

Review Article

Recent Advances in Image Compression Techniques: From Traditional Methods to Deep Learning Models

Akhilendra Kumar Pandey

Student, Rajagiri School of Engineering & Technology, Kochi, India

I N F O

E-mail Id:

akhilendrapandey38@gmail.com

How to cite this article:

Pandey A K. Recent Advances in Image Compression Techniques: From Traditional Methods to Deep Learning Models. *J Adv Res Image Proc Appl* 2026; 9(1): 19-24.

Date of Submission: 2026-04-21

Date of Acceptance: 2026-05-29

A B S T R A C T

Image compression is a fundamental and critical component of modern multimedia and communication systems, enabling efficient storage, processing, and transmission of visual data across various platforms such as mobile applications, cloud storage, streaming services, and remote sensing systems. The rapid growth of high-resolution imaging technologies has increased the demand for advanced compression techniques that can reduce data size while preserving perceptual image quality.

Traditional image compression techniques such as JPEG, JPEG2000, and various transform-based coding methods have been widely used for decades. These approaches rely on mathematical transformations including Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), quantization, and entropy coding to reduce redundancy in image data. Although these methods are computationally efficient and well-established, they often face limitations in achieving an optimal balance between compression ratio and visual quality, particularly in complex images with rich textures and fine details. Additionally, artifacts such as blocking and blurring are commonly observed at high compression levels.

In recent years, deep learning-based image compression techniques have emerged as a powerful alternative to traditional methods. Approaches based on Convolutional Neural Networks (CNNs), autoencoders, variational autoencoders (VAEs), and transformer architectures have demonstrated significant improvements in compression efficiency and reconstructed image quality. These data-driven models learn compact latent representations of images and optimize compression directly based on reconstruction loss and perceptual metrics, leading to superior performance compared to handcrafted transform-based methods.

Keywords: Image compression, JPEG, JPEG2000, deep learning, autoencoder, CNN, lossy compression, lossless compression, neural compression.

Introduction

With the exponential increase in digital image generation from diverse sources such as social media platforms, medical imaging systems, surveillance networks, autonomous vehicles, and remote sensing satellites, the need for efficient image compression has become more critical than ever. High-resolution imaging devices continuously produce large volumes of data, making storage, transmission, and processing increasingly challenging. Image compression addresses this issue by reducing the number of bits required to represent an image while maintaining acceptable visual quality and preserving essential structural information.

The primary objective of image compression is to eliminate redundancy in image data, which includes spatial redundancy, spectral redundancy, and psychovisual redundancy. Effective compression techniques aim to achieve a balance between compression ratio and reconstructed image quality, ensuring minimal perceptual loss.

Traditional compression standards such as JPEG and JPEG2000 have been widely adopted in real-world applications for several decades. These methods rely on transform-based coding techniques, including the Discrete Cosine Transform (DCT) in JPEG and wavelet transforms in JPEG2000. While these approaches are computationally efficient and standardized, they are fundamentally limited by fixed mathematical transformations and hand-engineered encoding pipelines. As a result, they often fail to adapt effectively to diverse image content and complex visual structures.

In contrast, recent advances in deep learning have introduced data-driven image compression methods that learn optimal representations directly from large-scale datasets. These methods leverage neural networks to jointly optimize encoding and decoding processes, significantly improving compression efficiency and reconstructed image quality. Deep learning-based compression frameworks have demonstrated superior performance in terms of rate-distortion trade-offs, visual fidelity, and adaptability to complex image distributions.^{4,6}

Traditional Image Compression Techniques

Traditional image compression techniques are broadly categorized into lossy and lossless compression methods, each serving different application requirements based on acceptable levels of information loss.

Lossy Compression

Lossy compression techniques reduce file size by selectively removing less perceptually important information from images. These methods are widely used in applications

where slight quality degradation is acceptable in exchange for significantly reduced storage requirements.

JPEG (DCT-Based Compression)

JPEG is one of the most widely used image compression standards. It operates by dividing the image into small blocks (typically 8×8 pixels) and applying the Discrete Cosine Transform (DCT) to convert spatial pixel values into frequency components. High-frequency components, which are less perceptible to the human eye, are quantized more aggressively, resulting in data reduction.

However, at higher compression ratios, JPEG often introduces visible blocking artifacts due to block-wise processing.

