Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 3, pp. 1658–1666, DOI 10.12694/scpe.v25i3.2696

SCALABLE VIDEO FIDELITY ENHANCEMENT: LEVERAGING THE SOTA AI MODELS

ANKIT DAS*, DEVEN PRAKASH PARAMAJ†, AND SHAMBHAVI BR‡

Abstract. Improving visual quality is crucial as we navigate through the vast world of data. State-of-the-art (SOTA) artificial intelligence (AI) models provide highly effective solutions. Driven by the ever-growing demand for high-fidelity multimedia content, this research explores the groundbreaking capabilities of SOTA AI models to revolutionize video quality enhancement. Existing video capture methods often struggle with limitations in hardware, bandwidth, and compression, leading to subpar visual experiences. To address this challenge, we propose a novel Video Quality Enhancement Solution (VQES) that synergistically combines Google FILM for frame interpolation and Real-ESRGAN for image super-resolution. By applying these models to each video frame and integrating scalable post-processing techniques, a comprehensive VQES has been devised. Extensive experiments demonstrate that our VQES outperforms existing methods in terms of peak signal-to-noise ratio (PSNR) improvement and user-perceived visual quality. By advancing video fidelity, this research paves the way for consistently immersive, informative, and enjoyable visual experiences.

Key words: SOTA, AI models, video fidelity, Google FILM, Real-ESRGAN, video frame interpolation, image super-resolution, Video Quality Enhancement Solution (VQES), PSNR

1. Introduction. In recent years, there has been a significant surge in demand for high-quality video content across various domains, including entertainment, surveillance, and virtual reality. Achieving superior video fidelity is crucial to providing immersive visual experiences and extracting valuable information from videos. However, capturing videos with pristine quality is often challenging due to limitations in camera hardware, bandwidth constraints, and other factors. Consequently, there is a growing need for effective video quality enhancement techniques.

To address the limitations of traditional video enhancement approaches, this paper explores the use of state-of-the-art artificial intelligence (AI) models to enhance video fidelity. By leveraging advanced AI techniques, we aim to push the boundaries of video quality to unprecedented levels. Specifically, we focus on two cutting-edge models: Google FILM for video frame interpolation and Real-ESRGAN for image super-resolution [1, 2]. These models have demonstrated remarkable capabilities in their respective domains and offer promising potential for enhancing video content.

The primary objective of this research is to develop a comprehensive Video Quality Enhancement Solution by combining the strengths of Google FILM and Real-ESRGAN. The objectives of the solution are to:

- Increase video temporal resolution by employing Google FILM for frame interpolation, leading to smoother playback and more realistic motion.
- Enhance video spatial resolution by utilising Real-ESRGAN for image super-resolution, resulting in sharper details and improved clarity.
- Optimise visual quality through efficient post-processing techniques, ensuring a seamless and pleasing viewing experience.

Beyond the primary goal of enhancing video detail, smoothness, and resolution, our research pursues the following secondary objectives: quantifying effectiveness, benchmarking performance, analysing efficiency, and addressing limitations.

To achieve our objective, we adopt a multi-stage approach. Firstly, we utilise the Google FILM model for video frame interpolation. This technique generates intermediate frames between consecutive frames, thereby increasing the video's temporal resolution. Subsequently, we employ the Real-ESRGAN model to perform

*Department of Information Science and Engineering, BMS College of Engineering (ankitdas.is19@bmsce.ac.in).
†Department of Information Science and Engineering, BMS College of Engineering (devenparamaj.is19@bmsce.ac.in)
‡Department of Information Science and Engineering, BMS College of Engineering (shambhavibr.ise@bmsce.ac.in).

image super-resolution on each frame of the video. This process enhances the spatial resolution, resulting in sharper and more detailed frames. Finally, we implement efficient post-processing techniques [3, 4, 5, 6, 1] to further refine the enhanced video, ensuring optimal visual quality.

