ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2025.0499

AI-DRIVEN REAL-TIME VIDEO ENHANCEMENT USING SUPER-RESOLUTION AND OPTICAL FLOW TECHNIQUE

Sonia Victor Soans[1], T. Prabakaran[2] and S. Vamshi Krushna[3]

[1]*Department of Information Technology, University of Technology and Applied Sciences, Sultanate of Oman*
[2]*Department of Computer Science and Engineering, Joginpally B.R. Engineering College, India*
[3]*Department of Computer Science and Engineering, Samskruti College of Engineering and Technology, India*

Abstract

Advancements in artificial intelligence have revolutionized real-time video processing, enabling enhanced visual quality for applications in surveillance, medical imaging, and entertainment. Traditional video enhancement methods often struggle with balancing computational efficiency and high-quality output, leading to degraded performance in real-time scenarios. The primary challenge lies in preserving details while reducing noise, motion artifacts, and frame inconsistencies, particularly in low-resolution and fast-motion videos. This study introduces an AI-driven real-time video enhancement framework that integrates super-resolution techniques with optical flow-based motion estimation. The proposed method employs a deep learning-based Super-Resolution Generative Adversarial Network (SRGAN) to upscale video frames while maintaining texture fidelity. Additionally, an enhanced optical flow algorithm refines motion estimation, minimizing temporal inconsistencies and improving frame transitions. The combination of these techniques enables effective noise reduction, sharper details, and smooth motion handling, making the framework suitable for real-time applications. Experimental evaluations demonstrate that the proposed approach significantly improves peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) compared to existing methods. The system achieves real-time performance with minimal computational overhead, making it suitable for deployment in live broadcasting, telemedicine, and security surveillance. The results highlight the efficiency of integrating AI-based super-resolution with optical flow in achieving superior video clarity and motion coherence in real-time environments.

Keywords:

Real-Time Video Enhancement, Super-Resolution, Optical Flow, Ai-Driven Video Processing, Motion Estimation

1. INTRODUCTION

The increasing demand for high-quality video content in various fields, including surveillance, medical imaging, and entertainment, has led to significant advancements in real-time video enhancement techniques [1-3]. Traditional video processing methods often struggle with issues such as low resolution, motion artifacts, and temporal inconsistencies, particularly in dynamic environments where real-time processing is required. Artificial intelligence (AI)-driven techniques have emerged as a promising solution, offering the ability to enhance video quality through deep learning-based super-resolution and motion estimation approaches. These AI-driven methods can restore fine details, reduce noise, and improve the perceptual quality of videos, making them ideal for applications that require both accuracy and efficiency.

Despite the progress in video enhancement, several challenges persist in achieving real-time performance without compromising quality. First, deep learning-based super-resolution techniques often require high computational power, making them unsuitable for real-time applications without significant optimization [4]. Second, motion artifacts and frame inconsistencies remain a major issue, especially in videos with rapid movements, as existing methods struggle to maintain coherence across frames [5]. Third, ensuring a balance between noise reduction and detail preservation is complex, as aggressive denoising can lead to blurring, while insufficient processing may retain unwanted distortions [6]. Addressing these challenges requires an integrated approach that combines super-resolution techniques with robust motion estimation for smooth frame transitions.

Real-time video enhancement requires a framework that can simultaneously upscale video frames, reduce noise, and maintain temporal consistency while operating within computational constraints [7]. Traditional super-resolution approaches, such as bicubic interpolation, fail to recover fine details, while deep learning-based methods often introduce artifacts when dealing with complex motion scenarios [8]. Optical flow techniques, which estimate pixel movements between frames, provide a potential solution for motion-aware enhancement but suffer from inaccuracies when dealing with occlusions and fast-moving objects [9]. The key challenge is to develop a computationally efficient AI-driven framework that integrates super-resolution with motion estimation to achieve real-time video enhancement with minimal artifacts and maximum perceptual quality [10].

1.1 OBJECTIVES

- Develop an AI-driven real-time video enhancement framework that integrates super-resolution and optical flow-based motion estimation.
- Optimize the framework for computational efficiency, ensuring real-time processing without compromising video quality.
- Evaluate the proposed approach against existing methods in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and real-time performance metrics.

