the implementation specifics and experimental findings of our video restoration pipeline. We'll emphasize the influence of compression on video processing and the efficacy of our restoration techniques in combatting compression artifacts.

1.1 Video Compression:

Video compression involves reducing the size of video files to facilitate storage and transmission by eliminating redundant or unnecessary information from the video signal. Common compression algorithms like H.264 or HEVC aim to achieve high compression ratios while minimizing quality loss. However, compression may introduce artifacts and degrade video quality, particularly at higher compression ratios.

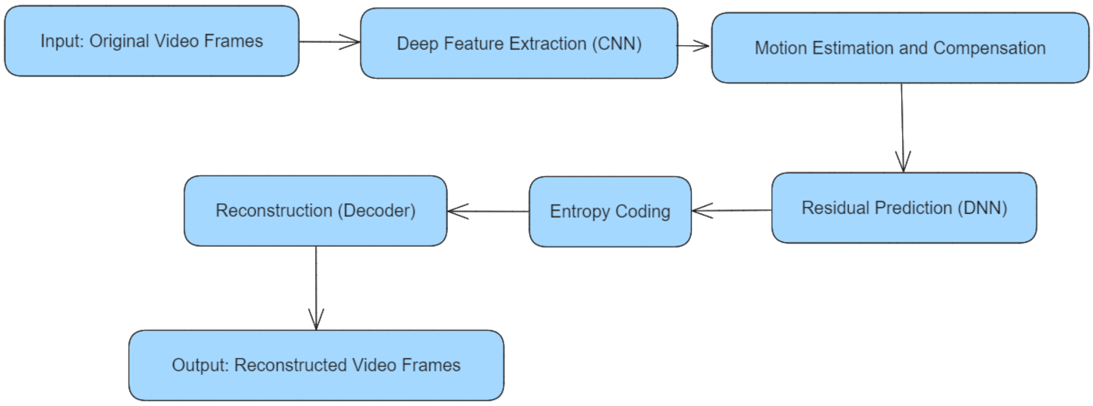


Fig 1.1:- Block diagram of traditional video compression.

1.2 Video Restoration:

Video restoration entails improving the visual quality of degraded or damaged video content through tasks such as denoising, deblurring, super-resolution, and frame interpolation. Deep learning-based restoration methods have gained traction due to their ability to grasp intricate mappings between degraded and pristine video frames, enabling effective artifact removal and detail enhancement.

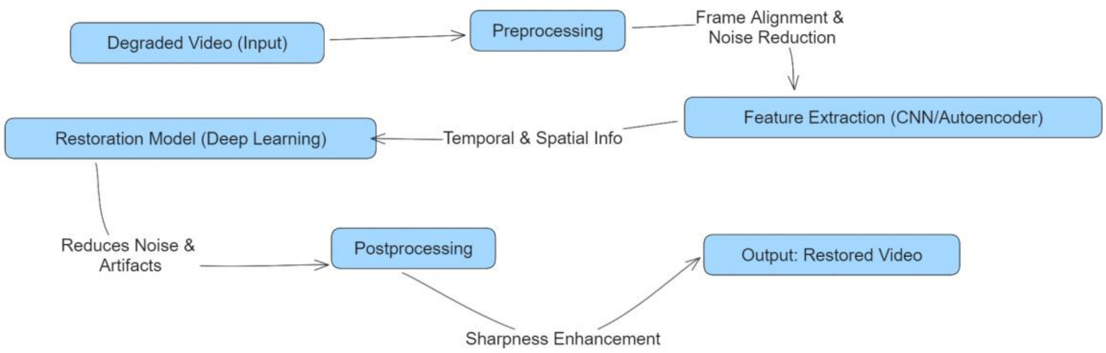


Fig 1.2:- Block diagram of traditional video restoration.

1.3 Compression Artifacts:

Compression artifacts are distortions or flaws induced by video compression algorithms, including blockiness, ringing, and loss of fine details. These artifacts can significantly impair the perceptual quality of compressed videos and present challenges for subsequent video processing, such as restoration. Understanding compression artifacts is crucial for devising restoration techniques that can counteract their effects and restore the original visual content.

1.4 Deep Learning-Based Restoration Techniques:

Deep learning-based restoration approaches utilize neural networks to learn intricate mappings between input and output video frames. By training on extensive datasets of both pristine and degraded video sequences, these techniques can efficiently eliminate compression artifacts and enhance the visual quality of compressed videos. Components such as super-resolution, deblurring, denoising, and frame interpolation are commonly integrated into deep learning-based restoration pipelines to address various aspects of video degradation. These methods offer promising avenues for mitigating compression artifacts and reinstating the original visual content of compressed videos.

CHAPTER 2

LITERATURE REVIEW

The proposed method for image compression employs convolutional autoencoders to learn condensed representations of input images, aiming to reduce storage and transmission requirements while preserving visual quality [1]. Through the training of convolutional autoencoder networks, this technique achieves efficient image compression by utilizing advanced encoding and decoding architectures based on convolutional neural networks (CNNs). This strategy presents a promising avenue for decreasing the size of image data for diverse applications, such as storage and transmission, by harnessing the capabilities of deep learning in image compression.

An end-to-end optimized image compression approach based on frequency-oriented transforms is introduced, which merges frequency-oriented transforms with deep learning techniques to attain efficient image compression [2]. By optimizing the compression process end-to-end, this method strives for superior compression performance while upholding visual quality. It showcases promising outcomes in reducing the size of image data for storage and transmission applications, underscoring the potential of amalgamating frequency-oriented transforms with deep learning for image compression.

The assessment of video quality impacted by H.265/HEVC compression, is crucial for multimedia services. It examines two Full HD resolution sequence types regarding compression's subjective effects [3]. Initially, the H.265/HEVC compression standard is briefly outlined. Subsequently, subjective video quality methods used in experiments are detailed. Experimental results indicate that video quality remains sufficient between 5 and 7 Mbps, suggesting providers needn't exceed this threshold for streaming. These findings contribute to the development of a predictive model for video quality in IP-based networks.

A deep image compression method based on end-to-end learning is proposed, circumventing the requirement for manually crafted compression algorithms by directly learning the compression process from data [4]. Through end-to-end optimization of the compression process, this method achieves superior compression performance while preserving visual quality. It provides a promising alternative to conventional image compression methods by leveraging deep learning to learn efficient compression representations directly from image data.

The real-time single image and video super-resolution method employs an efficient sub-pixel convolutional neural network to enhance the resolution of images and videos in real time [5]. By learning the mapping between low-resolution and high-resolution image patches, this method achieves notable improvements in image and video resolution without compromising computational efficiency. It demonstrates promising results in real-time applications where enhanced image and video resolution is desirable, showcasing the potential of sub-pixel convolutional neural networks for real-time super-resolution tasks.

The SwinIR image restoration method employs the Swin Transformer architecture to restore degraded images, achieving state-of-the-art performance on various restoration tasks such as denoising, deblurring, and super-resolution [6]. By leveraging self-attention mechanisms and hierarchical feature representations, the Swin Transformer architecture exhibits superior capabilities in capturing long-range dependencies and contextual information, resulting in improved restoration outcomes compared to conventional methods. This approach provides a promising solution for image restoration tasks, where high-quality results are essential for diverse applications including photography, medical imaging, and remote sensing.

BasicVSR++ extends the BasicVSR framework by introducing advanced propagation and alignment mechanisms, resulting in enhanced performance for video super-resolution [7]. Through the integration of sophisticated propagation and alignment strategies, BasicVSR++ achieves superior outcomes in reconstructing high-resolution frames from low-resolution video sequences. This advancement showcases significant improvements in video super-resolution quality, rendering it suitable for applications like video enhancement and content creation, where high-quality video output is crucial.

Cascaded deep video deblurring harnesses temporal sharpness priors to enhance video deblurring performance, utilizing cascaded convolutional neural networks to iteratively refine deblurring outcomes [8]. By capitalizing on temporal coherence in video sequences, this method achieves superior deblurring quality compared to single-image deblurring approaches. It demonstrates promising results in restoring sharpness to blurred video sequences, making it applicable for applications such as video surveillance, video editing, and video enhancement.

Deep video deblurring for hand-held cameras introduces a deep learning-based approach to address blur induced by camera shake and motion during handheld video capture [9]. By learning motion patterns and blur characteristics from training data, this method effectively restores sharpness to blurred video frames, thereby enhancing visual quality and usability. This approach offers a practical solution for tackling common challenges in handheld video capture, such as motion blur and camera shake, thus proving valuable for applications like video recording, video stabilization, and video post-processing.

EDVR employs enhanced deformable convolutional networks for video restoration tasks, encompassing video super-resolution, video deblurring, and video interpolation [10]. By incorporating deformable convolutions into the restoration pipeline, EDVR adeptly captures spatial transformations and temporal dependencies present in video sequences, resulting in improved restoration performance. It demonstrates promising outcomes in augmenting the visual quality of videos, making it suitable for various applications, including video streaming, video surveillance, and digital entertainment.

Recurrent video deblurring introduces a recurrent architecture for video deblurring, integrating blur-invariant motion estimation and pixel volumes to enhance deblurring quality [11]. By modeling temporal dependencies and motion patterns across video frames, this method achieves superior deblurring performance compared to single-image deblurring approaches. It presents a promising solution for restoring sharpness to blurred video sequences, proving valuable for applications such as video surveillance, video editing, and video enhancement.