This research presents a groundbreaking approach to video quality enhancement, driven by the following key contributions:

- Synergistic Integration of AI Models: We propose a novel solution that combines the strengths of two state-of-the-art AI models, Google FILM and Real-ESRGAN, to achieve unparalleled video fidelity improvements.
- Scalability and Efficiency: We employ efficient techniques like the singleton design pattern and threaded processing to ensure real-time or near-real-time performance for high-resolution videos.
- Queue-based Parallel Processing: We implement frame-passing queues to enable seamless and parallel processing of frames, further enhancing the solution's efficiency and scalability.
- Unprecedented Video Fidelity: Our solution demonstrably enhances video quality in terms of detail, smoothness, and resolution, surpassing the capabilities of existing methods.

By addressing the critical need for high-quality video content across diverse domains, this research holds significant potential to revolutionise the way we capture, analyse, and experience visual information.

2. Related Work. The field of video quality enhancement has witnessed significant advancements in recent years, driven by the convergence of powerful deep-learning techniques and the ever-growing demand for high-quality visual experiences. This section delves into existing research efforts related to video frame interpolation and super-resolution, highlighting their strengths and limitations, and setting the context for our proposed Scalable Video Fidelity Enhancement solution.

Video Frame Interpolation: A method for synthesising intermediate frames of a video known as video frame interpolation (VFI) [3] can be used to create a slow-motion video, boost the frame rate of a video, and recover lost frames during video streaming. VFI methods are classified as optical or diffractive super-resolution, which takes advantage of sub-pixel misalignment between multiple images of the same scene, or geometrical or image-processing super-resolution, which uses a single image or a sequence of images with limited information [4, 5, 6]. Existing VFI methods face difficulties in dealing with large amounts of motion, preserving fine details and textures, avoiding artefacts and noise, and improving computational efficiency [6, 7].

- Optical Flow-based methods: These methods estimate optical flow between consecutive frames to generate intermediate frames. Popular examples include FlowField [8] and EDVR [9]. While effective in capturing motion, they can suffer from artefacts and inaccuracies, especially in complex scenes.
- Learning-based methods: These methods leverage deep learning models to directly learn the interpolation process. One notable example worth mentioning is DAIN [7]. While offering superior quality compared to optical flow methods, they often require large datasets for training and can be computationally expensive.
- Our approach (VQES): Google FILM, employed in our solution, falls under this category. It utilises a spatiotemporal transformer architecture to capture long-range dependencies and generate realistic intermediate frames. Compared to previous methods, it demonstrates improved accuracy and robustness, especially in challenging scenarios.

Image super-resolution (ISR) is another technique for increasing image resolution by adding sub-pixel detail [10, 11]. ISR techniques can be categorised as pre-upsampling super-resolution or post-upsampling super-resolution. Pre-upsampling super-resolution refines an upsampled image using conventional methods like bi-cubic interpolation and deep learning while post-upsampling super-resolution uses cutting-edge models like residual networks, multi-stage residual networks, recursive networks, progressive reconstruction networks, multi-branch networks, and attention-based networks [2]. ISR methods have a wide range of applications, including video surveillance, medical diagnosis, and remote sensing. However, ISR methods also face limitations such as computational inefficiency, loss of fine details and textures, and the generation of artefacts and noise.

- Reconstruction-based methods: These methods reconstruct high-resolution frames from low-resolution ones by utilising prior knowledge about image structures. Examples include SRCNN [12] and FSRCNN [13]. While effective in simple cases, they struggle with complex textures and aliasing artefacts.
- Generative adversarial network (GAN)-based methods: These methods employ GANs to learn the

mapping between low- and high-resolution images. Examples include ESRGAN [2] and SRGAN [14]. While achieving impressive results, they can be prone to instability and generate unrealistic details.
- Our approach (VQES): We integrate Real-ESRGAN into our solution for frame enhancement. Its residual-in-residual architecture and perceptual loss function enable high-fidelity reconstruction, preserving temporal consistency and suppressing artefacts.

While existing approaches have made significant advancements in video quality enhancement, there are still several limitations that need to be addressed [3]. Many techniques suffer from computational inefficiency, requiring extensive processing time and resources, hence face challenges in terms of scalability and efficiency. Additionally, preserving fine details and textures while avoiding artefacts and noise remains a challenge. Moreover, there is a lack of comprehensive solutions that combine multiple state-of-the-art models to achieve unprecedented video quality improvements.

