International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EFFICIENT LOSSY IMAGE COMPRESSION AND RECONSTRUCTION WITH DCT-OPTIMIZED RESIDUAL COMPRESSION FRAMEWORK

[1]Rani Nandkishor Aher, [2]Dr. Mandaar B. Pande, [3]Nandkishor Daulat Aher, [4]Amit Ashok Aware

[1]Symbiosis Centre for Information Technology (SCIT),

Symbiosis International (Deemed University) (SIU),

Hinjewadi, Pune-411057, Maharashtra, India.

email: aher.rani@gmail.com

[2]Symbiosis Centre for Information Technology (SCIT),

Symbiosis International (Deemed University) (SIU),

Hinjewadi, Pune-411057, Maharashtra, India.

email: mandaar@scit.edu

[3]BITS WILP,

Birla Institute of Technology & Science, Pilani (Raj)

BITS, Pilani, Rajasthan-333031, India

Email: aher.kishor@gmail.com

[4]Symbiosis Centre for Information Technology (SCIT),

Symbiosis International (Deemed University) (SIU),

Hinjewadi, Pune-411057, Maharashtra, India.

email: amit_aware_shegaon@yahoo.co.in

Abstract

Image compression reduces file size while preserving visual quality, improving storage and transmission efficiency. Traditional lossy methods struggle at low bitrates, introducing artifacts that degrade image quality. While deep learning offers potential solutions, its adoption is hindered by challenges like training instability, sensitivity to hyper parameters, and mode collapse. To address these challenges, this research proposes a novel DCT-Optimized Residual Compression Framework (DCT-RCF) that integrates frequency domain transformations with an enhanced deep learning architecture. The framework leverages the Frequency-Enhanced residual Autoencoder (FERA-Net) which combines residual learning and an autoencoder with Discrete Cosine Transform (DCT) to efficiently separate essential visual structures from redundant spatial information. The proposed approach enhances compression efficiency by encoding high impact frequency components while discarding insignificant details, ensuring high quality reconstructions with minimal artifacts. The reconstruction process utilizes Inverse Discrete Cosine Transform (IDCT) and transposed convolution to restore spatial domain details while minimizing reconstruction errors. The framework is evaluates using Kodak and CLIC datasets with performance metrics including Mean squared Error, Peak Signal to Noise Ratio, Compression Ratio and Multi-Scale Structural Similarity Index Measure. Experimental results demonstrate that the proposed DCT-RCF achieves a compression ratio of 44.26:1 significantly outperforming the traditional and DL based compression methods by achieving PSNR of 42.38dB and MS-SSIM of 0.999 ensuring superior image reconstruction quality even at low bitrates making it highly effective solution for advanced image compression applications.

Keywords: Image Compression, Discrete Cosine Transform, Residual Learning, Autoencoder, Deep Learning, Lossy Compression, Reconstruction Quality.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

Image compression plays a vital role in digital media processing, primarily aimed at minimizing file size without compromising perceptual image quality. As the volume of digital content continues to rise rapidly, the need for optimized storage solutions and faster image transmission becomes increasingly important particularly in areas like multimedia applications, web platforms, and telecommunications [1, 2]. High-resolution images typically demand substantial bandwidth and storage, making efficient compression techniques critical for enhancing both system performance and user satisfaction. The effectiveness of any compression strategy lies in its ability to reduce data size while retaining the visual integrity of the image, ensuring smooth transmission with minimal resource consumption [3, 4].

Over time, a wide range of image compression methods has been introduced, generally falling into two main categories: lossless and lossy techniques. Lossless compression retains the complete integrity of the original image data, making it particularly suitable for scenarios where precise image recovery is essential such as in medical diagnostics or digital archiving. However, its ability to reduce file size is limited [5, 6]. In contrast, lossy compression achieves significantly higher compression ratios by selectively discarding image details that are less perceptible to the human eye, making it the preferred choice for real-world applications like streaming, photography, and social media [7, 8]. Despite its advantages, lossy compression still struggles to maintain high visual quality, particularly at low bitrates, where compression artifacts such as blurring, blocking, and color banding become prominent [9, 10].

Traditional lossy compression techniques, such as JPEG, work by transforming images into the frequency domain, quantizing coefficients, and discarding high-frequency components. While effective, these methods struggle with preserving fine details and maintaining high reconstruction quality at lower bitrates [11]. This results in severe degradation in visual quality, particularly in images with complex textures and intricate features. Furthermore, conventional compression techniques fail to adapt dynamically to varying image content, leading to suboptimal results in diverse real-world datasets. These shortcomings underscore the urgent need for more efficient and adaptive compression methods that can overcome the limitations of traditional approaches and optimize compression efficiency while preserving image fidelity [12].

The increasing complexity and diversity of modern image datasets further amplify the limitations of traditional compression techniques. Real-world images vary significantly in content, resolution, and quality, making it challenging for conventional models to handle this variability effectively [13]. The fundamental challenge lies in balancing compression efficiency (data reduction) and image quality (visual integrity preservation). This trade-off becomes particularly critical at low bitrates, where maximum compression is required while still maintaining perceptual quality. Existing methods struggle to achieve this balance, often resulting in either excessive data loss or suboptimal compression ratios [14].

Deep learning (DL) has recently emerged as a promising solution for overcoming these challenges. Unlike traditional methods that rely on handcrafted features and fixed algorithms, DL models can learn optimal compression strategies by extracting hierarchical feature representations [15]. Convolutional Neural Networks (CNNs) and Autoencoder have been used to enhance feature extraction, reduce redundant spatial information, and improve compression quality. However, despite these advancements, DL-based image compression still faces critical challenges such as training instability, Hyperparameter sensitivity, and mode collapse [16]. Additionally, existing DL-based methods struggle to efficiently separate essential structural details from irrelevant high-frequency components, limiting their ability to optimize image reconstruction quality. These challenges highlight the need for a more refined and stable DL model tailored for image compression.

By addressing these challenges, this research proposes a DL framework which enhances image compression by first transforming the images into frequency domain, where essential details are preserved and redundant information is efficiently reduced. This is followed by a novel FERA-Net which leverages residual learning and skip connections to enhance feature selection and image reconstruction. By combining traditional frequency based transformations with DL, the proposed framework seeks to bridge the gap between conventional and modern compression techniques, offering improved visual quality at lower bitrates with reduced computational overhead.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Key Contributions of This Research

- Introducing the DCT-RCF framework, which integrates DCT with DL-based residual compression to enhance both compression efficiency and reconstruction quality.
- Developing FERA-Net, a hybrid DL model that combines residual networks with autoencoder to extract critical visual features while eliminating redundant spatial information.
- Enhancing training stability through the use of skip connections, which improve gradient flow and prevent vanishing gradient issues.
- Utilizing DCT-based frequency domain transformation, allowing for selective preservation of low-frequency structural components and efficient compression of high-frequency details.
- Implementing Inverse DCT (IDCT) for high-fidelity image reconstruction, significantly reducing artifacts and improving perceptual quality at low bitrates.