JPEG2000 (Wavelet-Based Compression)

JPEG2000 improves upon JPEG by using Discrete Wavelet Transform (DWT) instead of block-based DCT. This allows for multi-resolution representation of images, enabling better scalability and improved visual quality, especially at low bitrates. It also reduces blocking artifacts but may still suffer from computational complexity and ringing effects in some cases.

Lossless Compression

Lossless compression techniques ensure exact reconstruction of the original image without any loss of information. These methods are essential in domains where data integrity is critical.

Common lossless formats include:

- PNG (Portable Network Graphics): Uses predictive coding combined with entropy coding techniques.
- GIF (Graphics Interchange Format): Utilizes Lempel-Ziv-Welch (LZW) compression.
- JPEG-LS: A lossless/near-lossless compression standard designed for high efficiency and simplicity.
- Lossless compression is widely used in medical imaging, scientific visualization, and technical documentation, where even minor distortions are unacceptable.³

Limitations of Traditional Methods

Despite their widespread use, traditional compression techniques suffer from several inherent limitations:

- Dependence on fixed mathematical transformations that do not adapt to image content
- Limited ability to model complex, high-dimensional image distributions
- Visible artifacts such as blocking (JPEG) and ringing (JPEG2000)
- Reduced performance at high compression ratios
- Lack of perceptual optimization aligned with human visual perception

These limitations have motivated the development of more adaptive and intelligent compression approaches using deep learning.

Deep Learning-Based Image Compression

Deep learning has introduced a paradigm shift in image compression by enabling end-to-end optimization of encoding and decoding processes. Unlike traditional methods, these approaches learn compact representations directly from data.

Autoencoder-Based Compression

Autoencoders are neural network architectures designed to learn efficient data representations in an unsupervised manner. In image compression, they consist of two main components:

- **Encoder:** Maps input images into a compact latent representation
- **Decoder:** Reconstructs the image from the compressed latent space

These models are trained using a rate-distortion optimization function, which balances compression efficiency (bitrate) with reconstruction quality (distortion). Autoencoder-based methods have demonstrated superior performance compared to traditional codecs in terms of PSNR and perceptual quality.¹⁶

CNN-Based Compression Models

Convolutional Neural Networks (CNNs) enhance image compression by learning spatial hierarchies and extracting meaningful features from image content. Unlike fixed transforms such as DCT, CNNs adaptively learn representations based on image characteristics.

Key advantages include:

- Improved feature extraction capability
- Content-aware adaptive compression
- Higher PSNR and SSIM compared to JPEG and JPEG2000
- Better preservation of edges and textures

CNN-based compression systems have shown 10–30% bitrate savings compared to JPEG2000 while maintaining similar or better visual quality.¹⁶

Transformer-Based Compression

Transformer-based models represent a recent advancement in image compression. These models use self-attention mechanisms to capture long-range dependencies across image regions.

Advantages include:

- Superior modeling of global image context
- Improved representation of complex textures and structures

- Better scalability for high-resolution images
- State-of-the-art performance in several benchmarks

However, transformers typically require higher computational resources and large-scale training datasets.

3.4 Learned Lossy Compression

Learned lossy compression refers to end-to-end trainable neural compression systems that directly optimize the trade-off between bitrate and distortion. These models integrate encoding, quantization, and decoding into a unified learning framework.

Key features include:

- End-to-end differentiable training pipelines
- Adaptive bitrate control mechanisms
- Learned quantization and entropy models
- Optimization based on perceptual loss functions

These approaches consistently outperform traditional codecs such as JPEG and JPEG2000 in terms of both objective metrics (PSNR, SSIM) and perceptual quality.^{4,17}

4. Hybrid Compression Techniques

Hybrid image compression techniques integrate the strengths of traditional compression methods with modern deep learning-based approaches to achieve a better balance between compression efficiency, computational complexity, and reconstructed image quality. The main motivation behind hybrid systems is to leverage the computational efficiency and standardization of classical codecs while enhancing their performance using data-driven neural networks that improve perceptual quality and reduce compression artifacts.

One common approach is JPEG combined with neural enhancement, where images are first compressed using the standard JPEG pipeline and then passed through a deep learning-based post-processing network. This neural enhancement model, often based on CNNs, is trained to remove compression artifacts such as blocking, ringing, and blurring. As a result, the final output exhibits significantly improved visual quality without modifying the existing JPEG standard, making this approach highly practical for real-world deployment.