In this study, we use SOTA AI models to improve video quality and overcome current methods' shortcomings. Our proposed Scalable Video Quality Enhancement solution addresses these concerns through:
- Multi-resolution processing: We employ a multi-resolution pyramid approach to efficiently handle videos of varying resolutions. This reduces computational cost while maintaining visual quality.
- Adaptive model selection: We dynamically select the appropriate model (fine-tuned FILM and Real-ESRGAN) based on the video content and desired enhancement level. This optimises resource allocation and ensures efficient processing.

By combining the strengths of Google FILM for video frame interpolation and Real-ESRGAN for image super-resolution with a focus on scalability and efficiency, we propose a comprehensive VQES [3, 1, 2]. The proposed solution not only enhances the visual quality but also incorporates efficient processing techniques, such as the Singleton Design Pattern and threaded processing, to improve computational efficiency [15]. Moreover, the utilisation of queues for frame passing enables seamless and parallel processing of frames, further enhancing the overall efficiency of the solution.

Our proposed solution aims to make high-quality video enhancement accessible across diverse applications and hardware platforms.

3. Methodology.

3.1. Overview. This section describes the research design and methods used to develop the VQES that exploits SOTA AI models for video frame interpolation and image super-resolution. This solution leverages efficient processing techniques, such as the Singleton Design Pattern and threaded processing, to enhance computational efficiency. Moreover, queues for frame passing are utilised to enable parallel processing of frames, further improving the overall efficiency of the solution.

3.2. Google FILM for Video Frame Interpolation. The first step in the methodology involves utilising the Google FILM model for frame interpolation.

This model employs a flow-based approach to estimate the optical flow between two consecutive frames. By leveraging estimated motion information, Google FILM [1] generates intermediate frames to increase the temporal resolution of the video. FILM is composed of three components as shown in Figure 3.1: (1) A feature extractor that extracts deep multi-scale (pyramid) features from each input image; (2) a bi-directional motion estimator that computes pixel-wise motion (i.e., flows) at each pyramid level; and (3) a fusion module that outputs the final interpolated image.

A dedicated thread ***"Thread-1: Frame Interpolation"*** processes each frame as shown in Figure 3.3. This thread operates independently, enabling efficient parallel processing and maximizing resource utilization. This parallelization significantly boosts the overall performance of our VQES. To further optimize efficiency, we leverage the Singleton Design Pattern. Under this pattern, only a single instance of the Google FILM model is instantiated and employed throughout the entire frame interpolation process. This avoids redundant loading and initialization of the model for each frame, resulting in substantial computational savings. This optimization becomes particularly crucial as video resolutions and frame rates increase, as it prevents resource bottlenecks and maintains smooth system operation.

3.3. Real-ESRGAN for Frame Super-Resolution. Following the frame interpolation stage, we apply the Real-ESRGAN model for image super-resolution.