The novelty of this work lies in the seamless integration of deep learning-based super-resolution with optical flow-based motion estimation for real-time video enhancement. Unlike conventional approaches that treat these techniques separately, the proposed framework leverages the strengths of both methods to improve video clarity, reduce motion artifacts, and enhance temporal consistency. Key contributions include:

- A deep learning-based Super-Resolution Generative Adversarial Network (SRGAN) optimized for real-time video upscaling.
- An enhanced optical flow algorithm that improves motion estimation accuracy, reducing temporal inconsistencies.

- A hybrid AI-driven framework that balances computational efficiency and high-quality enhancement for real-time applications.

2. RELATED WORKS

Several studies have explored AI-based video enhancement techniques, focusing on super-resolution, motion estimation, and real-time processing. Deep learning-based super-resolution methods have gained attention due to their ability to restore high-frequency details lost in low-resolution videos [7]. Among these, SRGAN has been widely adopted for video upscaling, leveraging adversarial learning to generate high-fidelity frames. Researchers have extended SRGAN with modifications such as feature extraction networks and attention mechanisms to enhance visual quality further [8]. However, these methods often suffer from computational inefficiency, limiting their applicability in real-time scenarios [9].

Optical flow-based motion estimation techniques provide an alternative approach for enhancing video clarity by predicting motion patterns between frames [10]. Traditional methods such as Farneback and Lucas-Kanade optical flow algorithms have been used for motion compensation but struggle with occlusions and fast-moving objects [11]. Deep learning-based optical flow models, such as FlowNet and RAFT, have significantly improved motion estimation accuracy, enabling smoother frame transitions and reducing motion artifacts [12]. However, these models still face challenges in real-time applications due to high computational demands.

Hybrid approaches combining super-resolution and optical flow have been proposed to leverage the strengths of both techniques. Some studies have explored the use of recurrent neural networks (RNNs) and temporal consistency constraints to refine frame transitions [13]. Others have integrated spatial-temporal attention mechanisms to enhance motion-aware super-resolution, achieving improved perceptual quality [14]. Despite these advancements, existing solutions still face trade-offs between computational efficiency and enhancement quality, highlighting the need for a more optimized framework [15].

The proposed study builds on these prior works by integrating an optimized SRGAN with an enhanced optical flow algorithm, ensuring real-time performance while maintaining high visual fidelity. By addressing the computational inefficiencies of existing models, this research contributes to the development of a practical AI-driven solution for real-time video enhancement.

3. PROPOSED METHOD

The proposed AI-driven real-time video enhancement framework integrates deep learning-based super-resolution with optical flow-based motion estimation to improve video quality while ensuring smooth frame transitions. A Super-Resolution Generative Adversarial Network (SRGAN) is employed to upscale low-resolution frames while preserving fine details and textures. To maintain temporal consistency and reduce motion artifacts, an enhanced optical flow algorithm estimates pixel movements between consecutive frames, refining motion prediction for smoother transitions. The framework operates in a multi-stage pipeline, where each frame undergoes super-resolution enhancement before optical flow refinement corrects any inconsistencies. The system is optimized for computational efficiency through parallel processing and lightweight neural network architectures, enabling real-time performance across various applications, including surveillance, medical imaging, and live broadcasting. The combination of super-resolution and motion estimation allows for noise reduction, improved frame coherence, and high-quality output with minimal latency.

- **Frame Preprocessing:** Input video frames are extracted and normalized to ensure uniform brightness, contrast, and feature consistency.
- **Super-Resolution Enhancement:** A deep learning-based SRGAN is applied to upscale each frame, generating high-resolution outputs while preserving texture details.
- **Optical Flow Estimation:** An enhanced optical flow algorithm computes pixel-wise motion vectors between consecutive frames, predicting movement patterns to maintain temporal consistency.
- **Motion Compensation:** The estimated motion vectors are used to align frames, reducing motion artifacts and improving coherence in fast-moving scenes.
- **Post-Processing and Refinement:** A lightweight CNN-based refinement module further enhances frame sharpness, corrects distortions, and ensures visual smoothness.
- **Real-Time Optimization:** Parallel processing techniques and model compression are applied to achieve low latency, enabling real-time video enhancement on hardware-constrained devices.
- **Output Generation:** The enhanced high-resolution video stream is reconstructed and output in real-time for display, storage, or further analysis.

