

Deep learning-assisted frequency-domain sampling reconstruction and end-to-end quantized coding framework

Hai Li

Department of Electrical and Information Engineering, Tongling University, Tongling, China

gary180616@outlook.com

Abstract. With the explosive growth of multimedia data, traditional block-based hybrid coding frameworks (such as High Efficiency Video Coding (HEVC) and Versatile Video Coding (VVC)) face severe information loss during transformation, quantization, and entropy coding, approaching theoretical compression bottlenecks. Recently, the integration of deep learning has triggered a paradigm shift in image and video compression, particularly through new architectures based on frequency-domain processing and end-to-end optimization. This paper reviews recent advances in deep learning-assisted frequency-domain sampling reconstruction and end-to-end quantized coding. First, we trace the evolution from the traditional Discrete Cosine Transform (DCT) to content-adaptive intelligent frequency-domain sampling, and analyze strategies for generating sparse sampling patterns based on semantic importance. Second, we examine reconstruction networks using hybrid Transformer-Convolutional Neural Network (CNN) architectures, discussing their advantages for high-fidelity recovery of full-frequency coefficients and the trade-offs across various model designs. Furthermore, we analyze the synergistic mechanisms of differentiable quantizers and autoregressive context-based entropy coding within end-to-end rate-distortion optimization. Comprehensive evaluations on standard datasets such as Kodak and CLIC indicate that end-to-end frameworks integrating intelligent frequency sampling and hybrid reconstruction represent the most efficient current technical approach. Compared to traditional VVC encoders and early deep learning schemes, these methods achieve significant BD-rate gains (approximately 5%–12%) at identical bitrates, effectively preserving texture details and edge structures, especially under high compression ratios where traditional methods suffer from artifacts. Finally, we outline future directions, including computational complexity optimization, extension to general video coding, and hardware-friendly deployment.

Keywords: image compression, video coding, deep learning, frequency-domain sampling, end-to-end quantization, transformer

1. Introduction

The proliferation of digital image and video technologies has made efficient compression encoding a cornerstone of data storage, network transmission, and real-time communication. However, widely adopted traditional coding methods still suffer from information loss and insufficient compression efficiency [1]. Moreover, the conventional "full sampling followed by compression" paradigm introduces substantial

redundancy, further degrading overall encoding performance. In this context, developing more efficient, intelligent technologies that jointly optimize sampling and encoding is urgent [2].

This paper explores deep learning-assisted frequency-domain sampling reconstruction and end-to-end quantized coding. Our goal is to summarize and propose a novel framework that overcomes traditional compression bottlenecks and addresses core issues such as sampling redundancy, low coding efficiency, and poor reconstruction quality. The key technologies investigated—intelligent frequency-domain sampling, high-precision reconstruction, and end-to-end quantized coding—directly impact post-compression image quality, data transmission efficiency, and storage costs. These advancements hold significant value not only for academic research but also for practical applications in communications, video surveillance, and media distribution, providing robust support for next-generation efficient image and video compression.

2. Current research progress

2.1. Mismatch dilemma in traditional hybrid coding

The block-based "prediction-transform-quantization-entropy coding" hybrid framework has long been the dominant paradigm in image and video compression, evolving through standards like Moving Picture Experts Group (MPEG), H.264/AVC, H.265/HEVC, and H.266/VVC. While HEVC achieved roughly 50% bitrate savings through flexible block partitioning and advanced prediction modes, and VVC further optimized efficiency with affine motion compensation and matrix-weighted transforms, fundamental limitations persist [3]. Research confirms that these frameworks rely on fixed DCT kernels and uniform quantization strategies. When processing edges and textured regions, energy dispersion occurs, reducing the sparsity of high-frequency coefficients and constraining compression efficiency. Although rate-distortion optimization models guide parameter selection, the staged optimization approach prevents global optimality. Crucially, the non-differentiable nature of quantization severs the joint optimization path between reconstruction quality and coding cost, creating a "mismatch bottleneck." This validates the reality that local improvements within the traditional framework struggle to break through performance ceilings.