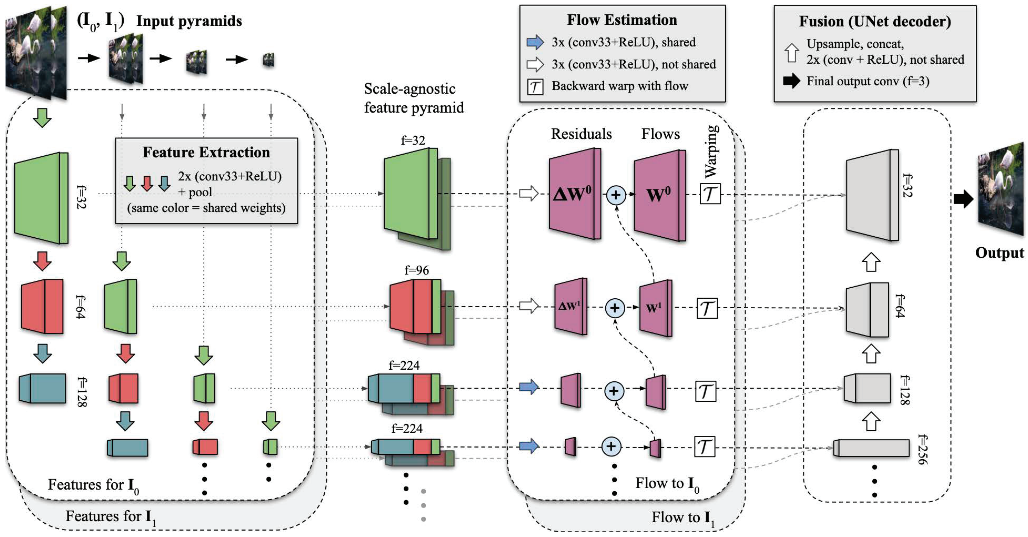

Fig. 3.1: **FILM architecture**

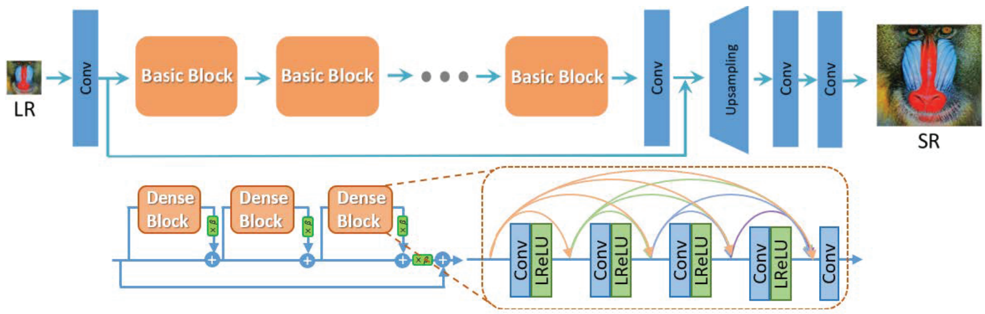

Fig. 3.2: **Real-ESRGAN Architecture for Image Super-Resolution**

Real-ESRGAN [2] employs deep convolutional neural networks to learn high-resolution image mappings from low-resolution inputs. As shown in Figure 3.2 Real-ESRGAN's architecture can be visualised as a deep, multi-layered convolutional neural network (CNN). It consists of convolutional, residual, and upsampling blocks that work together to enhance the resolution and visual quality of low-resolution images. It employs perceptual loss, adversarial loss, and feature loss, along with a pre-trained VGG network for perceptual quality assessment.

Following the crucial step of frame interpolation, each frame, including the newly synthesized ones, undergoes a dedicated image super-resolution process. As illustrated in Figure 3.3, a specially designated thread, aptly named ***"Thread-2: Frame Enhancement"*** assumes this responsibility. Operating in a queue-based manner, it tackles each frame individually, ensuring efficient throughput and resource utilization. Similar to the thread responsible for frame interpolation, ***"Thread-2: Frame Enhancement"*** leverages the Singleton Design Pattern for optimal performance. This design principle ensures that only a single instance of the Real-ESRGAN model exists and serves all frames.

The Real-ESRGAN model employed in this stage is specifically trained for image super-resolution, meaning it can intelligently upscale the resolution of each frame while preserving visual details and minimizing artefacts. This enhances the overall sharpness, clarity, and visual fidelity of the video, creating a more immersive and enjoyable viewing experience.

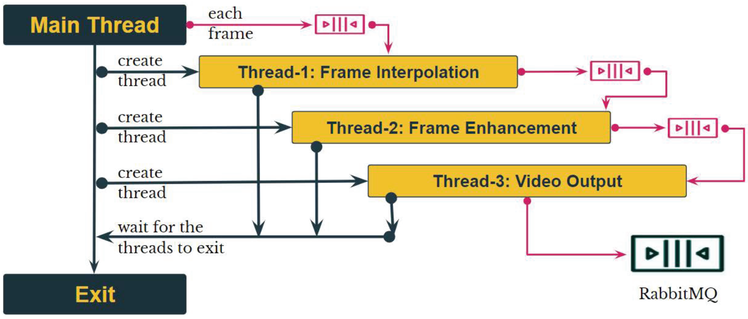

Fig. 3.3: **System Design for Video Quality Enhancement**

3.4. Queue-Based Frame Passing: Orchestrating the Workflow for Maximum Efficiency. The proposed video quality enhancement system leverages a meticulously designed queue-based frame passing mechanism, as illustrated in Figure 3.3. This mechanism plays a pivotal role in maximising processing efficiency, ensuring smooth video processing, and ultimately delivering unparalleled quality improvements.