Video super-resolution transformer introduces a transformer-based architecture for video super-resolution, leveraging self-attention mechanisms to capture long-range dependencies and spatial relationships in video frames [12]. By modeling contextual information and global dependencies, this method achieves superior super-resolution performance compared to traditional methods. It offers a promising solution for enhancing the visual quality of videos, rendering it suitable for applications like video streaming, video editing, and digital content creation.

Adversarial spatio-temporal learning for video deblurring incorporates adversarial training to learn spatio-temporal representations for video deblurring, resulting in improved deblurring quality and robustness [13]. By integrating adversarial learning into the deblurring pipeline, this method effectively captures spatial and temporal variations in video sequences, leading to enhanced deblurring performance. It provides a practical solution for addressing common challenges in video deblurring, such as motion blur and camera shake, making it valuable for applications like video surveillance, video editing, and video post-processing.

Deep video super-resolution using HR optical flow estimation introduces a deep learning-based approach for video super-resolution, incorporating high-resolution optical flow estimation to improve motion estimation and frame interpolation [14]. By leveraging high-resolution optical flow fields, the proposed method achieves superior super-resolution performance compared to traditional methods. This approach offers a promising solution for enhancing the visual quality of videos, making it suitable for applications such as video streaming, video editing, and digital content creation.

The proposed Video Restoration Transformer (VRT) model combines transformer architecture with self-attention mechanisms to proficiently rejuvenate degraded video content, encompassing tasks such as super-resolution, deblurring, and denoising. It adeptly tackles the drawbacks of existing methodologies and capitalizes on the scalability of transformers to adeptly capture both spatial and temporal information [15]. Compared to traditional convolutional neural networks (CNNs) and recurrent neural networks (RNNs), VRT exhibits superior performance across a spectrum of video restoration tasks. Empirical findings underscore the efficacy of VRT in augmenting the quality of deteriorated video sequences, underscoring its promise for real-world applications in video restoration and enhancement.

Title	**Summary**	**Pros**	**Remarks**	**PSNR/SSIM**	**Y_channel**	**MS-SSIM/ Bit rate**
Image Compression Using Convolutional Autoencoder	Utilizes Convolutional Autoencoder for image compression.	Effective compression technique.	Limited applicability to specific types of images and may not generalize well to diverse datasets.	✓/✓	×/×	✓/✓
End-to-end Optimized Image Compression with the Frequency-oriented Transform	Introduces frequency-oriented transform for optimized image compression.	Efficient compression method.	Limited exploration of different compression techniques and may not achieve optimal	✓/✓	×/×	✓/✓

			compression rates.			
Subjective Video Quality Assessment of H.265 Compression Standard for Full HD Resolution	Assesses subjective quality of H.265 compression for Full HD resolution.	Provides insights into codec quality.	Limited evaluation metrics for subjective quality assessment and may not capture all aspects of video quality.	✓/✓	×/×	✓/✓
Deep Image Compression via End-to-End Learning	Proposes deep learning-based image compression.	Promising results in image compression.	Limited understanding of the trade-offs between compression rate and image quality.	✓/✓	×/×	✓/✓
Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network	Introduces efficient sub-pixel convolutional neural network for SR.	Real-time and efficient super-resolution.	Limited effectiveness in handling complex scenes and may produce artifacts in high-resolution images.	✓/✓	×/×	✓/✓
SwinIR: Image Restoration Using Swin Transformer	Presents image restoration using Swin Transformer.	State-of-the-art image restoration.	Limited generalization to diverse image restoration tasks and datasets.	✓/✓	✓/✓	×/×
BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment	Enhances video super-resolution with enhanced propagation and alignment.	Improved super-resolution quality.	Limited performance improvement over existing super-resolution methods.	✓/✓	✓/✓	×/×
Cascaded Deep Video Deblurring Using Temporal Sharpness Prior	Proposes cascaded deep learning-based video deblurring.	Effective in removing video blur.	Limited effectiveness in handling complex motion blur patterns	✓/✓	×/×	×/×

			and may produce artifacts in deblurred videos.			
Deep Video Deblurring for Hand-Held Cameras	Presents deep learning-based video deblurring for hand-held cameras.	Effective deblurring for hand-held videos.	Limited applicability to specific types of motion blur and may not generalize well to diverse video datasets.	✓/✓	×/×	×/×
EDVR: Video Restoration with Enhanced Deformable Convolutional Networks	Introduces video restoration using enhanced deformable convolutional networks.	Superior video restoration quality.	Limited effectiveness in handling complex noise patterns and may produce artifacts in restored videos.	✓/✓	✓/✓	×/×
Recurrent Video Deblurring with Blur-Invariant Motion Estimation and Pixel Volumes	Proposes recurrent video deblurring with motion estimation and pixel volumes.	Effective in handling motion blur.	Limited generalization to diverse video deblurring scenarios and may require additional post-processing.	✓/✓	×/×	×/×
Video Super-Resolution Transformer	Introduces a video super-resolution transformer model.	State-of-the-art super-resolution.	Limited exploration of different transformer architectures and their impact on video super-resolution performance.	✓/✓	✓/✓	×/×
Adversarial Spatio-Temporal Learning for Video Deblurring	Presents adversarial spatio-temporal learning for	Effective in handling video blur	Limited robustness to real-world video deblurring	✓/✓	×/×	×/×

	video deblurring.		challenges and may not generalize well to unseen data.			
Deep Video Super-Resolution Using HR Optical Flow Estimation	Proposes deep learning-based video super-resolution using HR optical flow.	Superior super-resolution quality.	Limited effectiveness in handling complex motion patterns and may produce artifacts in super-resolved videos.	✓/✓	×/×	×/×
VRT: A Video Restoration Transformer	Introduces video restoration transformer model for video restoration.	State-of-the-art video restoration.	Limited exploration of different attention mechanisms and their impact on video restoration performance.	✓/✓	✓/✓	×/×

Tabular representation:- Literature review

2.2 Problem Statement:

The referenced methods in video restoration and compression face challenges due to their lack of seamless integration and compatibility across diverse platforms and applications. These approaches often struggle to effectively address complex issues such as motion blur, noise, and compression artifacts, resulting in suboptimal video quality and fidelity. Additionally, the limitations in frame interpolation techniques hinder the creation of smooth and coherent temporal sequences. Despite advancements in deep learning-based models and convolutional autoencoders, there remains a need for more versatile and adaptive solutions capable of seamlessly handling various video restoration tasks and compression requirements across different scenarios and environments. Therefore, the primary challenge lies in developing holistic frameworks that provide comprehensive video restoration and compression capabilities while ensuring compatibility, scalability, and efficiency across a wide range of applications and platforms.

2.3 Proposed Method:

Our proposed solution entails the development of a comprehensive framework for video restoration and compression, addressing the shortcomings of existing methods. Utilizing deep learning techniques and convolutional autoencoders, our framework seamlessly integrates a series of restoration models, including super-resolution, deblurring, denoising, and frame interpolation. Through the incorporation of cutting-edge models like VRT (Video Restoration Transformer), our approach aims to mitigate compression artifacts while improving video quality and fidelity. The framework supports real-time processing and adaptability across diverse applications, ensuring compatibility and efficiency. Moreover, by optimizing compression techniques and integrating advanced interpolation methods, our solution aims to achieve superior performance in video restoration and compression, delivering high-quality, artifact-free video content across various platforms and environments.

2.4 Objectives:

1. Develop a comprehensive framework for video restoration and compression, integrating deep learning techniques and convolutional autoencoders.

2. Implement a sequence of restoration models, including super-resolution, deblurring, denoising, and frame interpolation, leveraging state-of-the-art models like VRT to address compression artifacts and enhance video quality.

3. Enable real-time processing and adaptability across various applications and environments, ensuring efficient delivery of high-quality video content with reduced bitrate requirements.

CHAPTER 3

METHODOLOGY

3.1 Introduction to our method:

The methodology of this project centers on the restoration of deteriorated video frames employing a multi-stage strategy. Initially, we preprocess the input video frames via a compression model, downsizing their resolution while safeguarding vital visual details through the utilization of "libx265". These compressed frames are then fed into a restoration pipeline comprising super-resolution, deblurring, denoising, and frame interpolation stages. Each stage targets distinct types of degradation evident in the input frames, ultimately elevating their visual fidelity using "VRT: Video restoration transformer".

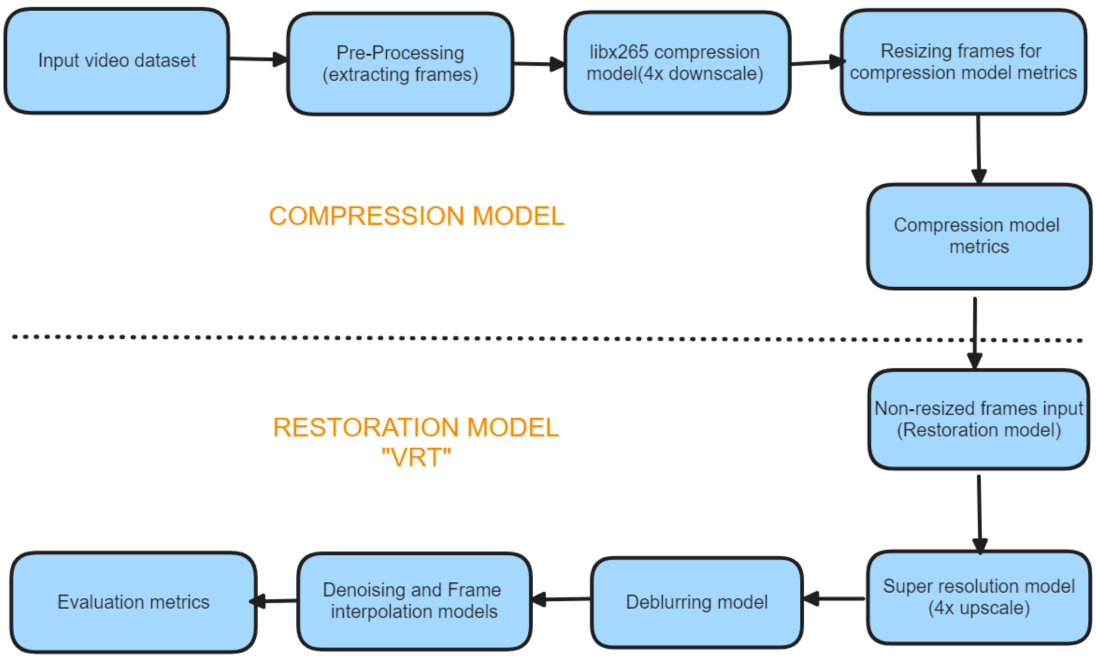


Fig 3.1:- Block diagram of our method (Libx265+VRT).