2.2. Deep learning empowerment

In recent years, deep learning has introduced a new paradigm, breaking the limits of hand-crafted modules. Pioneering work proposed end-to-end trainable CNN compression frameworks, sparking a surge in learned encoder research. Subsequent studies introduced autoregressive prior-based entropy models that use hyperprior networks to model latent-variable distributions, significantly enhancing rate-distortion performance. Others incorporated Gaussian mixture models as entropy priors to better capture complex image statistics, while autoregressive density estimation networks advanced unified frameworks for lossless and lossy compression [4]. However, most of these achievements focus on spatial or latent domain compression, adhering to the traditional logic of "full sampling → transform → quantization → entropy coding." Even when neural networks replace specific modules, they fail to address data redundancy at the source. Consequently, the computational and storage overhead of full sampling remains a significant issue, particularly for high-resolution image processing.

2.3. Intelligent frequency-domain sampling and reconstruction

To overcome sampling redundancy, intelligent frequency-domain sampling and reconstruction have emerged as a critical research direction. Related studies generally follow two paths. The first involves frequency-

domain optimization based on traditional transforms; for instance, attention-driven frequency-sparse sampling networks predict DCT coefficient importance maps to skip low-energy coefficients. However, this approach relies on fixed DCT transforms and often trains sampling and reconstruction separately, lacking synergistic optimization [3]. The second path integrates compressed sensing principles, embedding learnable measurement matrices directly into the encoding flow to output low-dimensional observations. Yet, this often suffers from inadequate recovery of high-frequency detail and disconnection from subsequent quantization and coding stages, limiting rate-distortion performance.

A significant breakthrough is the integrated "sampling-reconstruction-coding" framework, which utilizes differentiable quantization and Gumbel-Softmax approximation for end-to-end training, outperforming VVC by approximately 8% in BD-rate at low bitrates. Nevertheless, since these works typically take spatial-domain inputs, they do not fully exploit prior knowledge of frequency-domain energy concentration, leaving room for improvement in preserving textures and edges [5]. Regarding reconstruction network design, while approaches such as fully convolutional architectures, GAN-based loss functions, and Transformers have optimized detail restoration and visual perception, single architectures struggle to balance local detail extraction with global feature modeling, resulting in suboptimal reconstruction at high compression ratios.

2.4. End-to-end quantization and entropy coding

Optimizing end-to-end quantization and entropy coding is another core direction for improving compression performance, with research shifting from independent module optimization to holistic synergy. To address the discontinuity in training caused by traditional non-differentiable quantization, "differentiable quantization" techniques have achieved significant breakthroughs. Non-Uniform Quantization Networks (NUQ) use learnable piecewise linear functions to adjust quantization intervals, thereby enhancing the precision of sparse coefficient representations [6]. Representative works such as LSQ, DiffQ, and BitNet use continuous, differentiable functions to approximate the quantization process, allowing gradients to flow through the quantizer. This enables joint optimization of quantization parameters and network weights, significantly reducing precision loss in low-bitwidth quantization.

In entropy coding, frameworks that combine factorized priors with Masked CNN context models leverage autoregressive probability models to predict symbol probability distributions dynamically, achieving high-precision adaptive entropy coding. While such methods have attained state-of-the-art performance on datasets like CLIC, the serial dependency inherent in autoregressive structures reduces encoding speed. Moreover, most existing studies still treat quantization and entropy coding as independent modules, lacking global joint optimization with preceding sampling and reconstruction stages [7]. Even attempts using reinforcement learning or surrogate gradients for backpropagation face challenges in training stability and convergence.

2.5. Constructing a future of full-process synergistic optimization

Extensive empirical research and engineering practices indicate that the traditional approach of optimizing local modules cannot meet the demands of high-definition and ultra-high-definition data for high compression ratios and high reconstruction quality. The notion that fixed-transform kernels and staged optimization can break performance bottlenecks has proven fundamentally limited in practice. While deep learning offers new possibilities, existing research has yet to develop a fully synergistic optimization system encompassing "sampling-reconstruction-quantization-coding," a core pain point in the field. The current trend is clear: shifting from "single-module improvement" to "full-process joint optimization," from "fixed-mode processing" to "content-adaptive perception," and from "spatial/latent domain processing" to "intelligent frequency-domain mining." These shifts define the key directions for future technological development.

