Video Compression Optimization Techniques Using Artificial Intelligence: Review / Review Article

Assist. Prof. Amal Abbas Kadhim[1*],

Assist. Prof. Wedad Abdul Khuder Naser[2],

Lecturer Nada Abdulkareem hameed[3]

1,2 Mustansiriyah University, College of Education, Department of Computer Science, Baghdad/Iraq.

3 Al-Mansour University College, Department of Computer Science and Information Systems, Baghdad/Iraq

* Corresponding Author: abbas.amal@uomustansiriyah.edu.iq

تقنيات تحسين الضغط الفيديو باستخدام الذكاء الاصطناعي: مراجعة

استاذ مساعد امل عباس كاظم[1]،

استاذ مساعد وداد عبد الخضر ناصر[2]،

مدرس ندى عبد الكريم حميد[3]

1، 2 الجامعة المستنصرية- كلية التربية، قسم علوم الحاسبات، بغداد\العراق

3 كلية المنصور الجامعة، قسم علوم الحاسبات ونظم المعلومات، بغداد\العراق

Abstract

Compressing video is an important and essential function in today's multimedia systems as it enables the efficient storage and transmission of large volumes of video data. The proliferation of high-resolution videos that are being used in various application areas such as video streaming, video conferencing, surveillance, and autonomous systems has caused the strong demand for more efficient compression algorithms. This paper presents an in-depth review of video compression techniques with particular focus on AI methods. It also discusses traditional video coding standards including H. 264/AVC, H. 265/HEVC, and AV1, including motion estimation, transform coding quantization entropy coding, and rate-distortion optimization. The paper provides an extensive review of the latest transformative AI-based techniques which include machine learning, deep learning, and neural networks that are used to effectively enhance compression and visual quality. Lastly, the paper discusses the key challenges like computational complexity, latency, and real-time implementation, and at the same time, it suggests future research paths such as the study of lightweight AI models, hybrid compression systems, and optimization of perceptual quality.

Keywords: Video Compression, Artificial Intelligence, Machine Learning, Deep Learning, Optimization Techniques, H.264/AVC, HEVC, AV1, Rate–Distortion Optimization, Multimedia Systems, Neural Compression.

المستخلص

يعد الضغط الفيديو وظيفة مهمة وأساسية في أنظمة الوسائط المتعددة اليوم لأنه يتيح تخزين ونقل كميات كبيرة من بيانات الفيديو بكفاءة. أدى انتشار مقاطع الفيديو عالية الدقة التي يتم استخدامها في مجالات التطبيقات المختلفة مثل بث الفيديو ومؤتمرات الفيديو والمراقبة والأنظمة المستقلة إلى زيادة الطلب على خوارزميات الضغط الأكثر كفاءة. تقدم هذه الورقة مراجعة متعمقة لتقنيات ضغط الفيديو مع التركيز بشكل خاص على أساليب الذكاء الاصطناعي. ويناقش أيضًا معايير ترميز الفيديو التقليدية.

H. 264/AVC, H. 265/HEVC, and AV1,

بما في ذلك تقدير الحركة، وتحويل الترميز، وترميز الإنتروبيا، وتحسين معدل التشويه. تقدم الدراسة مراجعة شاملة لأحدث التقنيات التحويلية القائمة على الذكاء الاصطناعي والتي تشمل التعلم الآلي والتعلم العميق والشبكات العصبية التي تُستخدم لتحسين الضغط والجودة البصرية بشكل فعال . أخيرًا، تناقش الدراسة التحديات الرئيسية مثل التعقيد الحسابي، وزمن الوصول، والتنفيذ في الوقت الفعلي، وفي الوقت نفسه، تقترح مسارات بحثية مستقبلية مثل دراسة نماذج الذكاء الاصطناعي خفيفة الوزن، وأنظمة الضغط الهجينة، وتحسين جودة الإدراك الحسي.