3.1 FRAME PREPROCESSING

The first step in the proposed video enhancement method involves preprocessing the input video frames to standardize the visual properties, ensuring that the deep learning model can effectively process them. Preprocessing typically includes operations like resizing, normalization, and noise reduction. Resizing the frames to a uniform resolution ensures consistent input sizes, allowing the model to handle different video sources efficiently. Additionally, normalization of pixel values between 0 and 1 helps in stabilizing the model's training and inference, preventing issues related to scale disparities in input data. Noise reduction is performed to remove any background noise or compression artifacts present in low-resolution video frames, which can negatively affect both the super-resolution process and optical flow accuracy.

Formally, let's denote a low-resolution frame as $\mathbf{I}_{LR}$. After preprocessing, it is resized to a standardized resolution $\mathbf{I}'_{LR}$, which ensures that the frame's dimensions are consistent across all inputs:

$$\mathbf{I}'_{LR} = f(\mathbf{I}_{LR}, resize, normalize, denoise) \tag{1}$$

where, $f(\cdot)$ is the preprocessing function including resizing, normalization, and denoising.

3.2 SUPER-RESOLUTION ENHANCEMENT

After preprocessing, the super-resolution enhancement stage is applied to upscale the low-resolution frames to higher resolutions. The deep learning-based Super-Resolution Generative Adversarial Network (SRGAN) is used for this task. SRGAN employs a generator-discriminator architecture, where the generator learns to reconstruct high-resolution images, and the discriminator distinguishes between real and generated high-resolution images, improving the output's perceptual quality.

The SRGAN aims to optimize the following loss function, which balances pixel-wise accuracy and perceptual quality:

$$L = \lambda_1 \cdot L_{L2} + \lambda_2 \cdot L_{GAN} \tag{2}$$

where, $L_{L2} = \| \hat{\mathbf{I}} - \mathbf{I}_{HR} \|_2^2$ is the L2 loss, representing the pixel-wise difference between the generated high-resolution frame $\hat{\mathbf{I}}$ and the ground truth high-resolution frame $\mathbf{I}_{HR}$. L_{GAN} is the adversarial loss, which encourages the generator to produce realistic images that can fool the discriminator. λ_1 and λ_2 are hyperparameters controlling the balance between the L2 loss and GAN loss.

By training on large datasets of high-resolution and low-resolution frame pairs, SRGAN learns to upscale the input frames effectively. After applying SRGAN to the preprocessed low-resolution frame, the output is a high-resolution frame , which preserves fine details and textures.

Table.1. Super-Resolution Enhancement Results

Frame	Original Resolution	Low-Resolution (LR)	High-Resolution (HR)	PSNR (dB)	SSIM
1	1920x1080	480x270	1920x1080	28.2	0.85
2	1920x1080	480x270	1920x1080	27.5	0.83
3	1920x1080	480x270	1920x1080	29.1	0.87
4	1920x1080	480x270	1920x1080	28.7	0.84

In this table, the low-resolution frames (480x270) are enhanced using the super-resolution technique to match the original resolution (1920x1080). The PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) values indicate the quality of enhancement, showing a significant improvement over the low-resolution input frames. These metrics are crucial for evaluating the perceptual quality of the enhanced frames, with higher PSNR and SSIM values indicating better video quality post-enhancement. This process ensures that the video frames are not only upscaled to higher resolution but also exhibit improved visual fidelity and detail retention.

3.3 OPTICAL FLOW ESTIMATION

The optical flow estimation step involves calculating the pixel-level motion between consecutive video frames. The goal is to determine how each pixel moves from one frame to the next, allowing for the compensation of motion-induced distortions. The optical flow algorithm estimates motion vectors by comparing the intensity changes between consecutive frames. One of the most used methods for optical flow estimation is the Horn-Schunck method, which assumes that the intensity of each pixel remains constant between frames.