Another widely used strategy is Wavelet transform combined with CNN-based refinement. In this method, wavelet-based compression (such as JPEG2000-like decomposition) is first applied to reduce redundancy across multiple frequency bands. A CNN model is then used to refine the reconstructed image by enhancing edges, textures, and fine details. This combination improves both structural preservation and perceptual quality, especially at low bitrates.

A more advanced hybrid framework involves neural entropy models integrated with classical codecs. In this approach,

deep learning models are used to estimate probability distributions of latent image representations, enabling more efficient entropy coding. These learned probability models significantly improve bit allocation and reduce redundancy compared to traditional fixed entropy coding schemes. When combined with conventional compression pipelines, this results in higher compression ratios with minimal loss in visual quality.

Comparison of Compression Techniques

Compression techniques in image processing can be broadly compared based on their underlying methodology, compression efficiency, reconstruction quality, and computational complexity. Traditional methods such as JPEG and JPEG2000 rely on fixed mathematical transformations like DCT and wavelet transforms, making them

computationally efficient and easy to implement, but they often suffer from visible artifacts and limited adaptability to complex image structures. In contrast, deep learning-based compression methods, including autoencoders, CNN-based models, and transformer architectures, learn data-driven representations that significantly improve rate-distortion performance and produce higher visual quality, especially for high-resolution and complex images. However, these methods typically require higher computational resources and large training datasets. Hybrid techniques attempt to bridge this gap by combining classical codecs with neural networks, achieving a balance between efficiency and quality. While traditional methods remain widely used in resource-constrained systems, deep learning approaches are increasingly dominating applications where compression quality and adaptability are critical.

Table 1. Comparative Analysis of Image Compression Methods

Method Type	Technique	Strengths	Limitations	Applications
Traditional Lossy	JPEG	Fast, simple, widely supported	Blocking artifacts, lower quality	Web images
Traditional Lossy	JPEG2000	Better quality than JPEG	High complexity	Medical imaging
Lossless	PNG / JPEG-LS	Exact reconstruction	Larger file size	Medical, technical images
Deep Learning	CNN/Autoencoder	High compression efficiency	Requires training data	Multimedia, AI systems
Deep Learning	Transformers	Global feature modeling	High computation cost	Advanced compression research

Applications of Image Compression

Image compression plays a vital role in modern digital systems where large volumes of visual data need to be efficiently stored, transmitted, and processed. It is widely used across multiple domains, each with different requirements in terms of speed, quality, and computational efficiency.

In social media platforms, image compression is essential for optimizing storage and enabling fast image sharing over networks with varying bandwidth conditions. Platforms such as Instagram, Facebook, and WhatsApp rely heavily on compression algorithms to reduce image size while maintaining acceptable visual quality for end users.

In medical imaging systems, compression is used in formats such as DICOM to store and transmit large diagnostic images including MRI, CT, and X-ray scans. However, medical applications require a careful balance between compression ratio and diagnostic accuracy, as even minor loss of detail can impact clinical decision-making. Therefore, lossless or near-lossless compression techniques are often preferred in this domain.

In satellite and remote sensing applications, image compression is crucial due to the massive volume of data generated by Earth observation systems. Efficient compression enables faster transmission of data from satellites to ground stations and supports real-time environmental monitoring and analysis.

Video streaming services such as YouTube and Netflix rely on image and video compression techniques to deliver high-quality content with minimal buffering, adapting dynamically to network conditions using adaptive bitrate streaming.

In mobile and edge devices, compression is used to reduce storage requirements and improve processing efficiency, especially in resource-constrained environments such as smartphones, drones, and IoT devices.

Overall, image compression is a foundational technology that enables efficient handling of visual data across diverse real-world applications.

Challenges in Image Compression

Despite significant advancements in both traditional and deep learning-based compression techniques, several challenges remain unresolved.

One of the primary challenges is the trade-off between compression ratio and image quality. Increasing compression levels often leads to degradation in visual quality, introducing artifacts and loss of fine details, which can be critical in sensitive applications such as medical imaging.

Another major challenge is the high computational cost of deep learning-based models. Neural compression systems, especially those based on CNNs and transformers, require significant processing power and memory resources for both training and inference, limiting their deployment in real-time or resource-constrained environments.

The requirement for large training datasets is also a key limitation. Deep learning models depend heavily on diverse and high-quality datasets to learn effective representations, which may not always be available for specialized domains.