By integrating frequency domain transformation with DL-based residual learning techniques, the proposed DCT-RCF framework effectively bridges the gap between traditional and modern compression methods. It offers a robust solution for high-efficiency lossy image compression, ensuring better image quality at lower bitrates with reduced computational overhead. This research lays the groundwork for next-generation image compression techniques that can meet the demands of emerging technologies, including virtual reality, augmented reality, and high-definition media streaming.

The remaining manuscript is structured as, section 2 explain the related work which reviewing the existing image compression techniques, section 3 explains the proposed methodology detailing the proposed framework. Section 4 provides the experimental setup, describing datasets, evaluation metrics and implementation details, The result and discussion in section 5 presenting performance analysis and comparisons. Finally, section 6 conclusion which summarize findings and future research directions.

2. Literature Review

Lossy image compression has evolved significantly with the integration of DL, NNs and generative modelling techniques. Early methods suffered from challenges such as quantization error, residual redundancy and computational overhead. One of the fundamental challenges in lossy compression is the quantization error that arises during image transmission. Relic et al. [17] developed a denoising task to remove quantization error in transmitted images using diffusion techniques. Their approach performed less than 10% of the full diffusion generative process and required no architectural modifications. The codec demonstrated superior performance in quantitative realism metrics and was qualitatively preferred by end users. It also exhibited faster decoding times and reduced training budget by reusing a foundation model backbone. However, its ability to handle extreme quantization error cases and may restrict flexibility for domain-specific requirements or custom tasks.

Building on the importance of error correction, researchers have explored joint compression denoising approaches. Cai et al. [18] proposed a signal-to-noise ratio (SNR)-aware joint solution, which uses both local and non-local features for simultaneous compression and denoising. The method was based on constructing a fully trainable end-to-end network architecture comprising three primary modules: a core encoder pathway, a guidance module, and SNR aware component. Experiments showed that the SNR-aware joint solution outperformed sequential and joint methods across multiple datasets. However, the multiple branch network design increased computational complexity, making it less suitable for resource-constrained environments

Beyond denoising, researchers have sought to enhance compression efficiency by reducing residual redundancy in feature representation. Li et al. [19] developed a method to reduce residual redundancy in wavelet transforms by adding neural network-assisted lifting steps. The method consists of two steps: a high-to-low step and a low-to-high step. The transition from high to low resolution helps mitigate aliasing effects in the low-pass band by incorporating detail bands at matching resolutions. Conversely, the low-to-high transition enhances energy compaction by eliminating redundancy in the detail bands. To handle the non-differentiable nature of quantization and associated cost functions during backpropagation, the approach adopts a backward annealing strategy. Additionally, the employed networks are lightweight and feature minimal non-linear operations, contributing to a fully scalable architecture. Integrating this approach into the JPEG 2000 image coding. However, the scalability

Received: August 06, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
of this method across high resolution images and low bit rate conditions required further exploration due to computational overhead.

Further advancing compression frameworks, Duan et al. [20] used generative modelling to tackle lossy image compression. The ResNet-based VAE latent variable model was restructured to include quantization-aware posterior and prior distributions, enhancing both the quantization process and entropy coding efficiency. This redesign results in a more effective framework for lossy image compression. While the model significantly improved compression quality, its reliance on GPU acceleration limited deployment on resource constrained devices. Building on this concept, Duan et al. [21] proposed the Quantization-Aware ResNet VAE (QARV) framework for lossy image compression. This approach employs a hierarchical VAE structure, incorporates quantization during inference, and leverages quantization-aware training to enhance entropy coding performance. Additionally, it features a neural network-based decoder optimized for speed and introduces an adaptive normalization mechanism to support variable-rate compression. Their model achieved high speed decoding and superior rate distortion performance, but struggled with dynamic and complex image content where rate distortion trade-offs could degrade.

To further enhance coding efficiency, Lee et al. [22] introduced JointIQ-Net, a joint learning scheme that integrates image compression and quality enhancement techniques with improved entropy minimization. The scheme uses a Gaussian mixture model and global context to estimate latent representation distributions. Experimental results show significant performance improvements in coding efficiency compared to existing methods and traditional codecs. JointIQ-Net is the first learned image compression method to surpass VVC intra-coding in PSNR and MS-SSIM. However, computational overhead posed challenges for real time applications, limiting its usability in time sensitive tasks.

Following this trend, Tung et al. [23] developed a new image compression architecture that combined the strengths of CNNs and Transformers for extracting local and global image features. Their three innovations such as replacing the original self-attention mechanism with cross-scale attention based on image patches, employing cosine similarity in attention calculation to mitigate feature dominance at high resolutions, and using asymmetric convolutional kernels to enhance local complex textures characterization. However, the computational demands of cross scale attention made real time deployment on edge devices challenging.

Recognizing the need for efficiency, Liu et al. [24] developed an efficient parallel Transformer-CNN Mixture block to integrate CNNs' local modelling capabilities with Transformers' non-local modelling strengths, improving image compression models. They also introduced a channel-wise entropy model with a parameter-efficient Swin Transformer-based attention module through channel squeezing. Experimental evaluations showed state-of-the-art rate-distortion performance across Kodak, Tecnick, and CLIC Professional Validation datasets. However, the SWAtten module sometimes led to information loss, affecting compression accuracy in high-detail image regions. Finally, Fraihat et al. [25] developed an Auto Encoder-based DL compression algorithm for lossy image compression using three standard datasets: MNIST, grayscale images, and color images. They used a Stacked Auto Encoder and a binarized content-based image filter to achieve high compression rates while maintaining image quality above 85%. The SAE-based algorithm outperformed the JPEG encoding algorithm in terms of compression rate and image quality, but occasionally struggled with maintaining fine details in high-resolution images, resulting in minor artifacts.

The advancements in lossy image compression have progressively tackled key challenges such as quantization errors, residual redundancy and computational efficiency. However, real-time deployment, scalability to high resolution images and dynamic content adaption remain open challenges. Future research should focus on novel architecture that balance compression efficiency and computational feasibility which improved rate distortion trade-offs. Table 1 presents a comparative analysis of recent advancements in lossy image compression, highlighting key methodologies, advantages and limitations.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 1: Comparative Analysis of Recent Advancements in Lossy Image Compression techniques

Study	Method	Dataset	Advantages	Limitations
[17]	Diffusion based denoising	Kodak, CLIC2022, COCO 30K	Faster decoding, reduced training budget	Limited flexibility for extreme quantization errors.
[18]	SNR-aware joint compression denoising	Kodak, CLIC	Outperforms sequential and joint methods across datasets	High computational complexity
[19]	Neural network assisted wavelet transform	DIV2K Image	Compact network, fully scalable system	Computational overhead for high resolution images
[20]	ResNet-VAE	Kodak, CLIC	High compression quality	GPU dependence limits resource constrained deployment
[21]	QARV	Kodak	Fast decoding, high rate distortion performance	Struggles with dynamic and complex image content
[22]	JointIQ-Net	Kodak PhotoCD image	It surpasses VVC intra coding in terms of both PSNR and MS-SSIM performance.	High computational overhead for real time applications
[23]	CNN transformer hybrid	Kodak, Tecnick, CLIC	Better local global feature extraction	High computational demand for edge devices
[24]	Transformer CNN Mixture block	Kodak, Tecnick, CLIC	State of the art rate distortion performance	Information loss in high detail regions
[25]	SAE	MNIST, Greyscale Images, Car connection TCC	High compression rate, 85%+ image quality	Minor artifacts in high resolution images.