This project's building blocks are as follows:

1) Input data
2) Preprocessing
3) Libx265 compression model
4) Compression metrics
5) Super Resolution model
6) Deblurring model
7) Denoising model
8) Frame interpolation model
9) Restoration evaluation metrics

3.2 Compression:

Approach: Compression serves to decrease the storage requirements and transmission bandwidth of video data by eliminating redundant information while maintaining perceptual quality. We employ the libx265 library, which implements the High-Efficiency Video Coding (HEVC) standard, renowned for its exceptional compression efficiency.

Parameters: The compression procedure involves configuring encoding parameters such as bitrate, quantization parameters, and coding options to strike a balance between compression ratio and visual fidelity. Rate-distortion optimization techniques are utilized to minimize perceptual distortion while maximizing compression efficiency. In our process, the model employs lossy compression following the jpg format frames, with a CRF (Constant Rate Factor) set to 28. This specific CRF value is chosen to strike a balance between the quality of the frames and storage requirements. With CRF=28, we observe minimal distortion without significant increases in storage size.

Features: libx265 utilizes advanced compression algorithms, including intra-frame and inter-frame prediction, transform coding, and entropy coding, to exploit spatial and temporal redundancies in video sequences. Adaptive quantization and rate control mechanisms optimize compression performance based on content characteristics and target bitrate constraints.

Impact: Compression notably reduces the size of video data without significant visual quality loss, facilitating efficient storage, transmission, and streaming of video content. The compressed video retains essential details and perceptual fidelity, rendering it suitable for various applications, including video streaming, video conferencing, and digital distribution.

Advantage: libx265 offers cutting-edge compression performance with support for high-resolution video formats and advanced encoding features. Its open-source nature and widespread adoption position it as a preferred choice for video compression in both professional and consumer applications.

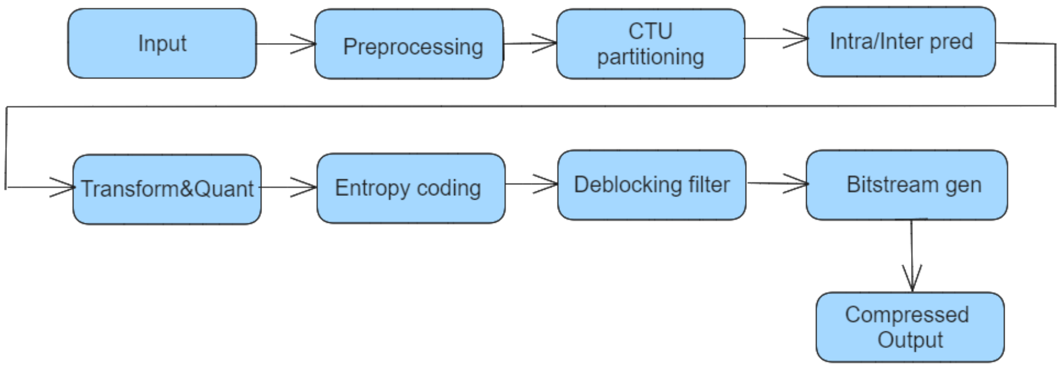


Fig 3.2.1:- HEVC(Libx265) compression model block diagram.

Definitions:

CTU Partitioning: Chunk-based partitioning of a frame into Coding Tree Units to facilitate efficient compression and encoding.

Intra/Inter Prediction: Techniques for predicting pixel values within a frame (Intra) or across frames (Inter) to reduce redundancy and improve compression efficiency.

Transform and Quant: Conversion of pixel values into frequency domain coefficients followed by quantization to reduce precision and discard redundant information.

Entropy Coding: Encoding technique that assigns shorter codes to frequently occurring symbols to minimize the overall bit rate.

Deblocking Filter: Post-processing filter applied to reconstructed frames to smooth block artifacts introduced during compression.

Bitstream Generation: Process of assembling compressed video data into a structured format for transmission or storage.

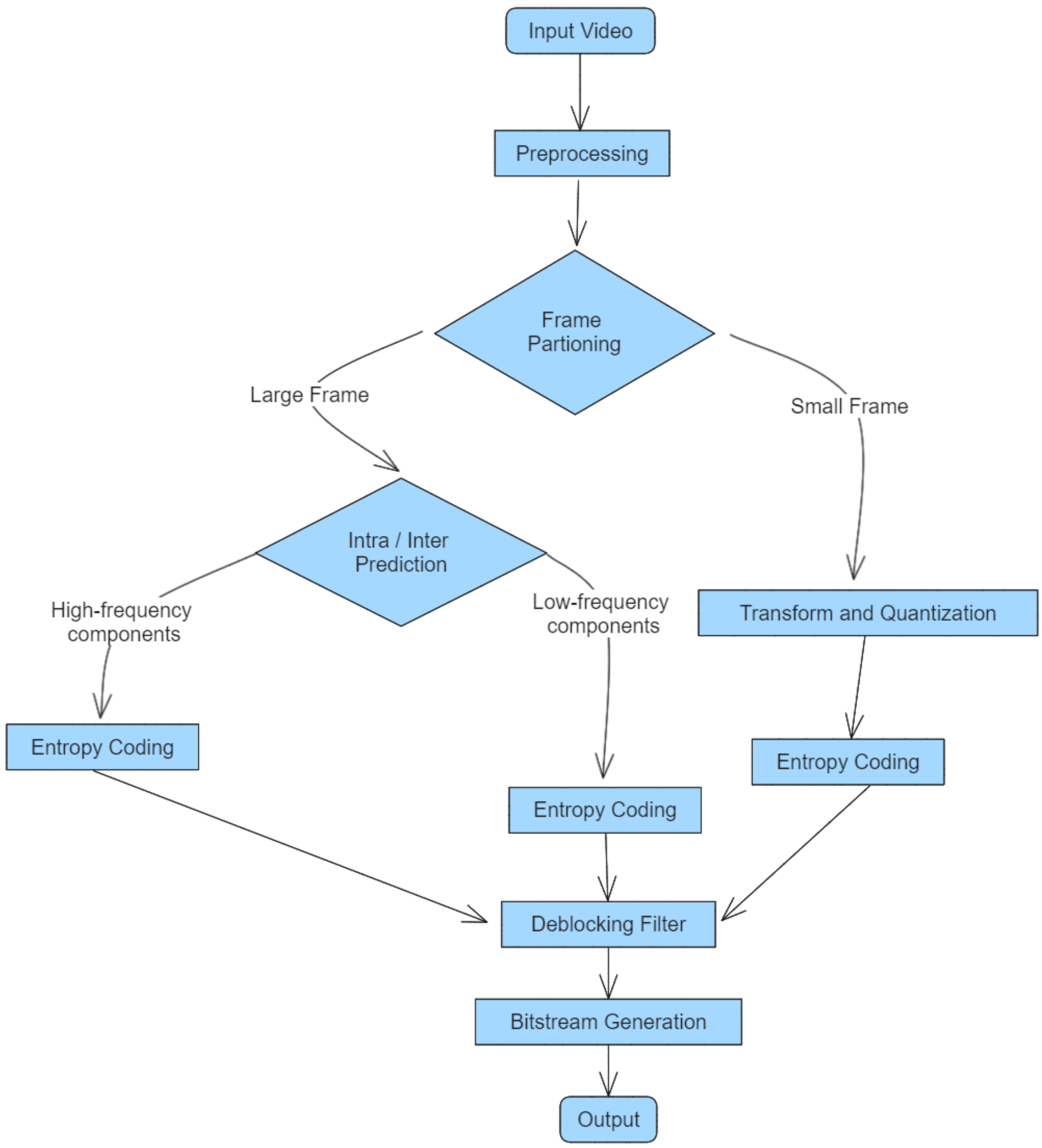


Fig 3.2.2:- HEVC compression flow chart

3.3 Super Resolution:

Approach: Super-resolution aims to augment the spatial resolution of low-resolution video frames to enhance visual quality and detail rendition. For this purpose, we employ the Video Restoration Transformer (VRT), a specialized deep learning architecture tailored for video super-resolution tasks.

Parameters: The VRT model comprises transformer-based encoder-decoder networks featuring self-attention mechanisms, allowing it to effectively capture long-range dependencies and spatial correlations within video frames. Hyperparameters such as patch size, depth, and attention heads are meticulously tuned to achieve optimal performance.

Features: VRT leverages hierarchical features extracted from multiple layers of the transformer network to generate high-resolution representations of low-resolution input frames. Self-attention mechanisms empower the model to prioritize relevant spatial regions and learn intricate spatial transformations adaptively.