The optical flow vector $\mathbf{v} = (v_x, v_y)$ at each pixel is computed by minimizing the following energy functional:

$$E(u,v) = \iint \left((I_x u + I_y v + I_t)^2 + \alpha (|u_x| + |u_y| + |v_x| + |v_y|) \right) \tag{2}$$

where:

u and v represent the horizontal and vertical components of the optical flow.

I_x, I_y, and I_t represent the spatial derivatives of the image in the x, y, and time directions, respectively.

α is a regularization parameter that controls the smoothness of the flow.

This minimizes the difference between the observed intensity change and the motion field while penalizing large variations in motion between neighboring pixels. The resulting motion vectors **v** capture the movement of pixels between frames, forming the basis for motion compensation.

3.4 MOTION COMPENSATION

Once the motion vectors are estimated, motion compensation is applied to align the frames and reduce artifacts such as ghosting or blurring due to misaligned pixels. The goal of motion compensation is to use the optical flow information to warp the previous frame to match the current frame, correcting for pixel displacements caused by motion. Mathematically, the motion compensation process can be expressed as:

$$\hat{I}_{MC}(x,y) = I_{t-1}(x - v_x, y - v_y) \tag{3}$$

where, $\hat{I}_{MC}$ is the compensated frame. I_{t-1} is the previous frame. v_x and v_y are the motion vectors from the optical flow estimation.

This shifts the pixels of the previous frame according to the computed motion vectors (v_x,v_y)to better align the frames. This compensates for the displacement caused by motion, helping to maintain temporal consistency and reduce motion artifacts in dynamic scenes.

3.5 POST-PROCESSING AND REFINEMENT

After motion compensation, a post-processing and refinement stage is applied to further enhance the video quality.

Table.2. Motion Compensation and Refinement Results

Frame	Original Frame	Estimated Optical Flow (Magnitude)	Motion Compensated Frame	PSNR (dB)	SSIM
1	1920x1080	2.5	1920x1080	30.4	0.86
2	1920x1080	3.1	1920x1080	31.2	0.88
3	1920x1080	1.8	1920x1080	29.8	0.84
4	1920x1080	2.4	1920x1080	30.6	0.87

This step involves refining the compensated frames to improve sharpness, reduce any remaining noise, and ensure smooth transitions between frames. A lightweight CNN is employed to correct any residual artifacts, such as blurred edges or inconsistencies caused by imperfect motion estimation. The CNN uses learned features to enhance the edges and texture details while ensuring that the refined frames are coherent with the rest of the video. This post-processing step removes remaining artifacts and enhances the visual quality of the frames, ensuring that the output video is both high-resolution and temporally consistent.

In this table, the motion compensation step utilizes optical flow vectors to align the frames, reducing motion artifacts and improving frame consistency. The PSNR and SSIM values show that motion compensation enhances video quality, with increased PSNR indicating reduced distortion and higher SSIM showing improved structural similarity with the ground truth. After the post-processing and refinement, the final frames exhibit enhanced sharpness and reduced temporal artifacts, leading to a smoother, higher-quality video stream suitable for real-time applications.

3.6 REAL-TIME OPTIMIZATION

The real-time optimization step is crucial to ensure that the entire video enhancement pipeline operates with minimal latency, enabling it to function efficiently in real-time applications. Given that video enhancement tasks like super-resolution, optical flow estimation, and motion compensation can be computationally intensive, real-time performance is achieved through model optimization techniques such as parallel processing, neural network compression, and optimized hardware utilization.

To achieve low latency, we use a model compression strategy that reduces the number of parameters in the neural networks used for super-resolution and post-processing. This is done by applying techniques such as pruning (removing less significant weights), quantization (reducing the precision of weights), and knowledge distillation (transferring knowledge from a large model to a smaller one). These techniques decrease the computational overhead without significantly compromising performance.

Let **M**c represent the compressed model, and **I** denote the input video frame. The optimization process can be represented as:

$$\mathbf{I}' = \mathbf{M}_c(\mathbf{I}) \tag{4}$$

The model compression and optimization techniques reduce the model size and accelerate processing time, ensuring that the entire video enhancement pipeline can run on hardware-constrained devices, such as mobile phones or edge devices, without significant delays.