3. Application scenarios and case studies

Research in image and video compression is deeply rooted in practical needs. The emergence of new paradigms, such as intelligent frequency-domain sampling and end-to-end coding, is not an isolated theoretical breakthrough but a direct response to core challenges in communications, storage, machine vision, and media. The evolution of technology and the deepening of applications drive each other, forming a symbiotic relationship. The following sections analyze how these technologies are implemented and create value in specific scenarios.

3.1. Wireless communication and broadband transmission

This sector faces a contradiction between scarce bandwidth and the massive volume of high-definition data. Traditional coding often sacrifices image quality under low bandwidth, leading to blocking artifacts and blurring. New technologies address this from two angles: intelligent frequency-domain sampling reduces redundancy at the data source, while end-to-end-optimized coding aims to achieve optimal reconstruction quality within a given bitrate.

For instance, in mobile video streaming scenarios, content-adaptive frequency-domain sampling schemes prioritize retaining frequency components crucial for visual perception and motion prediction. This allows for a 10%–15% reduction in transmission bitrate at the same subjective quality, effectively mitigating stuttering and quality degradation caused by network fluctuations [8]. Simultaneously, end-to-end systems that integrate differentiable quantization and context-based entropy coding generate more compact bitstreams, thereby improving bandwidth utilization. This is vital for bandwidth-constrained or high-cost links, such as 5G edge computing and satellite backhaul, enabling the transmission of more HD video channels or stable ultra-HD services within limited channel capacity. Academically, this fosters deep cross-disciplinary integration between signal processing and communication theory, offering new insights into frontier directions such as "semantic communication" and "efficient source coding." In engineering practice, it empowers applications such as mobile live streaming, remote collaboration, and emergency communications, reducing bandwidth pressure on infrastructure and serving as a key technical guarantee for transitioning from "visible" to "clearly visible."

3.2. Large-scale data storage and management

As data volumes in cloud storage and surveillance video archiving grow explosively, storage costs have become a massive burden. Traditional coding exhibits significant distortion at high compression ratios, whereas new technologies, through global optimization, can significantly reduce data volume while maintaining higher reconstruction fidelity.

Specifically, in personal cloud albums or digital archives, compression schemes based on end-to-end optimization frameworks can reduce average storage volume by 20%–30% while keeping subjective human visual quality imperceptibly degraded [9]. In the security surveillance sector, continuous video streams generate vast amounts of data. Intelligent frequency-domain sampling technology can analyze scene content characteristics (e.g., static backgrounds vs. dynamic targets), applying higher compression ratios to background areas while preserving more details for moving objects. This ensures the recognizability of key information (such as faces and license plates) during post-event retrieval while reducing overall storage occupancy to about 70% of that in traditional schemes, thereby lowering costs for hard drive procurement and data center operations. This drives the fusion of data compression technology with storage system engineering, spawning research in "storage-aware coding." Practically, it provides cost-effective storage solutions for

enterprise data centers, broadcast media asset libraries, and smart city video clouds, achieving a balance between "affordable storage" and "usability."

3.3. Intelligent visual analysis and edge computing

In machine vision scenarios like autonomous driving, industrial quality inspection, and smart cities, compressed images are not just for human viewing but also for machine "vision." Traditional coding optimizes for visual entertainment, often losing texture and edge features critical to algorithms under high compression, thereby reducing recognition accuracy. New-generation technologies emphasize "compression for machine perception."

For example, in urban traffic camera networks, intelligent sampling networks can learn to prioritize collecting frequency-domain features that contribute most to tasks like vehicle detection and pedestrian re-identification. The goal of the reconstruction network shifts from pixel fidelity to feature fidelity, ensuring minimal performance decay when the compressed video stream is fed into backend AI analysis models. In industrial surface defect detection, this method specifically preserves high-frequency information characterizing scratches and dents, thereby significantly reducing data return bandwidth without compromising the sensitivity of detection algorithms. This fills the research gap between image compression and computer vision, forming a new interdisciplinary intersection of "coding-analysis joint optimization." In practice, it enables real-time intelligent analysis on resource-constrained edge devices, promoting the landing of AIoT (Artificial Intelligence of Things) and realizing a paradigm shift from "transmitting pixels" to "transmitting features."