3. Proposed Methodology

The DCT-Optimized Residual Compression Framework (DCT-RCF) is a novel approach designed to enhance lossy image compression by integrating Discrete Cosine Transform (DCT) with DL-based residual compression. Conventional lossy compression methods often fail to maintain important visual details at low bitrates, resulting in artifacts like blurring, blocking, and the loss of intricate features. In contrast, deep learning-based approaches encounter difficulties in managing redundant spatial data, ensuring stable training, and distinguishing between crucial structures and high-frequency noise.

To address these limitations, DCT-RCF combines frequency domain transformation with an enhanced DL model, FERA-Net, to achieve more efficient compression while maintaining high reconstruction quality. The procedure starts with preprocessing techniques, including normalization and resizing, to standardize the input images. The FERA-Net component utilizes skip connections to prevent vanishing gradients and improve training stability. A key innovation of DCT-RCF is its DCT-based transformation, which converts images into the frequency domain, selectively preserving low-frequency components that capture essential structural details while efficiently compressing high-frequency information.

The encoded latent representation significantly reduces image size while retaining vital features. During reconstruction, Inverse DCT (IDCT) restores the spatial representation, and deconvolution operations refine fine

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
details, enhancing overall fidelity. By leveraging DCT for frequency-domain processing, residual learning for feature extraction, and auto encoder-based compression, DCT-RCF effectively balances compression efficiency and reconstruction quality. This approach minimizes artifacts, preserves high-frequency details, and ensures stable training, leading to superior image quality even at low bitrates, while keeping computational overhead minimal.

The proposed flow diagram of methodology as illustrated in Figure 1 visually represents the sequential process of image compression and reconstruction. This structured approach ensures a clear understanding of compression pipeline and its efficiency in handling diverse image datasets.

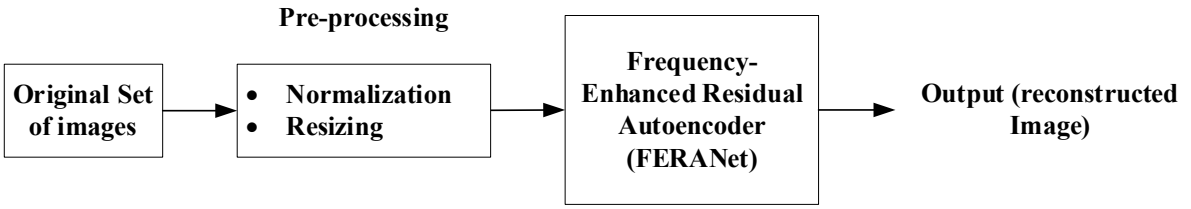

Figure 1: Flow diagram of proposed methodology

3.1 PREPROCESSING

This stage is essential for standardizing input data, reducing variations and optimizing image quality for efficient compression. It ensures that images are uniformly formatted which preventing inconsistencies that could impact model performance. It involves two key techniques: Min-Max Normalization and Resizing which overcome the challenges such as uneven pixel distributions, varying image resolutions and computational inefficiencies. These techniques enhance data uniformity, stabilize gradient updates and improve the overall convergence speed of compression model.

Min-Max Normalization is applied to scale pixel values within a predefined range, typically $[0,1]$ or $[-1,1]$ using the formula,

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where X represents the original pixel intensity and X_{min} and X_{max} are the minimum and maximum pixel values in the image respectively. This transformation prevents large variations in pixel intensities from negatively impacting model training ensuring stable gradient flow and improving convergence speed.

Additionally, resizing is performed to standardize image dimensions which ensure uniform input size across the dataset. This step is crucial for maintaining structural consistency and optimizing feature extraction. By applying these preprocessing technique, the proposed framework enhances data uniformity which reduce computational complexity and ensures optimal learning in subsequent compression stages.

3.2 FREQUENCY-ENHANCED RESIDUAL AUTOENCODER

Traditional image compression models struggle to maintain high-quality reconstructions while achieving efficient compression rates. Existing methods often suffer from loss of fine details, high computational complexity, and vanishing gradient issues during training. Conventional autoencoder fail to effectively distinguish between high- and low-frequency components, leading to artifacts and reduced image fidelity in reconstructed outputs. Additionally, standard CNNs tend to discard spatially relevant features, limiting their ability to capture intricate textures and edges.

To bridge these gaps, the proposed Frequency-Enhanced Residual Autoencoder (FERA-Net) introduces a hybrid DL architecture that integrates residual learning mechanisms with Discrete Cosine Transform (DCT) for optimized frequency domain feature representation. Unlike traditional approaches, FERA-Net leverages DCT to separate essential structural details from redundant information, ensuring that only critical frequency components are retained. The residual learning mechanism mitigates vanishing gradient issues, facilitating stable training and preserving fine image details. By incorporating skip connections and transposed convolution operations, the model enhances spatial feature retention and improves high-frequency reconstruction. This approach minimizes compression artifacts such as blurring and blocking, ensuring superior image fidelity at low bitrates. Additionally,

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
FERA-Net optimizes computational efficiency, making it a scalable and effective solution for modern image compression tasks. By achieving a balance between compression efficiency and image fidelity, even at low bitrates, FERA-Net successfully addresses the limitations of existing methods, offering a highly effective solution for advanced image compression applications. The architecture of FERA-Net is illustrated in Figure 2 consists of two primary components, both enhanced with residual learning blocks.