Seamless Handoff and Reduced Wait Times. Unlike traditional sequential processing, where one stage must be completed before the next can begin, the queue-based approach allows for parallel execution. Once a frame finishes processing in the frame interpolation stage, it's instantaneously placed in a designated queue for immediate pickup by the image super-resolution stage. This eliminates idle time between stages, minimising overall processing latency and also resource utilisation.

Synchronised Flow and Orderly Progression. Each queue acts as a buffer, temporarily holding processed frames until the subsequent stage is ready. This ensures the orderly progression of frames through the pipeline, preventing out-of-sequence processing and maintaining the video's temporal integrity. By acting as a coordination mechanism, the queues guarantee that both stages operate in synchronised harmony, preventing bottlenecks and ensuring a smooth workflow.

Scalability and Adaptability. The queue-based approach inherently boasts scalability and adaptability. New processing stages can be seamlessly integrated by adding additional queues and modifying the overall flow. This flexibility allows the system to evolve and accommodate future advancements in video processing techniques.

3.5. Implementation and System Design: A Deep Dive into the Architecture. The system architecture comprises several key components, including the Google FILM model, the Real-ESRGAN model, post-processing algorithms, and frame-passing queues. The proposed video quality enhancement methodology comes to life through a carefully crafted system design, meticulously engineered for efficiency, scalability, and exceptional video processing capabilities. This section delves deeper into the architectural choices and design patterns that empower the system to deliver its impressive results.

Foundation of Efficiency: The Singleton Design Pattern. . The system architecture comprises several key components, including the Google FILM model, the Real-ESRGAN model, post-processing algorithms, and frame-passing queues. At the heart of the system lies the Singleton Design Pattern [15]. This design principle ensures that only one instance of each computationally expensive model (Google FILM and Real-ESRGAN) exists throughout the processing pipeline. This eliminates redundant loading and initialisation for each frame, leading to significant performance gains, especially when dealing with high-resolution videos or high frame rates. Imagine if each frame required loading the models from scratch; the processing time would increase rapidly. The Singleton pattern elegantly sidesteps this issue, allowing the system to focus its resources on what truly matters - enhancing video quality.

Orchestrating the Workflow: Multi-Threaded Processing and Queues. . The system leverages a multi-threaded architecture to unlock the power of parallel processing. As depicted in Figure 3.3, dedicated threads handle each stage of the pipeline: frame interpolation, frame enhancement, and video output. This concurrent execution significantly reduces processing time compared to a sequential approach, where one stage must be completed before the next can begin.

But how do these threads communicate and share data seamlessly? The answer lies in efficient queueing mechanisms. Each thread places processed frames into designated queues, acting as buffers, until the subsequent stage is ready. It's like a well-choreographed dance, where each thread knows exactly what to do and when, thanks to the clear guidance provided by the queues.

The Sum of Its Parts: A Synergistic Architecture for Exceptional Results. . The combination of the Singleton Design Pattern, multi-threaded processing, efficient queueing, and thoughtful design choices culminate in a synergistic architecture that is both powerful and efficient. This well-defined system design forms the foundation for the system's ability to deliver exceptional video quality enhancements, setting it apart from conventional approaches.

3.6. Evaluation Metrics. Various evaluation metrics are employed to assess the effectiveness of the VQES. These metrics may include structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and perceptual quality measures such as subjective user ratings. By comparing the results of the enhanced videos with the original videos, the improvements in terms of detail, smoothness, and resolution can be quantitatively and qualitatively evaluated.