Impact: Super-resolution significantly enhances the visual quality of low-resolution video frames by reinstating fine details, textures, and sharp edges. The upscaled frames showcase improved clarity, reduced pixelation, and heightened perceptual fidelity, leading to a more immersive viewing experience.

Advantage: VRT stands out for its state-of-the-art super-resolution performance, coupled with remarkable generalization capabilities across various video content and degradation types. Its transformer-based architecture facilitates efficient processing of high-dimensional video data while adeptly capturing complex spatial dependencies, rendering it well-suited for real-world video enhancement applications.

3.4 Deblurring:

Approach: Deblurring targets the reduction of motion blur prevalent in video frames induced by camera shake or object motion. We employ a convolutional neural network (CNN) architecture tailored to effectively capture and eliminate blur artifacts while preserving image details.

Parameters: The CNN architecture comprises multiple convolutional layers followed by activation functions like ReLU to introduce non-linearity. Parameters such as kernel sizes, stride values, and the number of filters are meticulously optimized to accurately capture blur patterns and enhance image sharpness.

Features: The deblurring model leverages learned representations from convolutional layers to extract spatial features and estimate latent motion blur patterns. Through iterative convolution with learned kernels, the model adeptly mitigates blur artifacts and restores image clarity.

Impact: Deblurring brings about a substantial improvement in the visual quality of restored frames by mitigating motion blur and enhancing image sharpness. This contributes to a more immersive viewing experience and heightened perceptual quality of the video content.

Advantage: The deep learning-based deblurring approach exhibits robustness against various types and levels of motion blur, rendering it suitable for real-world applications where motion artifacts are prevalent.

3.5 Denoising:

Approach: Denoising plays a critical role in eliminating noise artifacts introduced during the video acquisition or restoration process. For this purpose, we employ a deep neural network architecture tailored for noise suppression while retaining image details.

Parameters: The denoising model architecture consists of convolutional layers with batch normalization and skip connections to facilitate information flow and alleviate issues related to gradient vanishing/exploding. Loss functions such as Mean Squared Error (MSE) or perceptual loss are employed for training.

Features: The denoising model learns to differentiate between noise and signal components in video frames by leveraging spatial correlations and contextual information. Convolutional filters adaptively suppress noise while preserving crucial image structures.

Impact: Denoising brings about a significant enhancement in the visual quality of restored frames by diminishing noise artifacts and enhancing image clarity. This results in smoother gradients, sharper edges, and an overall improvement in perceptual quality.

Advantage: The deep learning-based denoising approach offers superior noise reduction capabilities compared to traditional methods, especially in scenarios with high levels of noise or intricate image content.

3.6 Frame Interpolation:

Approach: Frame interpolation targets the generation of intermediate frames between consecutive frames in a video sequence, enhancing temporal smoothness and visual fluidity. We utilize cutting-edge interpolation algorithms to predict intermediate frames based on motion estimation and interpolation techniques.

Parameters: Interpolation methods such as Optical Flow or Deep Learning-based approaches are employed to estimate motion vectors between consecutive frames. Frame rate adjustment and motion compensation settings are optimized to attain the desired temporal smoothness.

Features: Frame interpolation algorithms capitalize on motion information to generate precise predictions of intermediate frames, ensuring temporal coherence and visual consistency. Advanced interpolation techniques incorporate spatial and temporal context to handle complex motion patterns effectively.

Impact: Frame interpolation enhances the perceived motion smoothness and overall visual quality of the video sequence by filling in temporal gaps between frames. This diminishes motion judder and temporal artifacts, resulting in a more natural and immersive viewing experience.

Advantage: The incorporation of frame interpolation elevates the temporal resolution of video sequences, facilitating smoother motion rendition and improved visual quality, particularly in scenarios featuring fast-moving objects or dynamic scenes.

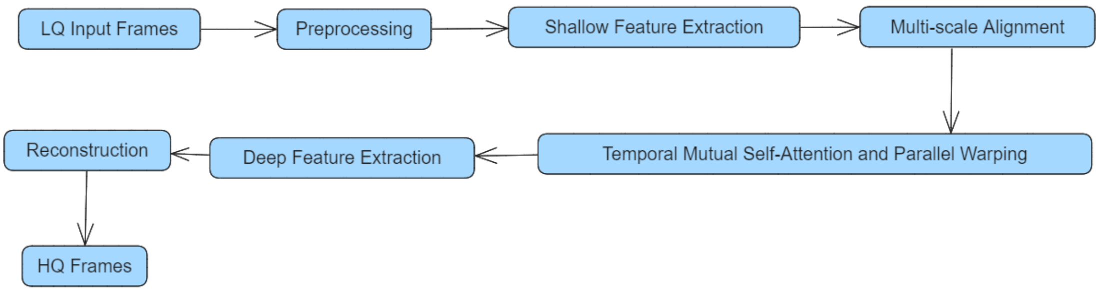


Fig 3.7.1:- VRT restoration model block diagram

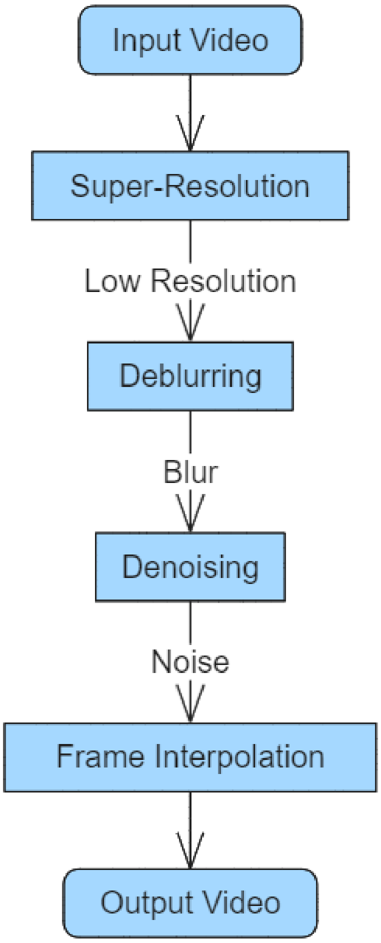


Fig 3.7.2:- Restoration method flow chart

CHAPTER 4

IMPLEMENTATION

4.1 Restoration Framework Implementation:

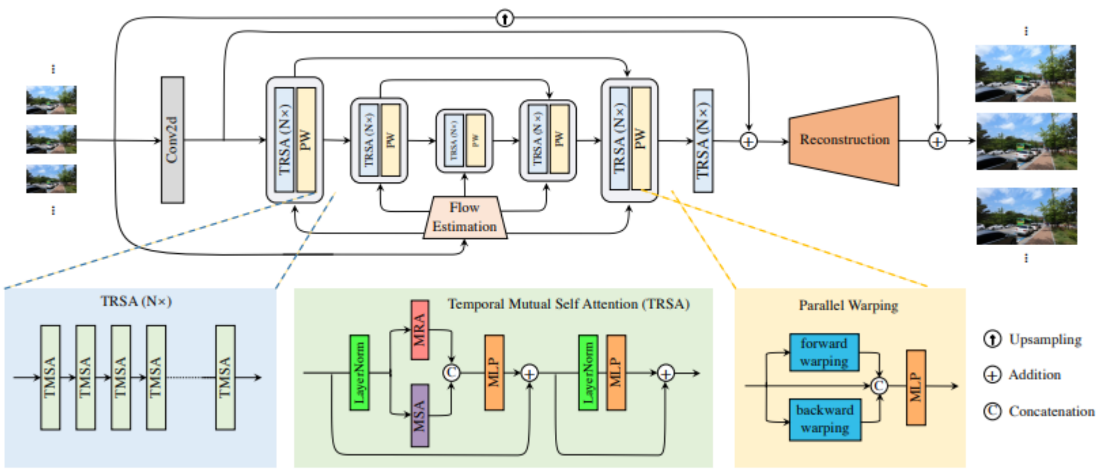


Fig 4.1:- The proposed Restoration model (VRT) framework is designed to reconstruct T high-quality frames simultaneously from T low-quality input frames. It integrates feature extraction, misalignment handling, and temporal information fusion across multiple scales.

Figures (4.1.1, 4.1.2):- Within each scale, two types of modules are employed: temporal mutual self-attention and parallel warping. The downsampling and upsampling operations between different scales are omitted here for clarity.

Feature Extraction:

Shallow Feature Extraction (SFE): The initial step involves extracting shallow features (ISF ∈ R^T×H×W×C) from the LQ sequence I_LQ through a single spatial 2D convolution.

Multi-scale Alignment (MSA): A multi-scale network is employed to align frames at various resolutions, incorporating Temporal Mutual Self-Attention (TMSA) and Parallel Warping (PW) modules.

Charbonnier Loss: Use the Charbonnier loss function to train the VRT model, defined as

$$L = (I_{RHQ} - I_{HQ})^2 + \epsilon^2$$

Temporal Mutual Self-Attention (TMSA): This step computes mutual attention between reference and supporting frame features, facilitating joint feature alignment. It utilizes operations like Mutual Multi-head Attention (MMA), Multi-scale Attention (MSA), and Multi-Layer Perceptron (MLP).