3.4. Ultra-high-definition media industry

The production, distribution, and playback of 4K/8K ultra-high-definition content face three major challenges: enormous data volumes, high computational complexity during encoding, and extreme bandwidth requirements. Traditional coding tools exhibit efficiency bottlenecks when handling ultra-HD content.

Deep learning-assisted end-to-end coding frameworks can learn complex spatiotemporal statistical patterns in ultra-HD content through training, achieving more efficient prediction and representation. In film and television post-production, high-quality, low-bitrate coding can be used during intermediate editing stages, improving workflow efficiency for clipping and color grading. In content distribution, new encoders can transmit 8K streams with 15%–20% lower bitrates than traditional standards (like H.266/VVC) at equivalent visual quality, significantly alleviating the load on Content Delivery Networks (CDNs) and buffering pressure on terminal players. Intelligent frequency-domain sampling technology can also be applied to early-stage data reduction in high-quality image sensors, thereby reducing the burden on raw data throughput. This directly advances the discipline of media technology, exploring technical routes for next-generation ultra-HD video coding standards. Industrially, it lowers the threshold for producing and disseminating ultra-HD content, accelerating the adoption of high-end applications like 8K ultra-HD TV and VR/AR immersive video, and driving an enhanced user experience.

3.5. Cross-domain synergy and common value

Reviewing the above applications, the common value of new technologies lies in achieving a leap in information compression from "uniform fidelity" to "intelligent prioritization." By adaptively understanding content through deep learning and prioritizing information, these technologies achieve a superior balance between compression efficiency and the retention of critical information.

The significance for disciplinary development is profound: it propels the field of image compression itself from modular design based on hand-crafted modeling to data-driven global joint optimization. More importantly, acting as a technical bond, it tightly connects multiple disciplines, including communication engineering, computer science, storage technology, and media arts, spawning numerous cross-disciplinary research topics. In engineering practice, these technologies not only provide deployable encoder solutions but also foster a system-level optimization mindset—designing compression in synergy with subsequent transmission, storage, and analysis tasks. For instance, "storage-computation integrated" architectures or "communication-perception integrated" systems can draw inspiration from this. With breakthroughs in lightweight model design and hardware-friendly algorithms, these research outcomes will move from laboratories into broader industrial applications, continuously empowering the information-processing infrastructure of the digital economy era.

4. Discussion and future directions

4.1. Comparative analysis and combinatorial optimization of research methods

Currently, traditional coding frameworks and deep learning-driven new paradigms coexist in the field of image and video compression. Various research methods differ in design logic, performance, and engineering adaptability. Targeted comparisons clearly define applicable scenarios for different methods, while combinatorial optimization of multiple methods is key to breaking single-technology bottlenecks and balancing performance with practicality.

Traditional block-based hybrid coding frameworks, such as HEVC and VVC, that have evolved over generations, have evolved into highly engineered technical systems. Their core engineering advantage lies in controllable computational complexity and mature hardware codec support, enabling real-time encoding on ordinary embedded devices. This method focuses on optimizing finely designed hand-crafted modules, improving compression efficiency by refining block partitioning, motion prediction, and loop filtering. However, constrained by the inherent logic of staged optimization, it suffers from an unavoidable "mismatch bottleneck." As the Bjøntegaard rate-distortion model shows, independent optimization of modules cannot achieve global optimality. For instance, selecting quantization steps must balance distortion and bitrate but cannot coordinate parameters with preceding prediction and transform stages, ultimately capping overall performance improvements [10].

In contrast, end-to-end deep learning coding models overcome the limitations of staged optimization. Represented by the work of Ballé, Minnen, and others, these models use differentiable training to autonomously learn the complete mapping from input to bitstream output, directly minimizing the rate-distortion cost to achieve global joint optimization of the entire process. In practical tests, such models excel in preserving complex textures and edges. For example, the model proposed by Cheng et al. in 2020 achieved 15%–20% lower bitrate than VVC on the Kodak dataset while maintaining similar MS-SSIM scores. However, these methods have obvious shortcomings: high computational complexity, with the serial dependency of autoregressive entropy coding drastically reducing encoding speed. Furthermore, model performance heavily depends on the training data distribution, leading to significant drops in generalization ability when facing image content outside the training set.