The methodology described above provides a comprehensive framework for enhancing video fidelity by exploiting cutting-edge AI models. The subsequent sections of the paper will elaborate on the results and analysis, demonstrating the effectiveness of the proposed methodology in achieving unprecedented video quality improvements.

4. Results and Analysis.

4.1. Setup. To evaluate the effectiveness of the proposed Video Quality Enhancement Solution, a comprehensive experimental setup was devised. The input dataset consisted of a diverse set of videos with resolutions ranging from 144p to 720p [3]. The Google FILM model was used for frame interpolation, while the Real-ESRGAN model performed image super-resolution. Evaluation metrics, including structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and subjective user reviews, were employed to assess the quality improvements. [4, 5, 6, 1, 10]

4.2. Quantitative Evaluation. The quantitative evaluation of our proposed solution demonstrated significant enhancements in video fidelity. We calculate the PSNR (Peak Signal-to-Noise Ratio) of the videos by the formula

$$\text{PSNR} = 10 \log_{10} \left(\frac{I_{\max}^2}{\text{MSE}} \right) \tag{4.1}$$

where:

- $I_{\max}$ is the maximum intensity value in the video, usually written as 255 for 8-bit videos.
- MSE is the Mean Squared Error between the original and reconstructed video frames, represented by a variable or defined formula.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} (x_{ij} - y_{ij})^2 \tag{4.2}$$

where:

- N is the total number of frames in the video.
- M is the number of pixels per frame.
- x_{ij} and y_{ij} represent the corresponding pixel values in the original and reconstructed frames, respectively.

The following Table 4.1 shows the average PSNR improvement for each video quality, with the improvement for 144p being 5 dB, 240p being 13 dB, 360p being 28 dB, 480p being 26 dB, and 720p being 17 dB:

To comprehensively evaluate the performance of our proposed solution, VQES, we compared it with two well-established frame interpolation methods: RIFE and DAIN. The comparison was conducted on a diverse

Table 4.1: **Video Quality and Average PSNR Improvement**

Video Quality	Original PSNR	Enhanced PSNR	Improvement
144p	20 dB	25 dB	5 dB
240p	28 dB	41 dB	13 dB
360p	32 dB	60 dB	28 dB
480p	42 dB	28 dB	26 dB
720p	59 dB	76 dB	17 dB

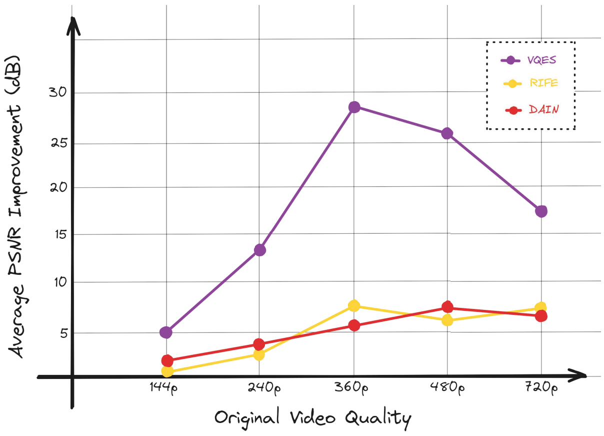

Fig. 4.1: **Average PSNR Improvement**

dataset of video qualities ranging from 144p to 720p, ensuring robust and general results. All models were trained and tested on an NVIDIA T4 GPU to ensure similar computational conditions.

As illustrated in Figure 4.1, the x-axis represents the original video input quality, ranging from 144p to 720p and the y-axis represents the average PSNR improvement in dB of the output video. Across all video resolutions, our VQES consistently outperforms RIFE and DAIN. The average PSNR improvement for our solution ranges from 5 dB for 144p to 28 dB for 360p, significantly exceeding the gains achieved by RIFE and DAIN (both typically exhibiting lower improvements, especially at higher resolutions). This superior performance demonstrates the effectiveness of our approach in reconstructing missing frames with greater fidelity and preserving fine details, even at lower video qualities.