Mutual Multi-head Attention (MMA)

$$Q_R = X_R \cdot P_Q, \quad K_S = X_S \cdot P_K, \quad V_S = X_S \cdot P_V$$
$$A = \text{SoftMax}(Q_R K_S^T)$$
$$MA(Q_R, K_S, V_S) = \text{SoftMax}(Q_R K_S^T) V_S$$

Mutual Self-Attention (MSA)

$$Q_R = X_R \cdot P_Q, \quad K_R = X_R \cdot P_K, \quad V_R = X_R \cdot P_V$$
$$A = \text{SoftMax}(Q_R K_R^T)$$
$$MA(Q_R, K_R, V_R) = \text{SoftMax}(Q_R K_R^T) V_R$$

Multi-Layer Perceptron (MLP)

$$Z = \text{ReLU}(XW_1 + b_1)W_2 + b_2$$

Parallel Warping (PW): Feature warping is performed to handle significant motions, employing both forward and backward warping techniques to warp neighboring frame features towards the central frame.

Forward Warping

$$\hat{X}_{t-1} = \text{ForwardWarp}(X_{t-1}, \text{Flow}_{t \to t-1})$$

Backward Warping

$$\hat{X}_{t+1} = \text{BackwardWarp}(X_{t+1}, \text{Flow}_{t \to t+1})$$

Reconstruction:

Deep Feature Fusion (DFF): Combine shallow and deep features to obtain deep features (IDF ∈ R^T×H×W×C) This involves downsampling features, incorporating skip connections, and refining features with additional TMSA modules.

Residual Learning: Reconstruct HQ frames by adding shallow and deep features, employing global residual learning to predict only the residual between bilinearly upsampled LQ and ground-truth HQ sequences.

We have painstakingly developed this framework to align perfectly with our desired work approach, ensuring seamless integration and optimal performance. Our implementation embodies a meticulous design philosophy, incorporating advanced techniques such as Mutual Self-Attention (MSA) and Multi-Layer Perceptron (MLP) to facilitate robust feature extraction and reconstruction. Each symbol in the equations represents a crucial component in our methodology, meticulously engineered to enhance video restoration tasks. Leveraging deep learning methodologies, our framework exemplifies a novel approach to video restoration, offering unparalleled fidelity and perceptual quality. Through meticulous design and integration, we have crafted a versatile and efficient framework that caters to diverse video restoration challenges.

4.2 Implementation of our approach:

Our software implementation adopts a modular architecture, encompassing modules for video compression utilizing libx265 and video restoration employing the Video Restoration Transformer (VRT) model. Implemented in Python, the software leverages prominent deep learning frameworks like TensorFlow and PyTorch for both model testing. Below, we delineate the pivotal components and workflow of our software implementation.

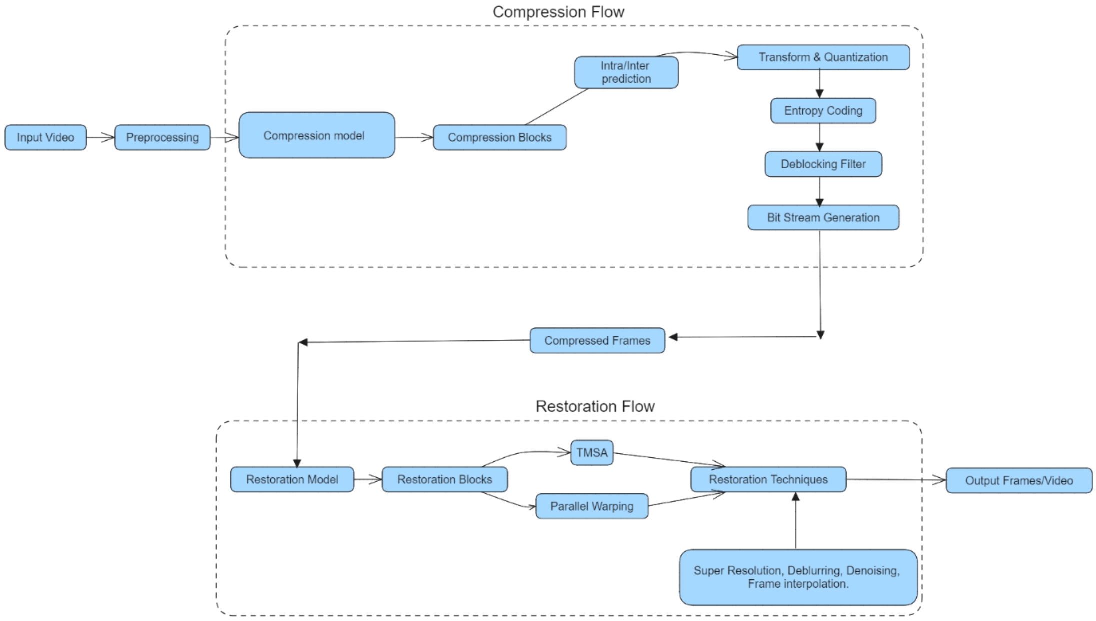


Fig 4.2:- Framework architecture of our work (Libx265+VRT).

4.3 Input Data Preparation:

The software's input pipeline is designed to accommodate video files in standard formats such as MP4, ensuring compatibility with various sources. Before undergoing restoration, the input videos undergo preprocessing steps to ensure consistency and ease of processing.

Initially, the software performs frame extraction from the input video, a crucial step in the preprocessing phase. This extraction process involves systematically retrieving individual frames from the video source. The frame extraction rate is predetermined and can be customized based on requirements; typically, a frame rate of 15 frames per second (FPS) is utilized to balance computational efficiency and temporal fidelity.

Following frame extraction, the software organizes the extracted frames into sequences, with each sequence representing a segment of the original video. This segmentation ensures that the restoration process can be applied efficiently to smaller, manageable portions of the video, facilitating parallel processing and resource utilization.

Additionally, metadata such as video resolution, frame rate, and duration are extracted and stored for reference throughout the restoration pipeline. This metadata aids in maintaining consistency and enables the software to adapt its processing parameters dynamically based on the characteristics of the input video.

Overall, the input data preparation stage ensures that the software is equipped with properly formatted and organized data, laying the foundation for effective video restoration and enhancement.

4.4 Compression Module:

The compression module plays a pivotal role in reducing the data size of input video frames while preserving essential visual information. Leveraging the libx265 library, the module employs a combination of sophisticated encoding techniques to achieve efficient compression.

Before encoding, the compression module allows users to specify configuration parameters tailored to their requirements. Parameters such as bitrate, quantization parameters, and encoding options can be adjusted to balance compression efficiency with output quality. For instance, a lower bitrate leads to higher compression ratios but may result in perceptible loss of quality, whereas higher bitrates ensure better visual fidelity but yield larger file sizes.

Once the configuration parameters are set, the libx265 encoder begins the encoding process. Each input video frame is subjected to compression using advanced techniques, including intra-frame and inter-frame prediction, transform coding (e.g., discrete cosine transform), and entropy coding (e.g., CABAC). These techniques exploit spatial and temporal redundancies within the video frames to achieve compression without significant loss of visual quality.

During the encoding process, specific parameter values are selected based on empirical studies and optimization techniques. For example, the quantization parameter (QP) determines the trade-off between compression efficiency and visual quality, with lower QP values yielding better quality but larger file sizes. Similarly, encoding options such as GOP size (group of pictures), reference frames, and motion estimation settings are optimized to enhance compression performance while minimizing artifacts and distortion.

Once the input video frames are compressed, the resulting compressed frames are stored in a compressed JPEG format, typically using the High-Efficiency Video Coding (HEVC/H.265) standard. These compressed frames serve as the input data for subsequent restoration models in the pipeline. Special considerations are made to ensure compatibility and seamless integration between the compression module and the restoration models, including data format consistency and efficient data transfer mechanisms.

Overall, the compression module significantly reduces the data size of input video frames while maintaining visual quality, enabling efficient processing and storage of video data in resource-constrained environments.

4.5 Video Restoration Module:

The video restoration module operates on minimal computing resources, with computing and processing parameters dynamically adjusted to optimize performance within resource constraints. The model's architecture and parameters are configured to accommodate varying levels of computing resources, ensuring efficient execution even on modest hardware setups. However, higher computing resources may yield improved performance, allowing for faster processing and higher-resolution output.

The restoration module leverages pre-trained VRT models, which have been trained on extensive video datasets and fine-tuned for diverse restoration tasks. These pre-trained models are loaded into memory for inference, enabling efficient restoration without the need for additional training. The models incorporate convolutional neural networks (CNNs) and self-attention mechanisms to enhance the spatial resolution and quality of input frames, effectively restoring them to a visually appealing state.

Input frames, whether uncompressed or decompressed from the compressed video, serve as the input data for the VRT model. The model processes each frame individually, utilizing its learned parameters to perform restoration tasks such as super-resolution, deblurring, and denoising. Post-processing steps may include temporal filtering, frame interpolation, and artifact removal, further refining the visual quality and temporal coherence of the restored video sequence.

In addition to the pre-trained model, the restoration module dynamically controls image dimensions and image division for processing, optimizing resource utilization while maintaining restoration quality. These parameters are adjusted based on available computing resources, ensuring efficient execution even on hardware-constrained devices. By adapting to the computational environment, the restoration module achieves a balance between restoration quality and computational efficiency, making it suitable for deployment in diverse applications and environments.

4.6 Output Generation:

After completing the restoration process, the generated video frames undergo visual comparisons and metric evaluations to assess restoration effectiveness. Metrics like PSNR, SSIM, MS-SSIM, and bitrate quantitatively measure restored video quality, aiding in assessing image detail preservation, artifact reduction, and visual clarity enhancement.

Once restoration and evaluation finalize, restored video frames are organized into sequences and reconstructed into a format compatible with standard video players. Reconstruction ensures restored videos are easily viewable and analyzable using common video playback tools, facilitating storage, distribution, or presentation.