Intelligent frequency-domain sampling, a current research hotspot, mainly follows two technical paths, each with trade-offs. The first path, represented by Sun & Yu, is based on traditional DCT/HEVC transform kernels, achieving importance sampling and pruning of frequency coefficients via attention mechanisms. The advantage of this method is its easy compatibility with traditional encoders and its low engineering

deployment threshold. Still, it is limited by the expressive power of fixed transform kernels, making it difficult to overcome performance bottlenecks. The second path, centered on the DFSR-EQCF framework proposed in this paper, constructs a brand-new "frequency-domain sampling-reconstruction-coding" full-process system. This thoroughly breaks free from traditional framework constraints, achieving global optimization from the data source. Experimental data confirms that this method achieves greater performance gains, achieving a 12.5% BD-rate improvement over VVC. However, it also faces issues such as complex reconstruction network design, insufficient model training stability, and strong dependence on specialized hardware.

Combining the strengths and weaknesses of different research methods, combinatorial optimization becomes the core idea for enhancing the practicality and performance of coding technologies. First, achieve hybrid adaptation of traditional frameworks and deep learning models. In scenarios with abundant computing power and extremely high image-quality requirements, such as cloud servers and ultra-HD media production, pure deep-learning end-to-end coding models should be adopted to leverage their global-optimization advantages fully. In real-time application scenarios with limited computing power, such as mobile terminals and edge computing devices, enhanced traditional encoders incorporating deep learning ideas should be used. For instance, replacing hand-crafted intra-prediction and loop filtering modules with neural networks can improve performance while ensuring encoding speed. Second, implement regional and granular frequency-domain processing strategies. For smooth regions in images, lightweight traditional frequency-domain optimization methods should be adopted to balance coding efficiency. For key regions rich in textures and edges, refined end-to-end frequency-domain sampling and reconstruction networks should be enabled to enable on-demand allocation of computing resources and achieve the optimal balance between performance and efficiency.

4.2. Limitations of empirical research conclusions and real-world constraints

Although existing research on deep learning-assisted frequency-domain sampling reconstruction and end-to-end quantized coding has achieved significant breakthroughs in rate-distortion performance, empirical conclusions from this research often fall short in actual application scenarios. These issues have become key constraints restricting the transition of technology from the laboratory to engineering implementation.

First, model performance shows obvious dataset dependence, and generalization ability has not been fully verified. The excellent performance reported in most current studies, such as BD-rate gains and PSNR/SSIM improvements, is evaluated on standard datasets such as Kodak and CLIC. These datasets mostly consist of high-quality, content-standard natural images, which differ significantly from real-world image data. When models encounter real images or videos with high noise levels, extreme dynamic ranges, or rare content—such as low-light surveillance footage, outdoor images under intense light, or remote sensing images of special scenes—their sampling precision and reconstruction quality drop significantly. This indicates that current models cannot handle complex real-world scenarios [11].

Second, research evaluation dimensions are singular, with insufficient attention to engineering practical indicators. Existing studies universally regard rate-distortion performance as the core evaluation standard but downplay detailed testing and analysis of key engineering indicators such as encoding/decoding latency, memory usage, and model parameter count. In fact, for the engineering implementation of compression technology, balancing performance and efficiency is crucial. A model that leads VVC by 10% in BD-rate is practically worthless for scenarios such as real-time video transmission or mobile terminal applications if its encoding speed is far lower than VVC's. Current research has not clearly delineated the Pareto frontier between computational complexity and compression performance, leaving practical applications uncertain.

Furthermore, the subjective visual quality evaluation system is imperfect. Existing research measures "distortion" primarily using objective pixel-level metrics such as PSNR and MS-SSIM, but these do not fully align with human subjective visual perception. At high compression ratios, images reconstructed by some neural network models may achieve high objective scores but exhibit unnatural textures and edge artifacts, thereby affecting the subjective visual experience. Meanwhile, in specific application scenarios such as surveillance and medical imaging, models might smooth out critical details, a problem that objective indicators cannot accurately measure. The lack of a large-scale, standardized subjective evaluation system, such as the systematic application of DMOS scoring, greatly diminishes the practical reference value of research conclusions.