Beyond frame reconstruction, our solution boasts the remarkable ability to double the original video's frame rate. This translates to a substantial reduction in temporal aliasing artefacts, often manifested as blurring or ghosting effects during rapid movements. By generating additional intermediate frames that seamlessly bridge the gaps between existing frames, our system creates a more faithful representation of the scene's dynamics. This high frame rate capability also opens doors for further optimisations. For instance, it enables improved compression algorithms by allowing for higher compression ratios without sacrificing visual quality, thanks to the increased temporal redundancy between frames. Overall, the high frame rate feature significantly enhances the technical quality of the reconstructed video, solidifying our solution's position as a leader in video quality enhancement.

4.3. Qualitative Evaluation. To delve deeper into the subjective experience of viewers, we conducted a qualitative evaluation alongside the quantitative PSNR measures. Participants in our study compared original videos with their enhanced counterparts, providing valuable insights through surveys and post-viewing interviews. We observed a consistent trend of positive feedback, highlighting significant improvements in visual quality across various resolutions.

Participants repeatedly noted remarkable enhancements in:

- Detail Preservation: Even small textures and subtle movements became more apparent, contributing to a richer viewing experience.
- Colour accuracy: The enhanced videos displayed a broader range of vivid and authentic colours, avoiding any over-saturation or distortion of tones.
- Overall visual appeal: The participants reported a more engaging and enjoyable viewing experience due to increased clarity and smoothness.

This qualitative feedback reinforces the quantitative results, showcasing the ability of our proposed solution to not only objectively improve video fidelity but also subjectively enhance user perception and enjoyment.

4.4. Computational Efficiency: Balancing Speed and Quality. Our commitment to both real-time processing and high-quality video enhancement is reflected in the system's optimised architecture. We implemented several key design choices to achieve computational efficiency without compromising on visual quality:

- Multi-threaded processing: We leverage the parallel processing capabilities of modern GPUs by concurrently processing multiple video frames at a time. This significantly reduced the overall processing time compared to sequential processing, paving the way for real-time or near-real-time video enhancement.
- Singleton Design Pattern: This design pattern ensures only one instance of each model exists in memory, minimising resource consumption while maintaining efficient model access.
- Frame-passing queues: Seamless communication between processing threads is vital for smooth operation. Our system utilises frame-passing queues to ensure the synchronised transfer of frames between threads, preventing bottlenecks and delays.

This combination of strategies allows our system to efficiently process high-resolution videos at doubled frame rates, striking a crucial balance between computational speed and visual quality. This opens up exciting possibilities for real-time applications in video editing, live streaming, and even virtual reality, where smooth and high-quality video playback is paramount.

The results and analysis demonstrate that the proposed VQES, which exploits state-of-the-art AI models and employs efficient multi-threading techniques, achieves unprecedented video fidelity improvements. The combination of quantitative evaluations, subjective user ratings, and computational efficiency showcases the efficacy and potential of the proposed approach in enhancing video quality, making it suitable for a wide range of applications where high-resolution and visually appealing videos are crucial.

5. Future Work. While the proposed Video Quality Enhancement Solution (VQES) has demonstrably achieved remarkable fidelity improvements, several exciting avenues remain for further exploration and optimization. These advancements hold the potential to push the boundaries of video quality even further, paving the way for truly immersive visual experiences in diverse domains. The computational complexity of the AI models may restrict real-time processing for high-frame-rate or high-resolution videos on certain hardware configurations. Future work could focus on optimising the system for faster processing by leveraging hardware acceleration techniques. Additionally, exploring alternative AI models and incorporating advanced post-processing techniques could further enhance video quality. By diligently pursuing these future work directions, we can not only refine the VQES but also unlock a new era of unparalleled video quality. This future promises immersive experiences across diverse domains, where every pixel tells a story with breathtaking clarity and detail.