Additionally, there is a lack of universal standards for neural image compression, unlike traditional codecs such as JPEG and JPEG2000. This makes interoperability and widespread adoption more difficult.

Finally, real-time deployment constraints remain a significant issue, particularly for applications such as live video streaming, mobile imaging, and edge computing, where low latency and high efficiency are critical requirements.

Future Directions

Future research in image compression is expected to focus on developing more intelligent, efficient, and adaptable systems that overcome the limitations of current approaches.

One major direction is the development of fully end-to-end learned compression systems, where encoding, quantization, and decoding are jointly optimized using deep neural networks. These systems aim to surpass traditional codecs in both efficiency and perceptual quality.

Transformer-based compression models are also gaining significant attention due to their ability to capture long-range dependencies and global image context, leading to improved compression performance for complex and high-resolution images.

Another important area is the design of lightweight neural codecs for mobile and edge devices, enabling real-time compression with reduced computational requirements through techniques such as model pruning, quantization, and knowledge distillation.

Research is also progressing toward lossless neural compression techniques, which aim to achieve exact reconstruction while maintaining high compression efficiency using probabilistic modeling and neural entropy coding.

Finally, adaptive bitrate streaming using artificial intelligence is emerging as a key research direction, where compression algorithms dynamically adjust quality and bitrate based on network conditions and user requirements, improving overall streaming performance.

Conclusion

Image compression has evolved significantly over the past decades, transitioning from traditional transform-based techniques such as JPEG and JPEG2000 to modern deep learning-based approaches. Traditional methods remain widely used due to their simplicity, efficiency, and standardization, but they are limited in their ability to adapt to complex image structures and often introduce visible compression artifacts at high compression ratios.

In contrast, deep learning-based compression methods, including CNNs, autoencoders, and transformer-based models, have demonstrated substantial improvements in compression efficiency, reconstruction quality, and adaptability. These data-driven approaches learn optimal representations directly from images, enabling better rate-distortion performance and enhanced visual fidelity.

Hybrid approaches that combine classical codecs with neural networks provide a practical compromise between efficiency and quality, making them suitable for real-world deployment. Looking ahead, transformer-based architectures and fully learned compression systems are expected to define the next generation of image compression technologies.

In conclusion, while traditional compression techniques continue to play an important role in current systems, deep learning-based and hybrid methods represent the future of image compression, offering superior performance, adaptability, and potential for integration into advanced multimedia and communication systems.

References

1. I. Jamil, "Review of Image Quality Assessment Methods for Compressed Images," *MDPI Journal of Imaging*, 2024.
2. A. Abd-Alzhra et al., "Image Compression Using Deep Learning Methods," *Iraqi Journal of Science*, 2022.
3. A. K. Singh et al., "Lossless Image Compression Techniques: A Survey," *MDPI Symmetry*, 2019.
4. Engineering Applications of AI, "Learning-driven Lossy Image Compression: A Survey," 2023.
5. Z. Cheng et al., "Deep Convolutional Autoencoder for Image Compression," *arXiv*, 2018.
6. G. Ma et al., "TinyLIC: Efficient Image Compression Method," *arXiv*, 2024.
7. J. Zhang et al., "Unified Learning-Based JPEG

- Compression," arXiv, 2023.
8. Z. Cheng et al., "Learned Lossless Image Compression," arXiv, 2020.
 9. S. Lefkimmatis, "Universal Denoising Networks," CVPR, 2018.
 10. K. He et al., "Deep Residual Learning for Image Recognition," CVPR, 2016.
 11. J. Ballé et al., "End-to-End Optimized Image Compression," ICLR, 2017.
 12. J. Minnen et al., "Joint Autoregressive and Hierarchical Priors for Learned Image Compression," NeurIPS, 2018.
 13. G. Toderici et al., "Variable Rate Image Compression with Recurrent Neural Networks," CVPR, 2016.
 14. O. Rippel et al., "Real-Time Adaptive Image Compression," ICML, 2017.
 15. T. Chen et al., "Convolutional Autoencoders for Compression," IEEE TIP, 2020.
 16. Z. Cheng et al., "Deep Autoencoder-Based Image Compression," arXiv, 2018.
 17. S. Mentzer et al., "High-Fidelity Neural Image Compression," NeurIPS, 2020.
-