الكلمات المفتاحية: الضغط الفيديوي، الذكاء الاصطناعي، التعلم الالي ،التعلم العميق، تقنيات التحسين، رقم المعيار الذي وضعه الاتحاد الدولي للاتصالات، ترميز الفيديو عالي الكفاءة، الإصدار الأول للفيديو لتحالف الإعلام المفتوح، تحسين معدل التشويه، انظمة الوسائط المتعددة، الضغط العصبي.

1. Introduction

The surge in digital video content creation has made efficient storage and transmission solutions highly desirable. Since video data usually results in very large files, it takes up a lot of bandwidth and storage if not compressed. Therefore, video compression has turned out to be an essential element of today's multimedia systems which support various applications including streaming services, video conferencing, surveillance systems, autonomous vehicles, and edge computing devices [1].

Traditional video compression methods are mainly concerned with removing spatial and temporal redundancies present in the video sequences. The major standard video codecs including H. 264/AVC, H. 265/HEVC, and AV1 mainly rely on combining several processes such as motion estimation, transform coding quantization entropy coding, and rate-distortion optimization to provide high compression efficiency. With significant enhancements in compression performance, these methods have succeeded in maintaining the visual quality within an acceptable range. However, some of the issues that continue to hamper these methods include increased computational requirements, long encoding times, and limited capability to handle different and changing video content [1], [2], [3].

Recently, artificial intelligence (AI) has revolutionized the field of video compression research. Machine learning and deep learning approaches facilitate data-driven coding of video content which results in more efficient compression than traditional methods. These AI approaches have shown better compression efficiency, visual quality, and redundancy reduction that surpass the performance of conventional codecs [4], [5].

In spite of these developments, the introduction of AI-driven compression in real-time systems continues to be a major challenge mainly as a result of high computational necessities and hardware restrictions. In fact, it is crucial to carry out a thorough examination of both conventional and AI-based video compression methods in order to grasp the present scenario, difficulties, and prospects for the future.

2. Literature Review: Video Compression Optimization Techniques

Several researchers have carried out quite recent research in video compression optimization techniques. The studies below are sorted alphabetically by the surname of the first author. (Ananthanarayanan, *et al.*, 2019).[6] took an in-depth look at video analytics in edge computing environments. They pointed out in their research that one of the core elements of achieving minimal latency and making AI inference in real-time at the distributed system level comes from optimizing video compression at the edge. (Bankoski, *et al.*, 2018).[3] presented the AV1 video codec created by the Alliance for Open Media. In terms of term of compression performance efficiency, AV1 greatly surpasses HEVC but this comes at the trade-off of higher encoding complexity.

(Chen, *et al.,* 2024).[7] explored the use of generative AI models for video compression. They demonstrated that generative networks are capable of synthesizing visually high-quality video even from very low-bitrate inputs, which makes them a very attractive choice for compression systems of the future. (Ding, *et al.*, 2021).[8] carried out a very thorough investigation of the different neural network models for video compression. They also acknowledged the fact that the neural network-based systems are able to take over the various traditional encoding processes, like prediction and transformation, thereby yielding better compression efficiency.

(Huang and Wu, 2024).[9] raised the concept of an image and video compression framework that is learned end-to-end. The authors' work proved that neural networks are capable of surpassing the performance of classical codecs in rate-distortion terms through experiments. (Klink, 2021).[10] investigated techniques for selecting codecs adaptively that are suitable for environments with limited bandwidth. This paper emphasized using dynamic bitrate adjustments for preserving video quality in the case of unstable network conditions.

(Liu, *et al.*, 2025).[11] have surveyed model compression techniques such as pruning and quantization, which are crucial for the deployment of AI-based video compression models on edge devices. (Mazumdar, *et al.*, 2024).[12] concentrated on perceptual video compression systems. Their work showed that adjusting compression according to human visual perception can greatly enhance perceived quality at low bitrates.

(Masykuroh, *et al.,* 2025). [13] focused their studies on video compression optimization in relation to video analytics applications. According to them, adaptive compression is the way to go, as it not only improves detection accuracy but also enhances bandwidth efficiency in a smart surveillance system. Besides, (Mochurad ,2024). [14] studied the efficacy of machine learning-based compression against traditional codecs. He foundthat AI-based approach provides good perceptual quality yet similar to them, it needs high computational resources .