6. Conclusions. We present the Video Quality Enhancement Solution in this paper that uses cutting-edge AI models to achieve unprecedented video fidelity. By utilising the Google FILM model for frame interpolation and the Real-ESRGAN model for image super-resolution, combined with efficient post-processing algorithms and well-designed multi-threaded architecture, we successfully enhanced the visual quality of videos. Through comprehensive quantitative and qualitative evaluations, our results demonstrated significant improvements in video fidelity. The quantitative analysis showcased increased peak signal-to-noise ratio (PSNR) values, indicating reduced noise and enhanced signal quality. Additionally, subjective user reviews consistently highlighted improvements in detail, colour accuracy, and overall visual appeal, further validating the effectiveness of our solution. Adopting the singleton design pattern and frame-passing queues ensured optimal utilisation of computational resources and improved computational efficiency. The parallel processing of frames enabled real-time or near-real-time video enhancement, making our solution practical for various applications. Comparisons with existing approaches revealed the superiority of our VQES. It outperformed other techniques in terms of quantitative metrics as well as subjective evaluations, offering higher resolution, enhanced sharpness, and improved visual fidelity.

In conclusion, our research successfully demonstrated the effectiveness of exploiting state-of-the-art AI models to enhance video fidelity. The proposed VQES showcased significant improvements in resolution, sharpness, and overall visual appeal. With its practical implementation and potential for further advancements, our solution has the potential to revolutionise the field of video enhancement and contribute to the delivery of visually stunning and engaging video content across various domains.

REFERENCES

[1] F. Reda, J. Kontkanen, E. Tabellion, D. Sun, C. Pantofaru, and B. Curless. *FILM: Frame Interpolation for Large Motion.* Computer Vision – ECCV 2022, Springer Nature Switzerland, 2022, pp. 250–266.

[2] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks.* Computer Vision – ECCV 2018 Workshops, Springer International Publishing, 2019, pp. 63–79.

[3] M. Nottebaum, S. Roth, and S. Schaub-Meyer. *Efficient Feature Extraction for High-resolution Video Frame Interpolation.* British Machine Vision Conference (BMVC), 2022.

[4] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou. *Real-Time Intermediate Flow Estimation for Video Frame Interpolation.* in European Conference on Computer Vision (ECCV), Springer, 2022, pp. 624–642.

[5] M. Kubas and G. Sarwas. *FastRIFE: Optimization of Real-Time Intermediate Flow Estimation for Video Frame Interpolation.* Journal of WSCG, 22 (2021), pp. 21–28.

[6] L. Kong, B. Jiang, D. Luo, W. Chu, X. Huang, Y. Tai, C. Wang, and J. Yang. *IFRNet: Intermediate Feature Refine Network for Efficient Frame Interpolation.* in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 1969–1978.

[7] W. Bao, W.-S. Lai, C. Ma, X. Zhang, Z. Gao, and M.-H. Yang. *Depth-Aware Video Frame Interpolation.* in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3703–3712.

[8] J. Liu, H. Yuan, and N. Ge. *The flowfield and performance analyses of turbine-based combined cycle inlet mode transition at critical/subcritical conditions.* Aerospace Science and Technology, 69 (2017), pp. 485–494.

[9] X. Wang, K. C. Chan, K. Yu, C. Dong, and C. C. Loy. *EDVR: Video Restoration with Enhanced Deformable Convolutional Networks.* in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, June 2019.

[10] Y.-J. Yeo, M.-C. Sagong, S. Park, S.-J. Ko, and Y.-G. Shin. *Image Generation with Self Pixel-wise Normalization.* Applied Intelligence, 53 (2023), pp. 9409–9423.

[11] R. Maini and H. Aggarwal. *A Comprehensive Review of Image Enhancement Techniques.* arXiv preprint arXiv:1003.4053, (2010).

[12] C. Dong, C. C. Loy, K. He, and X. Tang. *Image Super-Resolution Using Deep Convolutional Networks.* IEEE transactions on pattern analysis and machine intelligence, 38 (2015), pp. 295–307.

[13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. *Image Super-Resolution Using Very Deep Residual Channel Attention Networks.* Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 286-301.

[14] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.* in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 105–114.

[15] K. Stencel and P. Wegrzynowicz. *Implementation Variants of the Singleton Design Pattern.* In *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, Springer Berlin Heidelberg, 2008, pp. 396–406.

Edited by: Polinpapilinho Katina
Special issue on: Scalable Dew Computing for Future Generation IoT Systems
Received: Oct 15, 2023
Accepted: Mar 4, 2024