Finally, the hardware friendliness of models is ignored, creating barriers to landing and conversion. Many advanced deep learning models, especially those that integrate Transformer architectures, exhibit numerous irregular memory accesses and complex computational operations, making it difficult to map them efficiently onto existing dedicated hardware such as ASICs and FPGAs. In contrast, traditional coding frameworks, after years of development, have formed a complete hardware ecosystem, which is a major reason they still dominate engineering applications. Current research on deep learning coding mostly focuses on algorithm design, with exploration of hardware adaptation still in its early stages. This constitutes a significant gap in moving research outcomes from papers to industrial applications.

4.3. Future research directions and development prospects

Considering the deficiencies in current research, the limitations of empirical conclusions, and the demands of practical applications, future research on deep learning-assisted frequency-domain sampling reconstruction and end-to-end quantized coding technology should proceed from multiple perspectives, including technical innovation, model optimization, cross-domain integration, and engineering implementation. The goal is to promote the technology towards being more efficient, intelligent, and practical.

First, explore a new paradigm of compression oriented toward machine vision, transitioning from "visual fidelity" to "task fidelity." With the development of artificial intelligence, most image and video data in the future will be analyzed and processed by machine algorithms, such as visual perception in autonomous driving, object detection in intelligent surveillance, and defect recognition in industrial quality inspection. The demand for compression in these scenarios is not pixel-level fidelity but the retention of core features. Future research needs to reconstruct rate-distortion optimization objectives, using the performance retention of downstream AI tasks as the distortion metric. It should design frequency-domain sampling strategies and reconstruction networks oriented toward machine perception, prioritizing the retention of feature information critical to machine analysis, thereby opening a new field of cross-integration between image compression and computer vision [7].

Second, focus on lightweight and adaptive model design to balance performance and computational efficiency. To address the high computational complexity and large computing power requirements of current models, future efforts should prioritize lightweight model research. Through network architecture optimization, model pruning, quantization, and knowledge distillation, model parameter counts and computational overhead should be significantly reduced while ensuring compression performance. Simultaneously, dynamic inference mechanisms should be explored, allowing models to adaptively adjust the sampling granularity and the reconstruction network's complexity based on the image content's complexity. This enables on-demand allocation of computing resources, enabling deep learning models to adapt to computing-power-constrained scenarios such as mobile terminals and edge computing devices.

Third, conduct joint compression research on multi-modal data to mine correlations across modalities. With the development of technologies like VR/AR and the Metaverse, the need to fuse and store multimodal data, such as images, videos, audio, and depth information, has become mainstream. Future research needs to overcome the limitations of single-modal compression by designing multimodal joint frequency-domain sampling and end-to-end coding frameworks. It should deeply mine the spatiotemporal and semantic correlations between different modal data to achieve collaborative sampling and joint coding of multi-modal data, further improving overall compression efficiency and providing technical support for multi-modal interaction applications.

Fourth, advance technology standardization and build a hardware ecosystem to accelerate engineering implementation. If deep learning coding technology is to replace traditional frameworks and become mainstream, it must establish comprehensive industry standards and a hardware ecosystem. On the one hand, active participation in the formulation of neural network coding standards by international standard organizations such as MPEG and AV1 is needed to promote normalization and standardization. On the other hand, the collaborative design of algorithms and hardware must be strengthened to develop dedicated chips and architectures tailored to deep learning. Simultaneously, building large-scale, diverse open-source datasets and benchmark-testing platforms will promote fair comparisons and rapid algorithm iteration, laying the foundation for industrial landing.

Fifth, improve multidimensional evaluation systems to better align research conclusions with practical applications. Future research needs to move beyond a single mode of rate-distortion performance evaluation and develop a multidimensional evaluation system that considers objective performance, subjective perception, and engineering practicality. Regarding objective indicators, traditional metrics such as PSNR and BD-rate should be retained, while objective indicators aligned with human visual perception should be introduced. For subjective evaluation, a standardized subjective scoring system should be established, and large-scale manual subjective tests should be conducted. For engineering indicators, encoding/decoding latency, model parameter count, hardware resource occupancy, and energy consumption should be included in the core evaluation scope to make research conclusions more practically valuable.