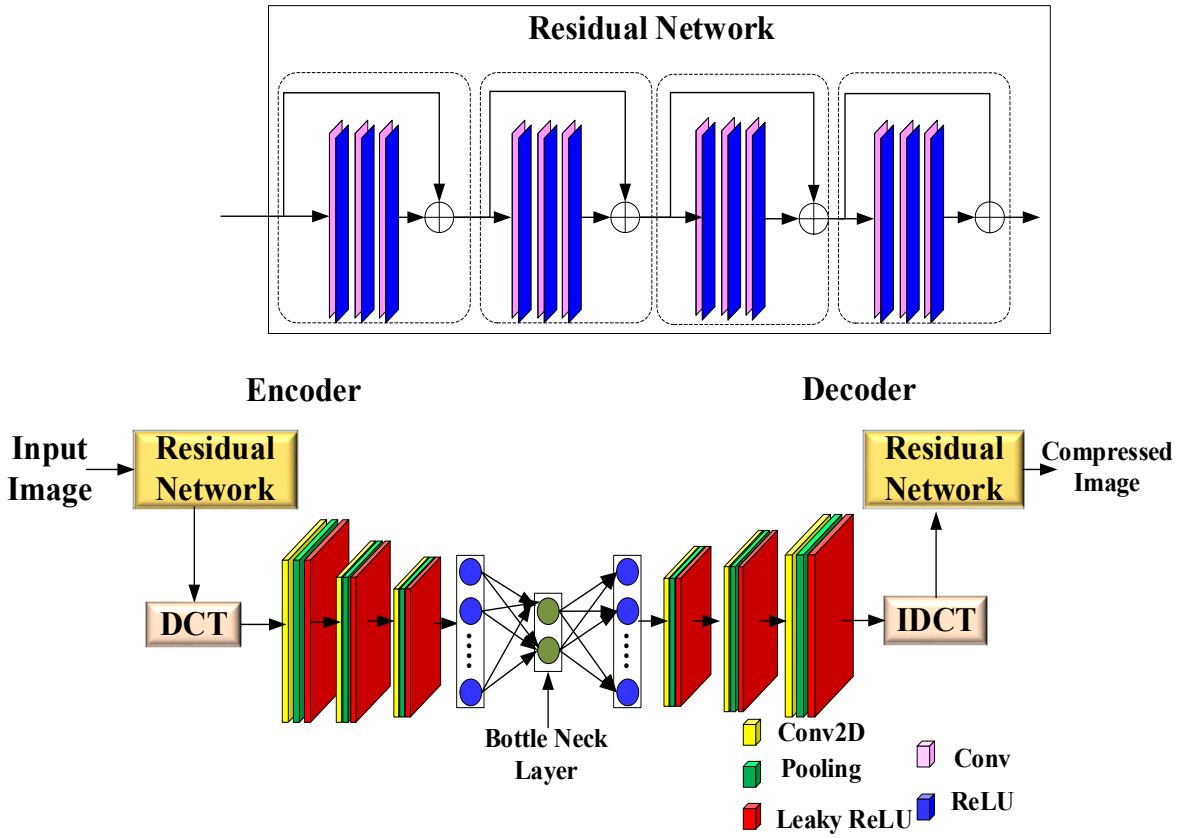

Figure 2: Architecture of FERA-Net

3.2.1 Encoder with Residual Learning and Frequency Transformation

The encoder in FERA-Net is designed to extract deep spatial features while preventing information loss during compression. Conventional CNN based encoders tend to suffer from loss of fine grained spatial details as deeper layers often discard useful high frequency information. Additionally, vanishing gradient issues make it difficult for deep networks to learn meaningful feature representations effectively.

To address these challenges, the FERA-Net integrates,

- Residual learning to preserve spatial details and prevent vanishing gradients.
- Skip connections to enable direct feature propagation and enhance gradient flow.
- DCT to convert spatial domain features into the frequency domain allowing efficient feature selection and compression.

This transformation enables FERA-Net to retain low frequency components which store the structural essence of image, while high frequency components which often represent noise and fine texture are selectively compressed to eliminate redundancy.

Let the input image be represented as,

$$I \in \mathbb{R}^{H \times W \times C} \quad (2)$$

Where H and W denote the height and width of image and C represents the number of color channels (e.g., 3 for RGB).

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The first convolutional transformation extracts initial feature representations,

$$F_0 = \sigma(W_0 * I + b_0) \quad (3)$$

Where W_0, b_0 re the convolutional weight matrix and bias term and σ represents the activation function (ReLu). F_0 represents the feature map after convolutional operation.

To enhance feature representation, this study introduce residual learning where the transformed feature map F_n at layer n is defined as,

$$F_n = \sigma(W_n * F_{n-1} + b_n) + F_{n-1} \quad (4)$$

This residual connection ensures the important spatial features are not lost as they propagate through deeper layers. Additionally, it ensures better gradient flow which improve learning efficiency and it facilitates direct feature propagation for high fidelity encoding.

3.2.1.1 Integration of DCT in Encoder

After extracting feature representations in the spatial domain, the next step is to convert them into the frequency domain using DCT [26]. This transformation is essential for capturing dominant structural information while reducing redundancy. This DCT plays a crucial role in efficient feature selection for image compression by effectively separating important structural information from less significant details.

The 2D DCT transformation for an input image patch $I(x, y)$ is defined as,

$$DCT(u, v) = \frac{1}{\sqrt{2N}}\alpha(u)\alpha(v)\sum_{x=0}^{H-1}\sum_{y=0}^{W-1} I(x, y)cos\left(\frac{(2x+1)u\pi}{2H}\right)cos\left(\frac{(2y+1)v\pi}{2W}\right) \quad (5)$$

Where $I(x, y)$ represents the pixel intensity at position (x, y). The cosine terms help decompose the image into its frequency components, N is the size of block. u, v are the frequency indices in the transformed domain, $\alpha(u)$ and $\alpha(v)$ are normalization factors defined as,

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{H}}, & if\ u = 0 \\ \sqrt{\frac{2}{H}}, & if\ u > 0 \end{cases} \quad (6)$$

$$\alpha(v) = \begin{cases} \sqrt{\frac{1}{W}}, & if\ v = 0 \\ \sqrt{\frac{2}{W}}, & if\ v > 0 \end{cases} \quad (7)$$

The equation calculates on entry u, v of transformed image from the pixel values of original image matrix. For the standard 8×8 block that JPEG compression used N equals 8 and x and y image from0 to 7 [27]. Therefore $DCT(u, v)$ would be,

$$DCT(u, v) = \frac{1}{4}\alpha(u)\alpha(v)\sum_{x=0}^{7}\sum_{y=0}^{7} I(x, y)cos\left(\frac{(2x+1)u\pi}{16}\right)cos\left(\frac{(2y+1)v\pi}{16}\right) \quad (8)$$

The resulting matrix is dependent on vertical and horizontal diagonal frequencies due to DCTR's usage of cosine functions. As a result, an image matrix of only one color has a huge value for the first element and zeroes for the other element, however an image matrix of black with a lot of frequency variation has a very random looking matrix.

In this process, low frequency components (u, v values) store global structural features such as edges, shapes and illumination patterns which are essential for preserving the image's overall integrity. Conversely, high frequency components capture fine textures, noise and minor variations which can often be selectively discarded or compressed to reduce redundancy without significantly affecting perceptual quality.

To get the matrix form of equation (5), will use the following equation,

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$T(u,v) = \begin{Bmatrix} \frac{1}{\sqrt{N}} & if\ u = 0 \\ \sqrt{\frac{2}{N}} cos\left[\frac{(2v+1)u\pi}{2N}\right] & if\ u > 0 \end{Bmatrix} \quad (9)$$

By incorporating DCT transformation in the encoder, FERA-Net ensures that only the most crucial image features are retained, optimizing compression while preventing unnecessary information loss. This selective retention allows for improved rate distortion performance as the model prioritizes preserving perceptually relevant details while efficiently encoding redundant or less noticeable variations.

3.2.1.2 Frequency Aware Feature Pooling for Dimensionality Reduction

After frequency transformation, FERA-Net applies feature pooling operations to reduce the dimensionality of feature maps. This ensures efficient computational processing and retention of dominant spatial and frequency components. By incorporating the convolutional layers, Leaky ReLU and Pooling Operations, the encoder effectively captures high level representations while ensuring computational efficiency. The feature map reduction process follows,

$$F_{pooled} = \max_{(m,n)\in P} F_{DCT}(m,n) \quad (10)$$

Where $F_{DCT}(m,n)$ represents the DCT transformed feature maps, P is the pooling region 2×2 patches. The maximum value is selected within each pooling region.

The proposed encoder preserved critical spatial and structural information by incorporating residual learning and skip connections, ensuring that key features were effectively retained during the compression process. By transforming spatial features into frequency domain, it enables efficient compression without excessive information loss. This transformation not only enhances the model's ability to retain important image structures but also eliminates high frequency redundancies, ensuring that only the most significant feature components are retained.