(Saberi, *et al.,* 2024). [15] conducted a thorough review of classical video compression techniques and outlined the shortcomings of traditional transform-based coding methods when it comes to ultra-high-definition video content . (Sullivan, *et al.*, 2012). [2] came up with the HEVC standard that significantly enhanced compression efficiency by about 50% over H. 264 with the help of improved block partitioning and motion prediction techniques.

(Wiegand, *et al.*, 2003). [1] were the first to propose the H. 264/AVC standard that has since evolved into the basis for the majority of modern video compression systems because of its high compression efficiency and global popularity .(Zhang, *et al.*, 2021). [16] explored several ways to optimize the coding of perceptual video and confirmed that perceptual models lead to better visual quality while the bitrate is only slightly increased.

According to (Zhou and Yang , 2024). [17] compressive sensing for video compression was the issue in their paper. They found that the sparse representation techniques which reduce redundancy also maintain the quality of the reconstruction.

3. Comparison Study of Video Compression Optimization Techniques:

Table 1 provides a comparative summary of previous studies on video compression optimization techniques. These details comprise the study identification writer's year of publication, research theme method main contributions, and drawbacks. The comparison points out the progression of video compression standards and methods, concentrating especially on block-based hybrid coding techniques, which are recognized for their high compression efficiency but also their limitations in handling high-resolution video content in today's world.

Table 1: Comparison Study Of Video Compression Optimization Techniques

No.	Author(s)	Year	Study Focus	Methodology	Key Contribution	Limitation
1	Wiegand, *et al.* [1]	2003	H.264/AVC standard	Block-based hybrid coding	High compression efficiency, global standard	Limited adaptability for modern high-res video
2	Sullivan, *et al.* [2]	2012	HEVC (H.265)	Coding Tree Units (CTU) + RDO	~50% bitrate reduction vs H.264	High computational complexity
3	Bankoski *et al.* [3]	2018	AV1 codec	Advanced block partitioning	Open-source, high compression efficiency	Very slow encoding speed

No.	Author(s)	Year	Study Focus	Methodology	Key Contribution	Limitation
4	Saberi *et al.* [15]	2024	Classical compression review	Comparative analysis	Overview of transform-based coding	Limited AI integration
5	Zhang *et al.* [16]	2021	Perceptual compression	Human visual modeling	Improved visual quality at low bitrate	PSNR mismatch with perception
6	Zhou & Yang [17]	2024	Compressive sensing	Sparse signal reconstruction	Reduced redundancy in video signals	Sensitive to noise
7	Klink [10]	2021	Adaptive codec selection	Network-based optimization	Better streaming adaptation	Limited AI usage
8	Ding *et al.* [8]	2021	Deep learning compression	Neural network-based encoding	End-to-end learning compression	High training cost
9	Mochurad [14]	2024	ML vs traditional codecs	Experimental comparison	AI improves perceptual quality	High computational demand
10	Mazumdar *et al.* [12]	2024	Perceptual compression	Human vision-based optimization	Better low-bitrate performance	Limited real-time use
11	Huang & Wu. [9]	2024	End-to-end compression	Deep neural networks	Better rate-distortion performance	Requires large datasets
12	Chen *et al.* [7]	2024	Generative compression	GAN-based modeling	High visual quality reconstruction	Training instability
13	Liu *et al.* [11]	2025	Model compression for AI codecs	Pruning & quantization	Enables edge deployment	Accuracy loss trade-off
14	Masykuroh *et al.* [13]	2025	Video analytics compression	Adaptive compression for AI systems	Improved analytics efficiency	System complexity
15	Ananthanarayanan *et al.* [6]	2019	Edge video analytics	Distributed systems	Reduced latency at edge	Limited compression focus

4. AI-Based Video Compression Techniques and Advanced Optimization Methods

Artificial Intelligence (AI) is one of the most useful technologies of the current time that is being used to improve the effectiveness of video compression significantly by implementing adaptive and data-driven optimization strategies. Unlike conventional compression methods that depend on fixed rules, AI-based approaches get trained on massive datasets to adjust encoding parameters in an optimal way dynamically [4], [8], [9].