Additionally, parallel processing is implemented to handle multiple frames simultaneously. This is achieved by distributing the workload across multiple processing units (e.g., GPUs or CPUs). Given that each frame enhancement task is independent, parallelism allows for simultaneous processing of multiple frames, leading to faster Thus performance. The parallel processing equation can be expressed as:

$$\mathbf{I}^{(i)}_{output} = \mathbf{M}^{(i)}_{\text{optimized}}(\mathbf{I}^{(i)}_{input}), \quad \forall i \in [1, N] \tag{5}$$

where, $\mathbf{I}^{(i)}_{output}$ is the output of the i^{th} frame after enhancement. N is the total number of frames being processed in parallel.

3.7 OUTPUT GENERATION

The final step involves generating the output video after applying all enhancement techniques. The output generation is the process of reconstructing the video stream from the enhanced frames. Once the real-time optimization and post-processing are completed for each frame, the frames are stitched together to form the final video sequence. The enhanced video is then ready for real-time display, storage, or transmission. This final step ensures that the enhanced video maintains high resolution, motion coherence, and smooth transitions between frames, all while being processed with minimal latency.

Table.3. Real-Time Optimization and Output Generation Results

Frame	Frame Resolution	Optimized Processing Time (ms)	Compressed Model Size (MB)	PSNR (dB)	SSIM
1	1920x1080	35	10	30.4	0.86
2	1920x1080	34	10	31.2	0.88
3	1920x1080	33	9.8	29.8	0.84
4	1920x1080	36	10	30.6	0.87

In this table, the optimized processing time for each frame has been significantly reduced by model compression and parallel processing. The compressed model size is also smaller, reducing the computational overhead. The PSNR and SSIM values reflect high-quality enhancement, with a slight decrease in SSIM due to the optimization but still maintaining acceptable quality for real-time applications. This optimization process ensures that the video enhancement system can function efficiently in environments with limited computational resources, achieving high-quality results with minimal processing time and latency.

4. RESULTS AND DISCUSSION

The experiments were conducted using Python with deep learning libraries such as TensorFlow and PyTorch. For video enhancement and real-time processing, we used the OpenCV library for frame extraction, optical flow estimation, and motion compensation. The super-resolution enhancement was implemented using pre-trained deep convolutional neural networks (CNNs) available in TensorFlow and PyTorch, while optical flow estimation was performed using the Farneback algorithm. The experiments were conducted on an Intel Core i9-10900K processor with 32 GB of RAM, along with an NVIDIA Tesla V100 GPU to accelerate deep learning computations. We tested the proposed method against two existing video enhancement methods: Deep Video Super-Resolution (DVR): A method based on a deep neural network designed to enhance video resolution using temporal information. Optical Flow-based Motion Compensation (OF-MC): A traditional method that employs optical flow for motion compensation followed by frame interpolation and enhancement techniques.

Table.4. Experimental Setup/Parameters

Parameter	Value
Video Resolution	1920x1080 (Full HD)
Input Frame Rate	30 FPS
Super-Resolution Model	Deep CNN (XceptionNet-based)
Optical Flow Estimation	Farneback Optical Flow
Model Compression	Pruning, Quantization, Distillation
Motion Compensation	Motion Vector Warping

Post-Processing Method	Lightweight CNN for Refinement
CPU Used	Intel Core i9-10900K
RAM	32 GB

4.1 PERFORMANCE METRICS

4.1.1 Peak Signal-to-Noise Ratio (PSNR):

PSNR is used to measure the quality of the enhanced video by comparing the pixel differences between the original and enhanced frames. Higher PSNR values indicate better quality, as they signify fewer differences in pixel intensities. PSNR is calculated as:

$$PSNR = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (5)$$

4.1.2 Structural Similarity Index (SSIM):

SSIM evaluates the perceptual quality of the image by comparing luminance, contrast, and structure between the original and enhanced frames. SSIM ranges from 0 to 1, where 1 indicates perfect similarity. It is calculated as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (6)$$

4.1.3 Execution Time:

Execution time measures the time taken to process one frame and generate the enhanced output. This is a critical metric for real-time applications, where low latency is required. The lower the execution time, the better the system can handle real-time video enhancement.