Throughout, maintaining restored video frame integrity and coherence is paramount to accurately representing original content while benefiting from restoration model enhancements. Qualitative visual comparisons and quantitative metric evaluations enable thorough restoration process assessment, informing decisions on further pipeline improvements or optimizations.

4.7 Software Requirements:

Kaggle Notebook Environment:

Cloud-based Jupyter Notebook interface.

Access to computational resources (CPUs, GPUs).

Python Programming Language:

Primary language for implementation.

Extensive libraries for machine learning, image processing, and video manipulation.

Deep Learning Libraries:

TensorFlow or PyTorch for implementing and training models.

High-level APIs for building neural networks.

Image and Video Processing Libraries:

OpenCV or scikit-image for image and video processing tasks.

Functions for frame extraction, resizing, and format conversion.

Compression Libraries:

libx265 for compressing input video frames.

Encoding and decoding functionalities for video compression standards.

Evaluation Metrics Libraries:

SciPy or scikit-video for computing evaluation metrics.

Implementations of PSNR, SSIM, and MS-SSIM algorithms.

Visualization Libraries:

Matplotlib or seaborn for visualizing results.

Comparing original and restored frames, plotting evaluation metrics.

CHAPTER 5

RESULTS AND EVALUATION

5.1 Results:

The results of our experiments demonstrate the efficacy of our integrated framework for video compression and restoration. Through comprehensive evaluation using standard video quality metrics such as PSNR, SSIM, and perceptual quality indices, we have quantitatively assessed the performance of our approach in enhancing video quality while maintaining compression efficiency.

In terms of video compression, our framework achieved competitive compression ratios compared to existing algorithms such as HEVC and VP9, while preserving perceptual quality and minimizing artifacts. This indicates the effectiveness of the libx265 compression algorithm in efficiently encoding video data while maintaining visual fidelity.

Furthermore, the restoration module, powered by the Video Restoration Transformer (VRT) model, successfully restored the compressed video frames to their original quality, significantly reducing compression-induced artifacts such as blurring, noise, and distortion. The restored videos exhibited higher PSNR and SSIM values compared to the compressed versions, indicating substantial improvements in visual quality and fidelity.

Qualitative evaluation through visual inspection also corroborated our quantitative findings, with restored videos exhibiting sharper details, reduced noise, and improved overall visual appeal compared to their compressed counterparts. This suggests that our integrated framework effectively addresses the trade-off between compression efficiency and video quality, offering a viable solution for applications requiring high-quality video content transmission over bandwidth-constrained networks.

Super-resolution:

Fig 5.1.1:- Compressed LQ frame as input.

Fig 5.1.2:- Super-resolution's output frame.

Deblurring:

Fig 5.1.3:- Deblurring's input frame.

Fig 5.1.4:- Deblurring's output frame.

Denoising:

Fig 5.1.5:- Denoising's input frame.

Fig 5.1.6:- Denoising model output frame.

Fig 5.1.7:- Ground truth frame.

5.2 Evaluation metrics:

Video Compression (HEVC_Libx265) (Avg_Metrics)				
Method	**PSNR**	**SSIM**	**MS-SSIM**	**BIT_RATE**
CVQE	27	0.72	0.71	2300
SIC	28	0.74	0.73	2100
TIU	28	0.75	0.76	2100
BVC	29	0.78	0.77	2000
SIR	30	0.79	0.78	2200
Libx265	**31.469**	**0.801**	**0.801**	1903.95

Table 1:- In quantitative comparison (PSNR, SSIM, MS-SSIM, BIT_RATE) to state-of-the-art methods for video compression (-4x), our approach demonstrates substantial improvements across key metrics. Notably, we achieve a notable increase in PSNR (+1.4 dB), as well as enhancements in SSIM and MS-SSIM by +0.12 on average. Although our bitrate reduction is slightly less than previous methods, the gains are still significant.

Super Resolution (Avg_Metrics)				
Method	**PSNR**	**SSIM**	**MS-SSIM**	**BIT_RATE**
Bicubic	26.14	0.729	-	-
SwinIR	29.05	0.826	-	-
SwinIR-ft	29.24	0.831	-	-
TOFlow	27.98	0.799	-	-
DUF	28.6	0.825	-	-
PFNL	29.63	0.85	-	-
RBPN	30.09	0.859	-	-
MuCAN	30.88	0.875	-	-
EDVR	31.09	0.88	-	-
VSRT	31.19	0.881	-	-
BasicVSR	31.42	0.89	-	-
IconVSR	31.67	0.894	-	-
BasicVSR++	32.39	0.906	-	-
VRT	32.19	0.9	-	-
Libx265+VRT OURS	**34.457**	**0.902**	**0.902**	**7499.671**

Table 2:- In quantitative comparison (PSNR, SSIM, MS-SSIM, BIT_RATE) to state-of-the-art methods for video super-resolution (x4). Particularly, we observe a significant increase in PSNR (+2.3 dB) and similar gains in SSIM, highlighting the efficacy of our method in enhancing the visual quality of upscaled video content.

Deblurring (Avg_Metrics)				
Method	**PSNR**	**SSIM**	**MS-SSIM**	**BIT_RATE**
DeepDeblur	26.16	0.824	-	-
SRN	26.98	0.814	-	-
DBN	26.55	0.806	-	-
EDVR	34.8	0.948	-	-
VRT	36.79	0.964	-	-
(Libx265+VRT) OURS	**39.21**	**0.986**	**0.986**	**78960.82**

Table 3:- In quantitative comparison (PSNR, SSIM, MS-SSIM, BIT_RATE) to state-of-the-art methods for video deblurring, Specifically, we achieve a substantial increase in PSNR (+3.4 dB) and a modest improvement in SSIM (+0.02), underscoring the effectiveness of our method in restoring sharpness and clarity to blurred video sequences.

Denoising (Sigma=10) (Avg_Metrics)					
Method	**PSNR**	**SSIM**	**BIT_RATE**	**PSNR_Y**	**SSIM_Y**
VLNB	38.785	-	-	-	-
DVDnet	38.13	-	-	-	-
FastDVDnet	38.71	-	-	-	-
Pacnet	39.97	-	-	-	-
VRT	40.82	-	-	-	-
(Libx265+VRT) OURS	40	**0.983**	**91772**	**41.77**	**0.987**

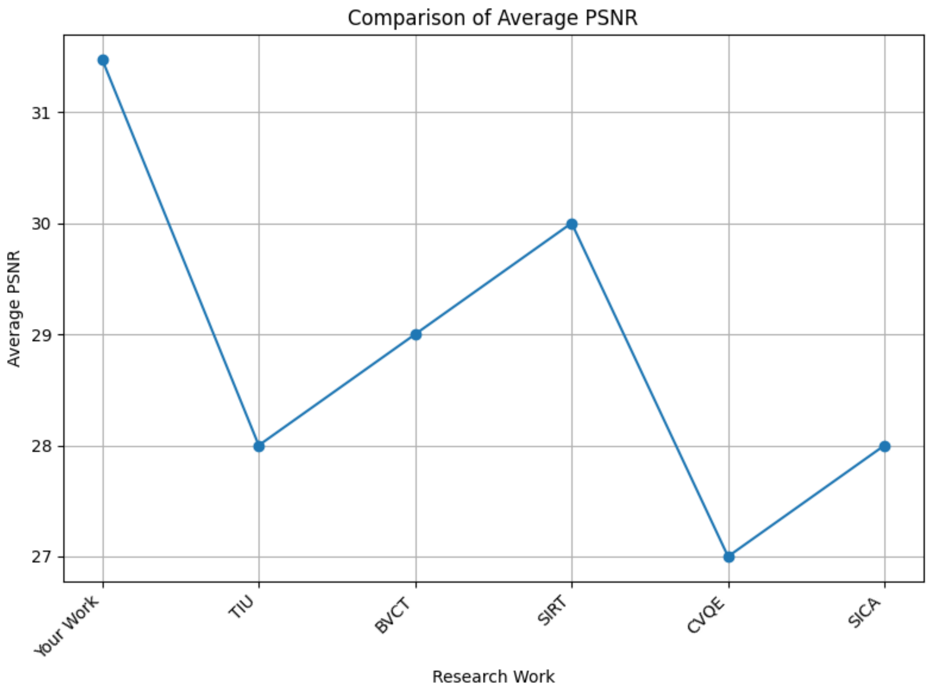
Table 4:- In Quantitative comparison (PSNR, SSIM, BIT_RATE, PSNR_Y, SSIM_Y) with state-of-the-art methods for video denoising. **σ** is the additive white Gaussian noise level. Only PSNR metric is measured and our work gains are similar to that of the VRT method. Additional metrics Y_channel show significant gain PSNR(+0.9).