4.1 Deep Learning-Based Compression

Deep learning methods, especially Convolutional Neural Networks (CNNs), have greatly advanced the extraction of spatial features from video frames. Besides, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are commonly utilized to model the temporal relationships between successive video frames. Such models facilitate more precise motion prediction and removal of redundancies, which can potentially result in enhanced compression effectiveness as well as visual quality over traditional codecs [8], [9], [10]. A number of recent works have reported that deep learning-driven methods surpass conventional codecs in terms of compression efficiency and visual quality, particularly in challenging and high-resolution video scenarios [9], [10].

4.2 Content-Aware Video Compression

Content-aware compression methods adjust the way they encode a video after recognizing its content. For instance, a video part with a lot of movement is compressed differently than a part depicting a still scene. AI systems can examine the video material and distribute bits appropriately, leading to better compression results. This approach yields higher PSNR and lower bitrate than conventional methods, thanks to its capability to keep significant visual features and discard redundant data [13], [16].

4.3 Neural Video Compression

Neural video compression is a change of concept in which deep learning models that work from end-to-end get rid of the old style encoding pipelines. Such systems automatically learn compact representations of video content, enabling them to compress at higher ratios while maintaining better perceptual quality. Even if neural compression techniques show very promising results, they need very large training datasets and significant computational power, which is a big impediment for their real-time deployment [9], [18].

4.4 Adaptive Bitrate Optimization Using AI

The adaptive bitrate technique will adapt streaming video quality on the fly based on network conditions. AI algorithms come with a prediction of changes in network capacities and can distribute bitrates in an optimized manner at any moment. This method reduces the negative effects and improves the overall experience of watching online videos by lessening rebuffering events and giving a steady level of picture quality [11], [19].

4.5 Computational Complexity Reduction

High computational complexity is one of the main issues in AI-based video compression. To deal with that drawback, the latest research has mainly been geared towards accelerating encoding, especially by model compression techniques with pruning and quantization as prime examples. Not only do these methods result in significant improvements from human perception point of view but also make it possible to bring AI-based video compression systems to the edge devices with very limited computational power [11], [19].

4.6 Evaluation Metrics and Theoretical Background

Several fundamental mathematical metrics and theoretical models are used in most of the literature for evaluating, analyzing the video compression performance. These metrics provide a quantitative measure of compression efficiency and visual quality. Compression ratio (CR) is a measure used to quantify how well a video is compressed. It is defined as:

$$CR = \frac{\text{Original Size}}{\text{Compressed Size}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Mean Squared Error (MSE) is one of the most common criteria, which is used to calculate how far apart the original and the reconstructed video frames are:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j) - K(i,j))^2 \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Peak Signal-to-Noise Ratio (PSNR) is one of the most popular measures for the quality assessment of the reconstructed videos:

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Rate-Distortion Optimization (RDO) is a fundamental idea in the current video coding standards allowing the trade-off between the bit rate and the distortion:

$$J = D + \lambda R \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Entropy coding being the most widely used technique for reducing statistical redundancies in video data is mathematically expressed as:

$$H = -\sum p(x)\log_2 p(x) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Such metrics are crucial for the comparative evaluation of conventional and AI-driven video compression methods and they have a great impact on the direction of recent multimedia research.

5. Conclusions

Video compression optimization techniques are very important to be able to effectively store, send, and handle digital video data in today's multimedia systems. In this article, both conventional and AI-based methods of video compression were thoroughly examined, with an emphasis on their advantages, disadvantages, and latest technological developments.

The conventional video coding standards like H. 264/AVC, H. 265/HEVC, and AV1 have shown great results in the areas of compression efficiency and standardization. They are based on the implementation of highly-regarded signal processing methods such as motion estimation, transform coding quantization entropy coding, and ratedistortion optimization. However, the downside of these codecs is their performance being hindered by the high computational complexity and the lack of flexibility to cope with various and changing video content.