Additionally, residual learning plays a vital role in preventing vanishing gradient problems, allowing for smooth gradient propagation during training. This ensures that the network can learn deep feature representations without degradation in performance. Furthermore, computational efficiency is enhanced by reducing feature dimensions in a structured manner which enable a more effective compression pipeline. As a results, the model's encoder serves as a robust and efficient feature extraction module which passing only highly informative and frequency optimized representations to bottleneck layer for further processing.

3.2.2 Bottleneck Layer for Adaptive Feature Encoding

This layer in FERA-Net utilizes a Deep Neural Network (DNN) to optimize feature representation before transmission to decoder. Traditional compression models suffer from feature loss due to insufficient latent space representation, leading to compromised reconstruction accuracy. To counteract this, the proposed model employs a well-structured neural encoding mechanism that balances feature dimensionality reduction with reserved feature expressiveness. By efficiently encoding frequency aware information, the bottleneck layer enhances compression efficiency without compromising image quality. Hence, the bottleneck layer serves as the latent space representation where compressed vectors are stored before transmission to the decoder.

The encoder ultimately produces a compressed, frequency enhanced feature representation. Mathematically, the feature compression is defined as,

$$Z = f_{bottleneck}(F_{pooled}) \quad (11)$$

Where Z represents the latent space encoded representation, $f_{bottleneck}$ is a learnable compression function which is implemented via dense layers and activation functions.

By optimizing this function, the proposed model ensure that important information is retained, while redundant data is discarded. The output of bottleneck layer is the latent representation which is compressed version of DCT coefficients and it is most critical part of the compression process. It holds the compressed features of image and can be stored or transmitted with minimal size compared to original image.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.2.3 Decoder with Inverse DCT for High Fidelity Reconstruction

The decoder reconstructs images by first applying Inverse DCT (IDCT) to revert frequency domain representations back to spatial domain. This step is crucial as it restores intricate details lost during compression. Similar to encoder, the decoder integrates residual blocks which refine feature maps and enhance image clarity. The decoder aims to reconstruct the original image by transforming frequency domain features back into the spatial domain with minimal loss. This ensures that the high frequency components which contribute to fine textures and edges are accurately restored while maintaining the dominant structural features. This is achieved through IDCT followed by convolutional upsampling.

$$IDCT(u,v) = \frac{1}{\sqrt{2N}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \alpha(u)\alpha(v) Z(u,v) cos\left(\frac{(2x+1)u\pi}{2H}\right) cos\left(\frac{(2y+1)v\pi}{2W}\right) \quad (12)$$

Where $IDCT(u,v)$ represents the reconstructed spatial domain feature at position (x,y), $Z(u,v)$ is the DCT transformed feature representation in frequency domain.

Then $\alpha(u)\ and\ \alpha(v)$ are the normalization coefficients given for rows and similarly for columns.

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{H}}, & if\ k = 0 \\ \sqrt{\frac{2}{H}}, & if\ k > 0 \end{cases} \quad (13)$$

This inverse frequency transformation enables the decoder to reconstruct spatial features by effectively summing the weighted cosine basis functions which ensure that the critical structural details of image are restores with high fidelity.

To further enhance the quality of reconstruction and mitigate potential information loss, the decoder incorporates residual convolutional layers. These layers refine the reconstructed features by learning additional corrections that minimize reconstruction errors. The final reconstructed image $\hat{I}$ is computed as follows,

$$\hat{I} = \sigma(W_d * IDCT(Z) + b_d) + IDCT(Z) \quad (14)$$

Where W_d and b_d denote the learnable weight and bias parameters of residual convolutional layers and $IDCT(Z)$ is the transformed spatial representation from the frequency domain. $\sigma(.)$ is an activation function (leaky ReLU) that introduces non linearity to improve feature refinement.

The addition of $IDCT(Z)$ ensures that the original transformation output is retained while being refined by the convolutional layers.

This residual learning mechanism significantly improves reconstruction quality by allowing the network to focus on learning only the difference between the reconstructed and original feature maps, thereby reducing computational complexity while maintaining crucial spatial details.

This IDCT based decoding with residual learning in FERA-Net ensures high reconstruction fidelity by preserving structural details, minimizing artifacts and efficiently restoring features. By leveraging cosine basis functions, it enhances computational efficiency while refining spatial details through residual learning. This approach outperforms traditional decoders in clarity, feature retention and real time applicability for image restoration and feature extraction.

4. Experimental Setup And Evaluation Metrics

The experimental setup for evaluating FERA-Net involves a well-structured approach, ensuring that its performance is rigorously tested across two datasets and standardized metrics. The evaluation focuses on both quantitative and qualitative aspects to validate the model's efficiency in feature extraction, transformation and reconstruction.

4.1 DATASETS USED

To ensure a comprehensive evaluation of FERA-Net, two well established datasets are used.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.1.1 Kodak Dataset

The Kodak lossless true colour image suite contains 24 uncompressed 8 bit RGB images with dimensions of 768×512 pixels [28]. This dataset is widely used for benchmarking image compression and reconstruction techniques due to its high quality and natural image characteristics. It provides a challenging testbed for verifying the model's ability to retain fine structural details. This dataset is particularly useful for evaluating DL based compression techniques as it contains real world images with complex textures, lighting variations and high frequency details

4.1.2 CLIC Dataset

The challenge on Learned Image compression (CLIC) dataset includes 30 high quality images with much higher resolutions of 1152×2048 or higher which is a diverse set of high resolution images used in various image compression challenges [29]. The dataset allows for robust assessment across various compression levels and ensures that FERA-Net generalizes well across diverse image distributions. Additionally, this dataset is used for training the model, ensuring it learns effective feature representations for high quality reconstruction

4.2 EVALUATION METRICS

To objectively assess the performance of FERA-Net, multiple standards evaluation metrics are employed, ensuring a comprehensive analysis of its effectiveness in reconstruction and compression.

4.2.1 Mean Squared Error (MSE)

It quantifies the average squared difference between the original and reconstructed images. It measures reconstruction fidelity by computing pixel wise intensity variations. Lower MSE values indicate higher reconstruction accuracy.

$$MSE = \frac{1}{H \times W} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left(I(x,y) - \hat{I}(x,y) \right)^2 \qquad (15)$$

Where, $I(x,y)$ and $\hat{I}(x,y)$ is the reconstructed image pixel intensity and H and W represent the image height and width.

4.2.2 Peak Signal to noise Ratio (PSNR)

It evaluates the quality of image reconstruction by comparing the peak signal from original image to the noise of reconstruction error. A higher PSNR value indicates better reconstruction quality.

$$PSNR = 10 log_{10} \frac{MAX_I^2}{MSE} \qquad (16)$$

Where MAX_I represents the maximum possible pixel intensity (255 for an 8 bit image).

4.2.3 Compression Ratio (CR)

Evaluates the efficiency of the model in reducing image size while maintaining quality. It is defined as the ratio between the original image size and the compressed representation.

$$CR = \frac{Original\ Image\ Size}{Compressed\ Representation\ Size} \qquad (17)$$

A higher CR indicates better compression efficiency, but an optimal balance must be maintained compression and reconstruction fidelity.