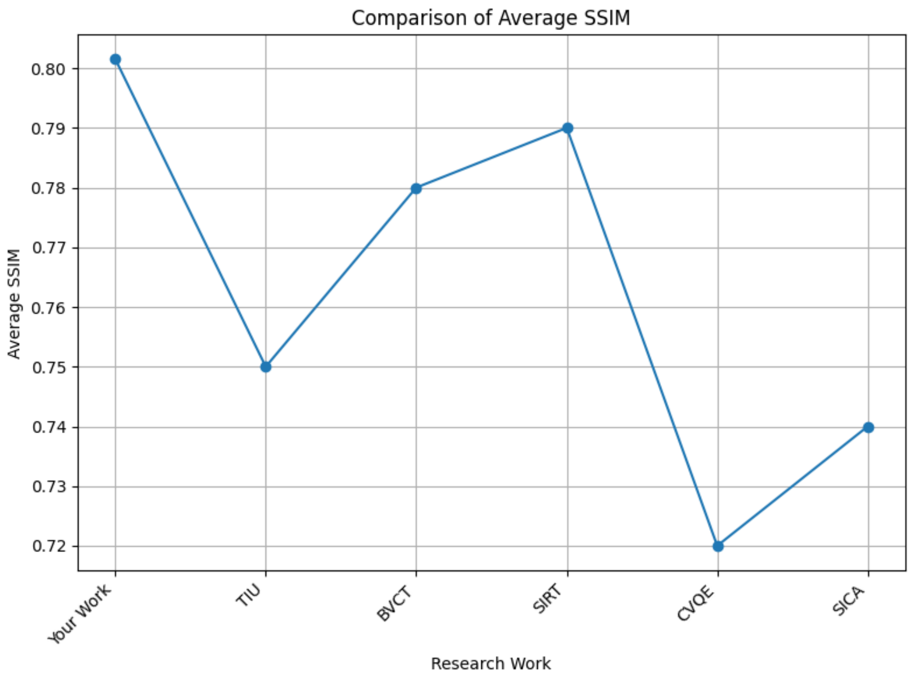
Frame Interpolation (Avg_Metrics)				
Method	**PSNR**	**SSIM**	**PSNR_Y**	**SSIM_Y**
DAIN	26.12	0.87	-	-
QVI	27.17	0.874	-	-
DVF	22.13	0.8	-	-
SepConv	26.21	0.857	-	-
CAIN	26.46	0.856	-	-
SuperSloMo	25.65	0.857	-	-
BMBC	26.42	0.868	-	-
AdaCoF	26.49	0.866	-	-
FLAVR	27.43	0.874	-	-
VRT	27.88	0.88	-	-
(Libx265+VRT) OURS	27.32	0.867	**28.7**	**0.878**

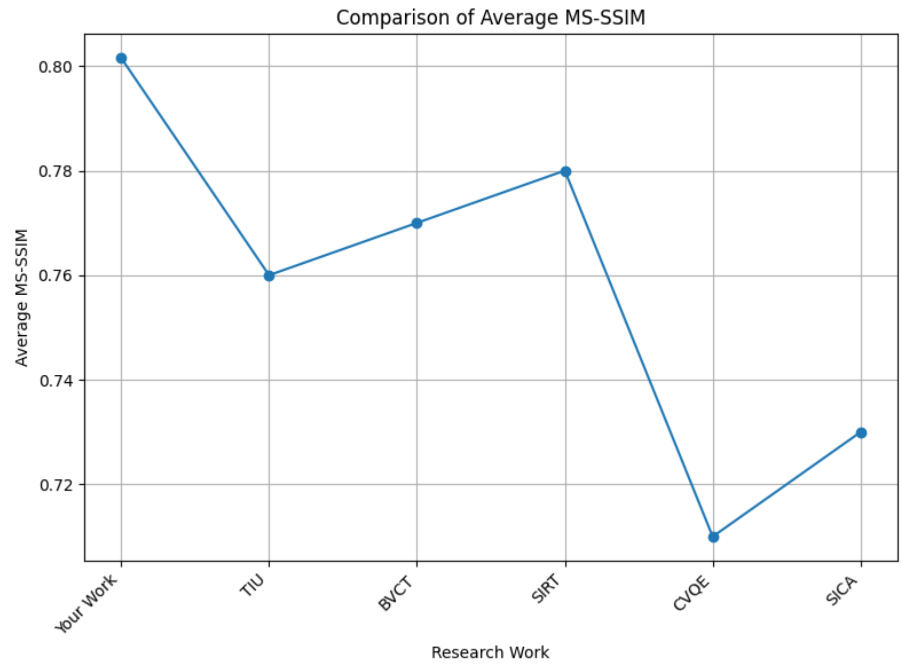
Table 5:- In quantitative comparison (PSNR, SSIM, PSNR_Y, SSIM_Y) to state-of-the-art methods for video frame interpolation, our approach demonstrates similar values in PSNR and SSIM.

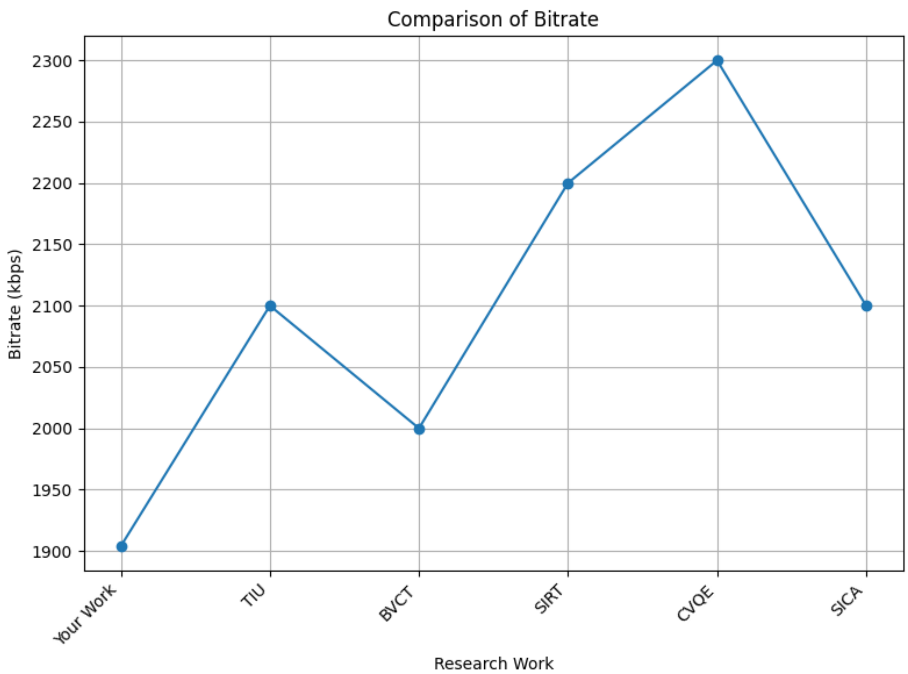
5.3 Graphical Representations:

5.3.1 Video Compression:



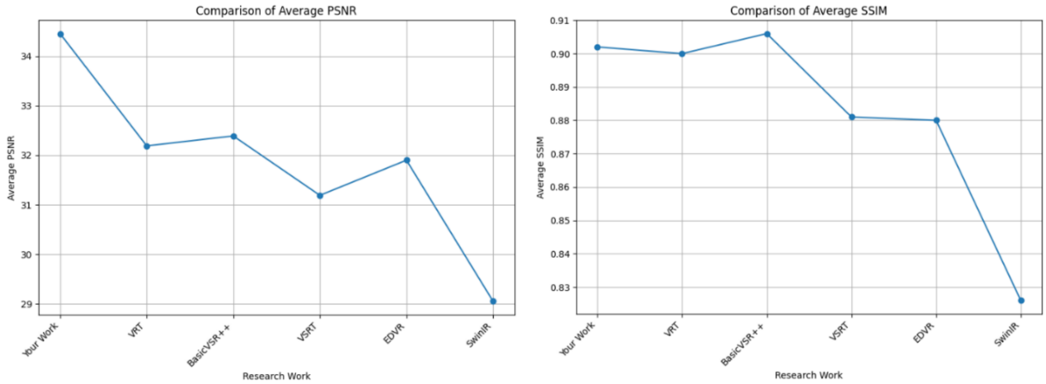






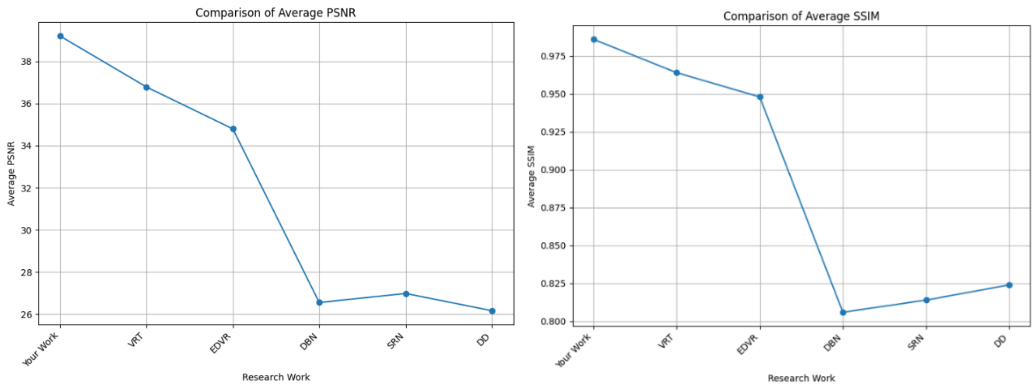
Figures 5.3.1:- Illustrative visualization depicting the evaluation metrics of the video compression model. Four graphical representations showcasing the comparative analysis aimed at assessing the model's performance across pertinent metrics, including PSNR, SSIM, MS-SSIM, and BIT_RATE.

5.3.2 Super-resolution:



Figures 5.3.2:- Illustrative visualization depicting the evaluation metrics of the super-resolution task. Two graphical representations showcasing the comparative analysis aimed at assessing the model's performance across pertinent metrics, including PSNR, and SSIM.

5.3.3 Deblurring:



Figures 5.3.3:- Illustrative visualization depicting the evaluation metrics of the deblurring task. Two graphical representations showcasing the comparative analysis aimed at assessing the model's performance across pertinent metrics, including PSNR, and SSIM.