4.1.4 Frame Rate (FPS):

Frame rate measures the number of frames processed per second. For real-time video applications, achieving a high frame rate is crucial. Higher FPS values ensure that the system can process video smoothly without lags.

Table.5. Performance Comparison

Method	PSNR (dB)	SSIM	Execution Time (ms/frame)	Frame Rate (FPS)
Proposed Method	32.1	0.91	35	28.57
DVR	30.4	0.86	50	20.00
OF-MC	31.2	0.88	45	22.22

This table shows the comparison of the proposed method with the existing methods. The proposed method achieves higher PSNR and SSIM values, indicating superior video quality. It also demonstrates faster execution time and a higher frame rate, making it more suitable for real-time applications compared to the DVR and OF-MC methods.

Table.6. PSNR Comparison

Method	Pruning	Quantization	Distillation
Proposed Method	32.5 dB	32.1 dB	31.9 dB
DVR	30.8 dB	30.4 dB	30.1 dB
OF-MC	31.2 dB	30.9 dB	30.5 dB

The proposed method shows a slight decrease in PSNR values when applying model compression techniques, with pruning resulting in a slight drop of 0.4 dB compared to the base performance (32.5 dB). Even with model compression, the proposed method still outperforms both DVR and OF-MC methods by around 1.7-2.0 dB.

Table.7. SSIM Comparison

Method	Pruning	Quantization	Distillation
Proposed Method	0.91	0.90	0.89
DVR	0.85	0.84	0.83
OF-MC	0.88	0.87	0.86

The proposed method shows consistently high SSIM scores even with compression techniques, with a slight decrease as pruning, quantization, and distillation are applied. The SSIM values for the proposed method range from 0.89 to 0.91, significantly higher than DVR (0.83-0.85) and OF-MC (0.86-0.88), indicating superior image structural preservation.

Table.8. Execution Time (ms/frame)

Method	Pruning	Quantization	Distillation
Proposed Method	36	35	34
DVR	51	50	48
OF-MC	46	45	43

The execution time for the proposed method is significantly lower compared to DVR and OF-MC methods, even after applying model compression. Pruning reduces the processing time slightly, with the proposed method achieving around 36 ms/frame. The optimized methods lead to reduced execution times compared to the baseline methods, making the proposed method more efficient.

Table.9. FPS Comparison

Method	Pruning	Quantization	Distillation
Proposed Method	27.78	28.57	29.41
DVR	19.61	20.00	20.83
OF-MC	21.74	22.22	23.26

The FPS results show that the proposed method maintains a higher frame rate even with model compression. Distillation gives the highest FPS (29.41), followed by quantization (28.57) and pruning (27.78), while DVR and OF-MC methods are slower, achieving around 19-23 FPS, demonstrating the efficiency of the proposed method in real-time processing.

5. CONCLUSION

The proposed method for real-time video enhancement using super-resolution and optical flow algorithms demonstrates significant improvements in both performance and efficiency over existing methods like Deep Video Super-Resolution (DVR) and Optical Flow-based Motion Compensation (OF-MC). Despite applying model compression techniques such as pruning, quantization, and distillation, the proposed method consistently outperforms the existing methods in terms of PSNR, SSIM, execution time, and frame rate. The results show that even with slight reductions in PSNR and SSIM values due to model compression, the proposed method maintains superior video quality and structural similarity compared to DVR and OF-MC. Moreover, the method significantly reduces execution time, allowing for real-time processing without compromising on performance, achieving higher FPS values than both DVR and OF-MC across all compression techniques. These results highlight the efficiency of the proposed approach in real-world applications where real-time video enhancement is critical. The ability to achieve high-quality video enhancement while maintaining low latency and high frame rates makes the proposed method an ideal solution for applications in fields such as live streaming, video conferencing, and surveillance systems, where both quality and real-time processing are paramount. Thus, this method successfully combines state-of-the-art techniques in super-resolution and optical flow estimation for practical, high-performance video enhancement.

REFERENCES

[1] D.V. Manjunatha, M. Maindola, R. Jose, S. Kaliappan, M. Patel and R. Maranan, “AI-Powered Multi View Face Video Super-Resolution Techniques for Real-Time Video Processing”, *Proceedings of International Conference on Integrated Intelligence and Communication Systems*, pp. 1-7, 2024.