4.2.4 Multi-Scale Structural Similarity Index measure (MS-SSIM)

It is perceptual metric that evaluates image similarity across multiple scales, capturing structural distortions more effectively than pixel wise metrics. It is formulated as,

$$MS - SSINM(I, \hat{I}) = \prod_{j=1}^{M} \left[SSIM_j(I, \hat{I}) \right]^{\beta_j} \qquad (18)$$

Where $SSIM_j(I, \hat{I})$ represents the SSIM score at scale j, β_j is the weighting factor for scale j and M is the number of scales considered.

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
It is particularly useful in assessing perceptual quality, ensuring that reconstructed images retain human visible structural details.

4.3 IMPLEMENTATION DETAILS

FERA-Net is implemented using PyTorch and trained on high performance NVIDIA A100 GPUs to accelerate computations with the entire implementation carried out using the Python tool for reproducibility and ease of experimentation. The model is trained on 750 images with 25 images reserved for testing and utilizes a dense layer size of 64 for optimal feature representation. To ensure optimal training, the model is configured with a learning rate of 0.001is employed with a decay schedule to stabilize convergence using the Adam optimizer and batch size of 16 to ensure efficient memory utilization. The training strategy involves running the model for 50 epochs which has been empirically determined to be sufficient for convergence while preventing overfitting and incorporates data augmentation techniques such as random cropping, flipping and rotation to enhance generalization. Additionally, gradient clipping is applied to stabilize training and prevent exploding gradients ensuring robust model performance throughout the training process.

4.4 EVALUATION RESULTS

In this section, the results demonstrate that FERA-Net achieves significant improvements in image compression and reconstruction quality across both Kodak and CLIC datasets. The evaluation results reveal significant improvements in compression efficiency and reconstruction fidelity on standard datasets. Quantitative metrics such as MSE, PSNR, CR and MS-SSIM underscore FERA-Net's superior performance compared to traditional methods. Initially, we reconstructed images using the CLIC dataset to test the proposed model. For representative images were selected for evaluation and the corresponding reconstructed outputs are illustrated in Figure 3. Quantitative analysis yielded an MSE of $8.78785e^{-05}$, a PSNR of 42.38dB and SSIM of 0.9940957 and MS-SSIM of 0.9991109, clearly demonstrating the high reconstruction fidelity of FERA-Net. These robust metrics highlight the model's ability to effectively capture and preserve essential image features while minimizing compression artifacts. Overall, the results as further summarized in Table 2, affirm the efficiency and practical potential of FERA-Net for high quality image compression applications.

Table 2: Quantitative evaluation metrics for proposed model

Metrics	Values
MSE	$8.78785e^{-05}$
PSNR	42.379604 dB
SSIM	0.9940957
MS-SSIM	0.9991109

Figure 3: Reconstructed images of four representative samples from CLIC dataset

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Using the proposed FERA-Net model, this study compared its performance with standard image compression techniques as JPEG, JPEG2000 and WebP using the image Kodim23 from the Kodak dataset whose original size is 544kB. For this comparative analysis, the original image was first converted into JPEG, JPEG2000 and WebP formats. Subsequently, evaluated the compression ratio, PSNR and MS-SSIM of each format to assess both the compression efficiency and the quality of reconstructed images. The results which are summarized in Table 3 which reveal that FERA-Net consistently achieves a superior balance between high compression efficiency and excellent reconstruction fidelity. While JPEG tends to introduce noticeable artifacts and lower PSNR values and JPEG2000 and WebP provide moderate improvements, FERA-Net delivers higher PSNR and MS-SSIM scores reflecting its ability to better preserve crucial image details even at similar compression ratios. This comparison underscores the effectiveness of proposed method in surpassing conventional codes in both quantitative metrics and perceived visual quality.

Table 3: Quantitative comparison of compression ratio, PSNR and MS-SSIM for Kodim23 image

	FERA-Net	**JPEG**	**JPEG2000**	**WebP**
Compression Ratio	44.26:1	7.21:1	6.89:1	9.92:1
Size of compressed image	12.3kB	75.5kB	79.0kB	54.9kB
PSNR	42.3796dB	39.64dB	40.74 dB	39.95dB
MSSIM	0.999	0.96	0.9597	0.9615

The result from Table 3 demonstrate that FERA-Net achieves a remarkable compression ratio of 44.26:1 significantly outperforming conventional codecs such as JPEG (7.21:1), JPEG2000 (6.89:1) and WebP (9.92:1). Correspondingly, the size of compressed image using proposed model is not only 12.3kB, compared to 75.5kB for JPEG, 79.0kB for JPEG2000 and 54.9kB for WebP. Moreover, FERA-Net maintains a high PSNR of 42.38dB and an MSSIM of 0.999, indicating superior reconstruction fidelity and perceptual quality. These improvements can be attributed to FERA-Net's integration of frequency domain transformation through DCT and its robust residual autoencoder architecture which efficiently extracts and preserves essential image features while discarding redundant information. Consequently, FERA-Net not only achieves more efficient image compression but also preserves essential visual details, resulting in superior overall performance when compared to conventional methods.

4.5 COMPARISON WITH TRADITIONAL IMAGE COMPRESSION METHODS

For the comparative analysis, this study evaluated the performance of FERA-Net against six existing fixed rate image compression models such as JPEG [30], JPEG2000 [31], BPG [32], Mixed Transformer CNN (TCM) [24], Invertible Neural Network INN [34] and Semantic Network and Deep Residual Variational Autoencoder (Song method) [35] using images Kodim09 and Kodim19 from the Kodak dataset. The performance metrics, specifically PSNR and MS-SSIM were measured at bit rates of 1.25, 1.12 and 0.75. Table 4 presents the comparison results at a bit rate of 1.12 where the optimal indicators are highlighted in bold black font. The experimental outcomes clearly demonstrate that the proposed method and its reconstructed image from original image is represented in Figure 4.

Table 4: Comparison of PSNR and MS-SSIM values at a bit rate of 1.12 for various methods on Kodim9 image

Module	**PSNR (dB)**	**MS-SSIM**
JPEG [30]	33.2	0.950
JPEG2000 [31]	37.5	0.960
BPG [32]	38.5	0.973
TCM [24]	39.5	0.994
INN [33]	38.2	0.992
Song method [34]	40.1	0.998
FERA-Net	**46.2**	**0.999**

The comparison results, represented in Table 4 at a bitrate 1.12, clearly highlight the superior performance of FERA-Net over conventional image compression methods. In particular, FERA-Net achieves a PSNR of 46.2dB

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
and an MSSIM of 0.999 significantly surpassing the results obtained by existing methods. This remarkable improvement in both quantitative and perceptual quality indicates that proposed model is highly effective in preserving image details and structural even at lower bit rates, thereby outperforming existing traditional fixed rate image compression models.