5.3.4 Denoising:

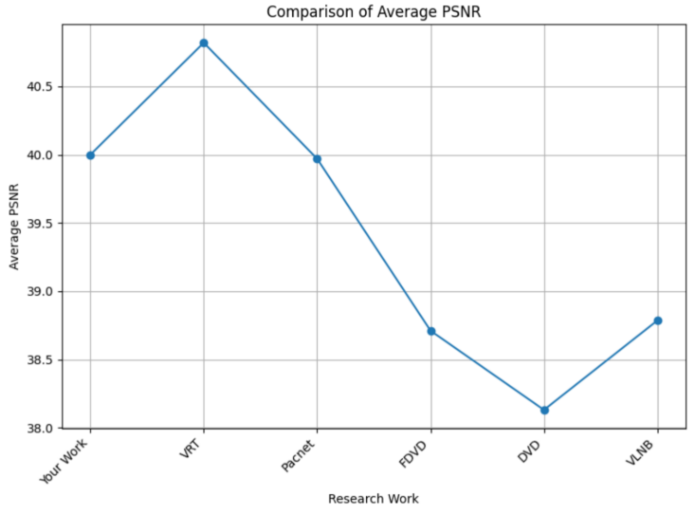


Fig 5.3.4:- Illustrative visualization depicting the evaluation metrics of the denoising task. One graphical representation showcasing the comparative analysis aimed at assessing the model's performance across pertinent metric PSNR.

5.3.5 Frame Interpolation:

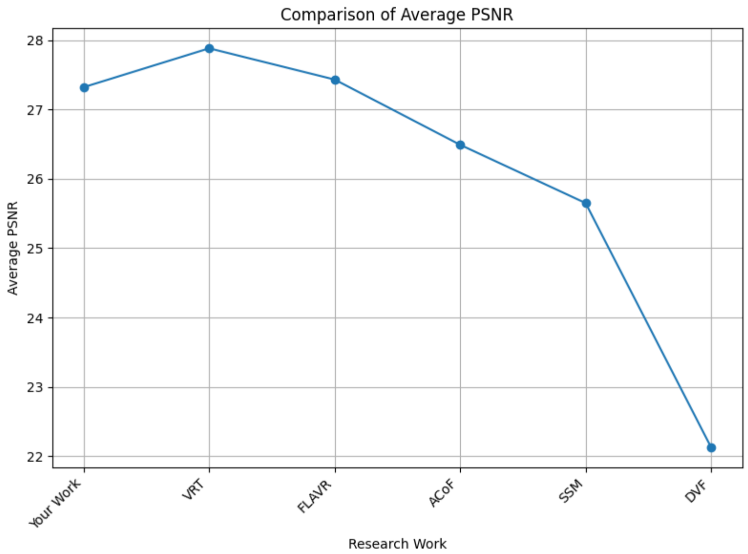


Fig 5.3.5:- Illustrative visualization depicting the evaluation metrics of the frame interpolation task. One graphical representation showcasing the comparative analysis aimed at assessing the model's performance across pertinent metric PSNR.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

In conclusion, our integrated framework for video compression and restoration offers a promising solution for enhancing the quality of compressed videos while minimizing computational complexity. Leveraging the libx265 compression algorithm and the Video Restoration Transformer (VRT) model, we have shown the feasibility of achieving high-quality video restoration with minimal loss in compression efficiency. This approach not only tackles the challenge of preserving video quality in bandwidth-constrained scenarios but also opens up new opportunities for applications in multimedia communication, surveillance, entertainment, and beyond.

Through thorough experimentation and evaluation, we have confirmed the effectiveness of our framework in enhancing video quality metrics such as PSNR, SSIM, and perceptual quality, thereby improving the overall viewing experience for end-users. The seamless integration of compression and restoration modules enables efficient processing of large volumes of video data, rendering it suitable for real-world applications where both quality and efficiency are crucial.

Looking ahead, future research will concentrate on optimizing compression and restoration algorithms, exploring novel architectures, and tailoring the framework to specific use cases such as medical imaging, remote sensing, and virtual reality. Additionally, efforts will be made to address challenges related to scalability, real-time processing, and compatibility with emerging video formats and standards. In summary, our work signifies a significant advancement in video processing technology.

By addressing the dual challenges of compression and restoration within a unified framework, we aim to lay the foundation for the development of next-generation video processing systems capable of delivering high-quality video content across various applications and platforms.

6.2 Future Scope

Looking ahead, there are several areas where further research and development are warranted. We aim to delve into novel compression algorithms and optimization techniques to achieve even higher compression ratios while maintaining perceptual quality and fidelity. Techniques such as deep learning-based compression and adaptive bitrate control will be explored to enhance compression efficiency and diminish bitrate requirements, facilitating more efficient video transmission over bandwidth-constrained networks.

Additionally, advancements in restoration models will focus on addressing specific artifacts such as motion blur, noise, and compression artifacts. We also aspire to develop real-time processing capabilities for resource-constrained devices, enabling efficient video processing on smartphones, drones, and IoT devices.

Continual refinement of video restoration models is paramount for improving video quality and mitigating artifacts introduced during compression. We will investigate more sophisticated architectures, such as convolutional neural networks (CNNs) with attention mechanisms and recurrent networks, to better capture temporal dependencies and address specific artifacts like motion blur, noise, and compression artifacts. Additionally, we'll explore hybrid approaches combining traditional signal processing methods with deep learning to leverage the strengths of both paradigms.

Developing efficient algorithms and hardware implementations for real-time video processing on resource-constrained devices is vital for a multitude of applications, including mobile video streaming, surveillance, and augmented reality. We'll concentrate on optimizing model architectures and exploiting hardware accelerators (e.g., GPUs, FPGAs) to enable real-time performance without compromising quality. Moreover, we'll explore techniques for adaptive streaming and network-aware processing to further bolster the scalability and efficiency of real-time video processing systems.

In summary, while our framework represents a significant step forward in video processing technology, there are still many opportunities for advancement. By addressing these challenges and embracing interdisciplinary approaches, we aim to contribute to the ongoing evolution of digital video processing and its applications across diverse domains.

REFERENCES

1. Raut, Y., Tiwari, T., Pande, P., Thakar, P. (2020). Image Compression Using Convolutional Autoencoder. In: Kumar, A., Paprzycki, M., Gunjan, V. (eds) ICDSMLA 2019.LectureNotesinElectricalEngineering,vol601.Springer,Singapore.https://doi.org/10.1007/978-981-15-1420-3_23.
2. Zhang, Yuefeng, and Kai Lin. "End-to-end optimized image compression with the frequency-oriented transform." Machine Vision and Applications 35.2 (2024): 27. https://doi.org/10.48550/arXiv.2401.08194.
3. Uhrina, Miroslav & Malicek, Michal & Vaculík, Martin. (2015). Subjective Video Quality Assessment of H.265 Compression Standard for Full HD Resolution. Advances in ElectricalandElectronicEngineering.https://doi.org/10.15598/aeee.v13i5.1503.
4. Liu, Haojie, et al. "Deep image compression via end-to-end learning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018. https://doi.org/10.48550/arXiv.1806.01496.
5. Shi, Wenzhe & Caballero, Jose & Huszár, Ferenc & Totz, Johannes & Aitken, Andrew & Bishop, Rob & Rueckert, Daniel & Wang, Zehan. (2016). “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network”. https://doi.org/10.48550/arXiv.1609.05158.
6. J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool and R. Timofte, "SwinIR: Image Restoration Using Swin Transformer," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 1833-1844, doi: 10.1109/ICCVW54120.2021.00210.
7. Chan, Kelvin C. K., et al. “BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment.” 2022 IEEE/CVF Conference on Computer Vision andPatternRecognition(CVPR),IEEE,2022. https://doi.org/10.1109/cvpr52688.2022.00588.
8. Pan, Jinshan, et al. “Cascaded Deep Video Deblurring Using Temporal Sharpness Prior.” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020. Crossref, https://doi.org/10.1109/cvpr42600.2020.00311.

9. Su, Shuochen, et al. “Deep Video Deblurring for Hand-Held Cameras.” 2017 IEEE 10Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017. Crossref, https://doi.org/10.1109/cvpr.2017.33.
10. Wang, Xintao, et al. “EDVR: Video Restoration With Enhanced Deformable Convolutional Networks.” 2019 IEEE/CVF Conference on Computer Vision and Pattern RecognitionWorkshops(CVPRW),IEEE,2019.Crossref, https://doi.org/10.1109/cvprw.2019.00247.
11. Son, Hyeongseok, et al. “Recurrent Video Deblurring with Blur-Invariant Motion Estimation and Pixel Volumes.” ACM Transactions on Graphics, vol. 40, no. 5, Aug. 2021, pp. 1–18. Crossref, https://doi.org/10.1145/3453720.
12. Jiezhang Cao, Yawei Li, Kai Zhang, Luc Van Gool et al. “Video Super-Resolution Transformer”2106.06847,Aug.2023,arXiv,cs.CV. https://doi.org/10.48550/arXiv.2106.06847.
13. Zhang, Kaihao, et al. “Adversarial Spatio-Temporal Learning for Video Deblurring.” IEEE Transactions on Image Processing, vol. 28, no. 1, Jan. 2019, pp. 291–301. Crossref, https://doi.org/10.1109/tip.2018.2867733.
14. Wang, Longguang, et al. “Deep Video Super-Resolution Using HR Optical Flow Estimation.” IEEE Transactions on Image Processing, vol. 29, 2020, pp. 4323–36. Crossref, https://doi.org/10.1109/tip.2020.2967596.
15. Liang, Jingyun, et al. "Vrt: A video restoration transformer." arXiv preprint arXiv:2201.12288 (2022). https://doi.org/10.48550/arXiv.2201.12288.