[2] B. Deng, “The Application of AI Video Generation Technology in Virtual Reality (VR) and Augmented Reality (AR)”, *Proceedings of International Conference on Artificial Intelligence, Deep Learning and Neural Networks*, pp. 168-173, 2024.

[3] D. Narsina, N. Richardson, A. Kamisetty, J.C.S. Gummadi and K. Devarapu, “Neural Network Architectures for Real-Time Image and Video Processing Applications”, *Engineering International*, Vol. 10, No. 2, pp. 131-144, 2022.

[4] W. Shi, Q. Li, Q. Yu, F. Wang, G. Shen, Y. Jiang and G.M. Muntean, “A Survey on Intelligent Solutions for Increased Video Delivery Quality in Cloud-Edge-End Networks”, *IEEE Communications Surveys and Tutorials*, pp. 1-6, 2024.

[5] O.V. Dubolazov, O.G. Ushenko, I.V. Soltys, M.O. Ohirko, S.V. Foglinskiy, R.A. Zaplitniy and A.O. Karachevtsev, “Researching the Possibilities of using AI Technologies for Digital Image Processing: Review and Applications”, *Optics Electronic Information and Energy Technologies*, No. 2, pp. 78-87, 2024.

[6] A. Ghasemieh and R. Kashef, “Advanced Monocular Outdoor Pose Estimation in Autonomous Systems: Leveraging Optical Flow, Depth Estimation and Semantic Segmentation with Dynamic Object Removal”, *Sensors*, Vol. 24, No. 24, pp. 1-6, 2024.

[7] A. Agrahari Baniya, T.K. Lee, P. Eklund and S. Aryal, “Frame Selection using Spatiotemporal Dynamics and Key Features as Input Pre-processing for Video Super-Resolution Models”, *SN Computer Science*, Vol. 5, No. 3, pp. 1-6, 2024.

[8] G. Sheeba and M. Maheswari, “HEVC Video Quality Enhancement using Deep Learning with Super Interpolation and Laplacian Filter”, *IETE Journal of Research*, Vol. 69, No. 11, pp. 7979-7992, 2023.

[9] J. Sun, B. Yang, N. Koukourakis, J. Guck and J.W. Czarske, “AI-Driven Projection Tomography with Multicore Fibre-Optic Cell Rotation”, *Nature Communications*, Vol. 15, No. 1, pp. 1-6, 2024.

[10] A.C. Tsai, C.H. Tsou and J.F. Wang, “EESRGAN: Efficient and Effective Super-Resolution Generative Adversarial Network”, *IETE Technical Review*, Vol. 41, No. 2, pp. 200-211, 2024.

[11] Y. Liu, X. Wang, C. Yuan, J. Xu, Z. Wei and J. Nie, “AttGAN: Attention Gated Generative Adversarial Network for Spatio-Temporal Super-Resolution of Ocean Phenomena”, *International Journal of Digital Earth*, Vol. 17, No. 1, pp. 1-6, 2024.

[12] F.M. Senalp and M. Ceylan, “A New Approach for Super-Resolution and Classification Applications on Neonatal Thermal Images”, *Quantitative InfraRed Thermography Journal*, Vol. 21, No. 3, pp. 172-189, 2024.

[13] B.S. Neyigapula, “Enhancing Image-based Rendering through Intelligent Machine Learning: Realism, Immersion and Future Directions”, *International Journal of Artificial Intelligence and Machine Learning*, Vol. 3, No. 2, pp. 45-56, 2023.

[14] C.Y. Yi, N.S.M. Sin and A.N.M. Isa, “Mapping the Landscape of VFX and AI: A PRISMA-Guided Systematic Review”, *International Journal of Business and Technology Management*, Vol. 6, No. 3, pp. 114-132, 2024.

[15] Y. Xiao, Q. Yuan, J. He and L. Zhang, “Remote Sensing Image Super-Resolution via Cross-Scale Hierarchical Transformer”, *Geo-Spatial Information Science*, Vol. 27, No. 6, pp. 1914-1930, 2024.