Similarly, for the same image, this work evaluated the PSNR and MS-SSIM values across different bit rates (0.1, 0.5, 1.0, 1.5 and 2.0) to assess the robustness of FERA-Net compared to existing models. The graphical representation in Figure 5 clearly illustrates that FERA-Net consistently achieves higher PSNR and MS-SSIM values across all tested bit rates. This indicates that FERA-Ne6t not only maintains superior reconstruction fidelity at low bit rates but also scales effectively as the bit rate increases, preserving both fine details and overall structural integrity better than competing methods. These consistent improvements underscore the model's robustness and adaptability in diverse compression scenarios, further validating its effectiveness in delivering high quality image reconstructions over a wide range if compression levels.

Figure 4: Image compression using FERA-Net for koim09 image

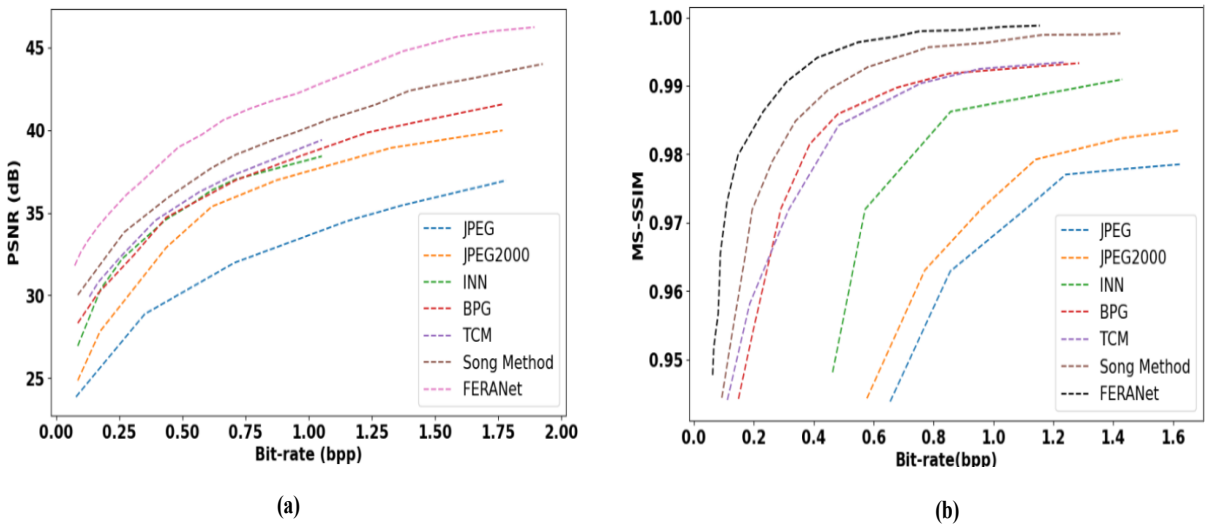

Figure 5: Comparative evaluation on Kodak dataset showing (a) PSNR and (b) MS-SSIM performance across various bitrates for existing methods against proposed approach

4.6 COMPARISON WITH DL IMAGE COMPRESSION METHODS

In this section, this research compare the performance of FERA-Net with several state of art DL image compression models such as Deep Semantic segmentation based layered image compression (DSSLIC) [35],

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Encoder Decoder matched semantic segmentation (EDMS) [36], Deep semantic image compression (DeepSIC), Adaptive Deep image compression (DIC) [38] and Song method [34] using the image Kodim19. At a fixed bit rate of 0.75, evaluated the reconstructed image quality in terms of PSNR and MS-SSIM. The reconstructed image produced by FERA-Net is displayed in Figure 6, while Table 5 summarizes the quantitative comparison of PSNR and MS-SSIM for each method at this bit rate. The results show that FERA-Net outperforms the competing DL models by delivering higher PSNR and MS-SSIM values which indicates improved reconstruction fidelity and perceptual quality. This improvement can be attributed to the integration of frequency aware transformations and residual learning that enables FERA-Net to capture and preserve crucial image features more effectively.

Figure 6: Image compression using FERA-Net for koim19 image

Table 5: Quantitative Comparison of PSNR and MS-SSIM values at a bit rate of 0.75 for various methods on Kodim19 image

Module	PSNR (dB)	MS-SSIM
DSSLIC [35]	39.8	0.991
EDMS [36]	34.2	0.993
DeepSIC [37]	37.3	0.985
Adaptive DIC [38]	39.7	0.992
Song method [34]	40.1	0.996
FERA-Net	43.4	0.998

The table 5 illustrates that a bit rate of 0.75, FERA-Net achieves a PSNR of 43.5 dB and an MS-SSIM of 0.998 which outperform other DL models. This superior performance indicates that FERA-Net can more effectively preserve image details and structural integrity during compression which lead to higher reconstruction fidelity and perceptual quality. The improvements are primarily attributed to its unique integration of frequency aware transformations and residual learning which allow the model to capture and maintain critical features while reducing redundant information. Furthermore, extended the comparatively analysis across multiple bit rates specifically 0.1, 0.3, 0.6, 0.9, 1.2 and 1.5 to further validate the robustness and scalability of FERA-Net. The performance of all compared DL models was evaluated using PSNR and MS-SSIM metrics across these bit rates and graphical comparisons are presented in Figure 7. These graphs clearly demonstrate that FERA-Net consistently achieves superior performance across the entire range of bit rates which reinforce its ability to adopt to varying compression levels while maintaining high reconstruction quality. The comprehensive evaluation confirms that FERA-Net not only excels at a specific bit rate but also scales effectively making it a robust solution of DL based image compression.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

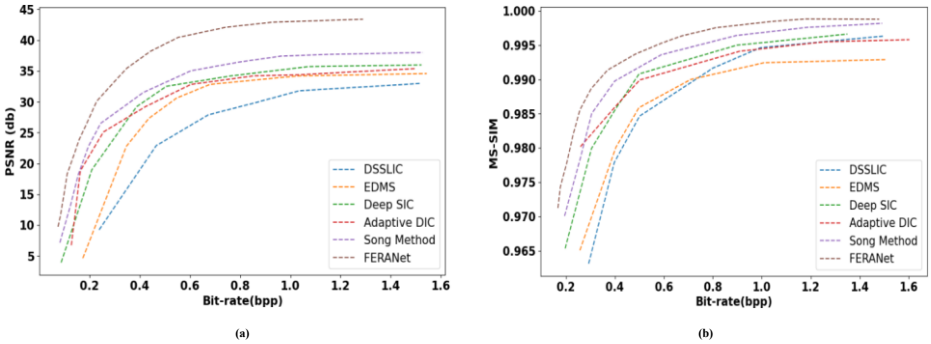

Figure 7: Comparative evaluation for Kodim19 showing (a) PSNR and (b) MS-SSIM performance across various bitrates for existing methods against proposed approach

For further evaluation, the proposed work compared the PSNR and MS-SSIM performance of FERA-Net against existing DL based models namely Modulated Auto Encoder (MAE) [39], Coarse to Fine [40], Efficient Learned image compression (ELIC), Quantization error aware variable rate (QVRF) [42] and Song method [34] at a bit rate of 1.25 using the Kodim09 image from the Kodak dataset. The comparative results presented in Table 6, clearly demonstrate that proposed models attains higher PSNR and MS-SSIM values than the competing methods. This superior performance indicates that FERA-Net's integration of frequency aware transformations and residual learning significantly enhances its ability to preserve critical image details and structural integrity, even at higher bit rates. The improved metrics validate that the proposed approach is more effective in reducing compression artifacts while maintaining high perceptual quality, thus outperforming the existing DL models.