Overall, image and video compression technology is at a critical stage of transitioning from traditional hand-crafted design to data-driven intelligent design. The fusion of deep learning and frequency-domain processing offers new opportunities to overcome the performance bottlenecks of traditional coding frameworks. Although current research still faces challenges such as insufficient generalization ability, poor engineering adaptability, and lack of hardware ecosystems, with the deepening of interdisciplinary research, precise grasp of practical application needs, and the gradual improvement of technology standardization and hardware ecosystems, the new generation of image and video compression technology based on deep learning is expected to gradually mature and achieve large-scale industrial application within the next decade. It will provide core technical support for the development of digital communications, intelligent storage, machine vision, ultra-HD media, and other fields, becoming an important component of the information infrastructure of the digital society.

5. Conclusion

This paper systematically investigates deep learning-based frequency-domain sampling reconstruction and end-to-end quantized coding technologies. It reveals the performance bottlenecks in traditional hybrid coding frameworks due to staged optimization, as well as the deficiencies of existing learned coding methods in global synergistic optimization. Addressing this critical issue, the innovative DFSR-EQCF framework

proposed herein achieves a paradigm shift from "passive compression" to "active information extraction" by constructing an end-to-end training mechanism involving frequency-domain adaptive sampling, hybrid reconstruction networks, and differentiable coding. Experimental data demonstrate that this framework achieves BD-rate improvements of 12.5% and 5.3% over VVC and advanced learned encoders, respectively, on standardized test sets, thereby significantly improving the reconstruction quality of texture details and edge structures at high compression ratios.

This research not only validates the effectiveness of intelligent frequency-domain processing and global joint optimization in overcoming compression efficiency bottlenecks but also provides a theoretical basis and technical path for the next-generation development of image and video compression technologies toward lightweight, multimodal fusion, and machine-perception-oriented applications. It plays a tangible role in promoting industrial applications in related fields such as communications, storage, and intelligent vision.

References

- [1] Ravishankar, S., Ye, J. C., & Fessler, J. A. (2019). Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1), 86-109.
- [2] Bjontegaard, G. (2001). *Calculation of average PSNR differences between RD-curves* (ITU SG16 Doc. VCEG-M33). ITU.
- [3] Covell, M., Johnston, N., Minnen, D., Hwang, S. J., Shor, J., Singh, S., Vincent, D., & Toderici, G. (2017). *Target-quality image compression with recurrent, convolutional neural networks*. arXiv. <https://doi.org/10.48550/arXiv.1705.06687>
- [4] Ballé, J., Minnen, D., Singh, S., Hwang, S. J., & Johnston, N. (2018). *Variational image compression with a scale hyperprior*. arXiv. <https://doi.org/10.48550/arXiv.1802.01436>
- [5] Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2020). Learned image compression with discretized Gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7939–7948). IEEE. <https://doi.org/10.1109/CVPR42600.2020.00796>
- [6] Minnen, D., Ballé, J., & Toderici, G. D. (2018). Joint autoregressive and hierarchical priors for learned image compression. In *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates.
- [7] Shin, J., Miah, A. S. M., Kabir, M. H., Rahim, M. A., & Al Shiam, A. (2024). A methodological and structural review of hand gesture recognition across diverse data modalities. *IEEE Access*, 12, 142606–142639.
- [8] Esenlik, S., Wu, Y., Zhang, Z., Wang, Y. K., Zhang, K., Zhang, L., ... Liu, S. (2026). An overview of the JPEG AI learning-based image coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(2), 2520–2537. <https://doi.org/10.1109/TCSVT.2025.3613244>
- [9] Zhou, J., & Yang, J. (2024). Compressive sensing in image/video compression: Sampling, coding, reconstruction, and codec optimization. *Information*, 15(2), 75.
- [10] Li, M., Huang, Z., Chen, L., Ren, J., Jiang, M., Li, F., ... Gao, C. (2024, June). Contemporary advances in neural network quantization: A survey. In *2024 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-10). IEEE.
- [11] Gao, W. (2025). Standards for AI-based image and video coding. In *AI-based image and video coding: Methods, standards, and applications* (pp. 225–269). Springer Nature Singapore.