AI-based video compression technologies designed by deep neural networks, generative models, and other deep learning methods have brought a great revolution in video compression research. They are changing the compression techniques completely as a result of their ability to reproduce as well as human perception, high compression efficiency, and therefore low bitrate requirements. Apart from that, these techniques have several other advantages such as being very flexible, greatly enhancing visual quality, and requiring data-driven optimization. Nevertheless, AI-based video compression solutions are expected to be confronted with issues like high computing power requirements, incapability for real-time rendering, and lack of universal standardization.

The review of previous works shows that no one method traditional or AI-based is able to fully solve this issue. On the contrary, mixtures of conventional codecs and AI-enhanced optimization modules are the ones paving the way to the future, most potential-wise. To sum up, video compression is likely to be dominated by smart, flexible, and hybrid compression methods that achieve a good trade-off between compression rate, image quality, and computational burden. Developing small-sized AI models, boosting real-time performance, and covering the distance between objective measurement and human visual perception are the main areas in which further research is needed.

6. Future Work

One of the directions for future research could be building lighter AI models which can process video compression in no time. Besides, the combination of conventional codecs and deep learning methods in a single workflow, i. e. hybrid approaches can be a great way to enhance the efficiency and at the same time, decreases the computational cost of the deep learning methods. There is a need for further efforts to improve perceptual quality and reduce latency in edge and cloud-based systems as well.

Acknowledgement

The authors would like to thank Al-Mustansiriyah University (www. uomusiriyah.edu.iq), Baghdad-Iraq for its support in the present work.

References

[1] G. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, (2003), "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand ,(2012), "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668.

[3] J. Bankoski *et al.* , (2018), "AV1: A video codec for the next generation of video delivery," *Alliance for Open Media*.

[4] T. Chen *et al.*, (2022), "A survey on deep learning-based video compression," *IEEE Access*, vol. 10, pp. 100–130.

[5] Z. Wang, A. Bovik, (2024), "Neural network-based video compression and enhancement techniques," *IEEE Access*, vol. 12, pp. 4500–4515.

[6] G. Ananthanarayanan *et al.*, (2019), "Video analytics in edge computing systems," *ACM/IEEE Edge Computing*.

[7] B. Chen, X. Zhang, (2024), "Generative AI for video compression: A review," *IEEE Access*, vol. 12, pp. 2024–2035, 2024.

[8] D. Ding *et al.*, (2021), "Deep learning-based video compression techniques: A survey," *Signal Processing: Image Communication*, vol. 95.

[9] C. Huang and J. Wu., (2024), "End-to-end learned image and video compression," *IEEE Transactions on Image Processing*, vol. 33, pp. 2024–2038.

[10] J. Klink, (2021), "Adaptive bitrate streaming and codec selection strategies," *Sensors*, vol. 21, no. 5.

[11] D. Liu *et al.*, (2025), "Model compression techniques for deep neural networks: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36.

[12] A. Mazumdar *et al.*, (2024), "Perceptual video compression: Human visual system-based optimization," *IEEE Access*, vol. 12.

[13] K. Masykuroh *et al.*, (2025), "Adaptive video compression for intelligent video analytics," *IEEE Access*, vol. 13.

[14] L. Mochurad, (2024), "Machine learning-based video compression versus traditional codecs," *Multimedia Tools and Applications*.

[15] Y. Saberi *et al.*, (2024), "A comprehensive review of classical video compression techniques," *Journal of Visual Communication and Image Representation*.

[16] Y. Zhang *et al.*, (2021), "Perceptual video coding optimization techniques," *IEEE Transactions on Image Processing*, vol. 30.

[17] J. Zhou and J. Yang, (2024), "Compressive sensing approaches for video compression," *Signal Processing*.

[18] S. Liu, H. Li, (2025), "Lightweight neural networks for real-time video compression," *IEEE Access*.

[19] L. Mochurad, (2024), "Machine learning-based video compression versus traditional codecs," Multimedia Tools and Applications.