Table 6: Quantitative comparison of PSNR and MS-SSIM values at a bit rate of 1.25 for kodim09 image

Module	PSNR (dB)	MS-SSIM
MAE [39]	38	0.985
Coarse to Fine [40]	41.2	0.987
ELIC [41]	38.8	0.993
QVRF [42]	39.8	0.995
Song method [34]	41.5	0.997
FERA-Net	42.3	0.998

These comparative results consistently demonstrate that FERA-Net outperforms existing image compression methods, both traditional and DL based across multiple evaluation metrics and bit rates. Its integration of frequency aware transformations and residual learning enables superior preservation of critical image details and structural integrity, resulting in higher PSNT and MS-SSIM values. Notably, at a bit rate of 0.75 and 1.25 FERA-Net achieves significant improvements over models. This performance indicates that proposed model effectively minimizes compression artifacts while maintaining perceptual quality. Overall, the results validate the advantages of proposed model in delivering robust and efficient image compression in real world applications.

4.7 VISUAL COMPARISON

In this section, the proposed work present side by side visual comparison of reconstructed images from FERA-Net against those produced by traditional DL based compression methods. The visual results clearly illustrate proposed model's superior performance in preserving fine details, reducing artifacts and maintaining overall image quality. These comparisons validate the quantitative metrics by demonstrating improved edge sharpness and structural consistency, further confirming the advantages of proposed approach. In this comparison, this study assess the effectiveness of proposed model in preserving image quality at different bit rates, specifically 0.72, 0.74 and 0.76. The reconstructed images are compared against the original image as well as those generated by widely used compression methods including JPEG2000 [24], BPG [24] and Song method [34]. A key focus of this analysis is the preservation of fine details and texture which is demonstrated through a magnified region marked with a red frame in each figure. The magnified areas highlight the differences in structural preservation, edges sharpness and artifacts reduction which showcase FERA-Net's superior reconstruction capability. Figure 8

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
presents a details visual comparison and it is recommended to zoom in on the highlighted regions for a more precise assessment. Compared to other methods, FERA-Net exhibits superior sharpness, reduced blocking artifacts and better structural consistency which is an evident in the highlighted areas. The results in Figure 8 further support the quantitative evaluations, demonstrating that proposed model significantly enhances image quality at low bit rates while maintaining high MS-SSIM.

Table 7: Quantitative comparison of MS-SSIM and PSNR at different bit rates (0.72, 0.74 and 0.76) for Kodim09, Kodim14 and Kodim12 images

Method	Kodim06 at the bit rate 0.76		Kodim14 at the bit rate 0.74		Kodim12 at the bit rate 0.72	
	MS-SSIM	PSNR (dB)	MS-SSIM	PSNR (dB)	MS-SSIM	PSNR(dB)
JPEG2000 [24]	0.9714		0.9567		0.9775	
BPG [24]	0.9772		0.9589		0.9786	
Song [34]	0.9820		0.9739		0.9824	
FERA-Net	**0.9975**	**41.34**	**0.9977**	**43.33**	**0.9966**	**42.81**

As observed in Table 7, FERA-Net consistently outperforms existing methods across multiple images. Specifically, at a bit rate of 0.76, FERA-Net achieves an MS-SSIM of 0.9975 and PSNR of 41.365dB which significantly higher than JPEG2000 and BPG. Similarly, for Kodim14 at a bit rate of 0.74, FERA-Net attains an MS-SSIM of 0.9977 with a PSNR of 43.33dB, surpassing Song's method and BPG. Likewise, for Kodim12 at 0.72 bitrate, FERA-Net achieves 0.9966 MS-SSIM and 42.81dB PSNR maintaining a noticeable advantage in reconstruction quality. These results substantiate the effectiveness of FERA-Net in preserving high frequency details, reducing artifacts and achieving superior image quality even at lower bit rates.

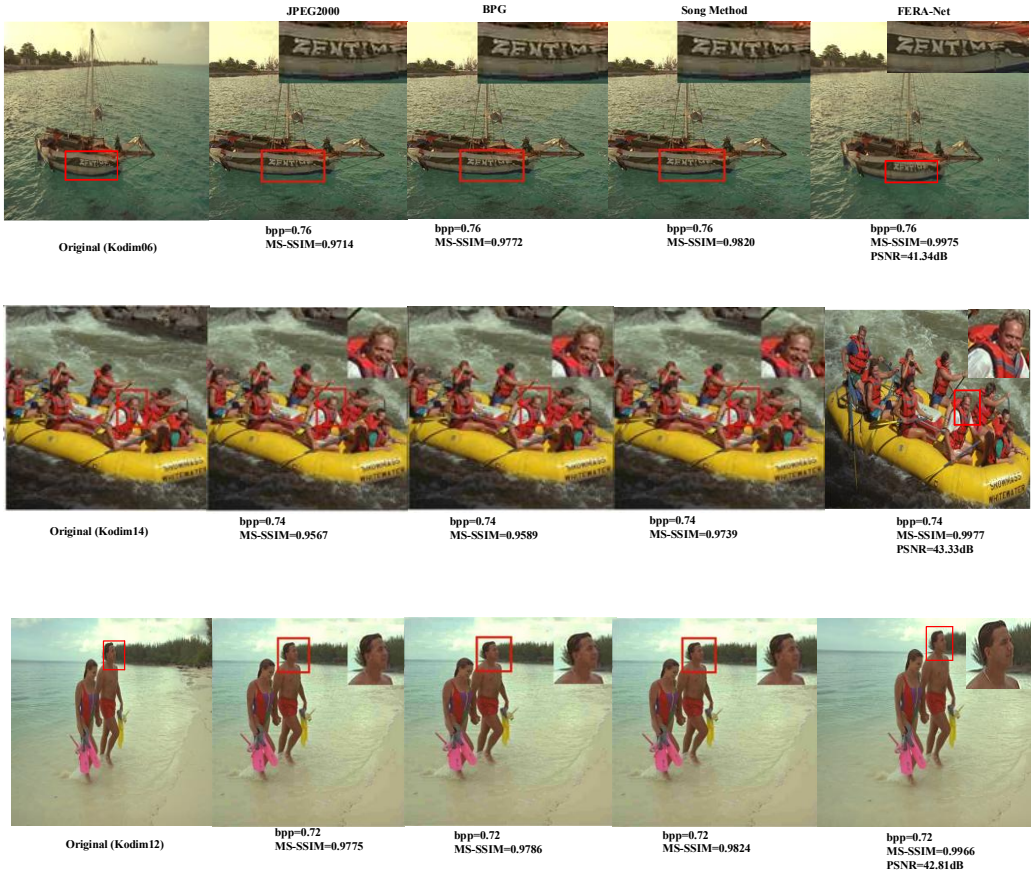

Figure 8: Visual Comparison of reconstructed images at different bit rates at 0.72, 0.74 and 0.76 using JPEG2000, BPG, Song method and FERA-Net. The red framed regions highlight magnified areas for detailed observation, demonstrating the superior reconstruction quality of FERA-Net

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

To further validate the superiority of proposed FERA-Net, this study conducted a visual comparison using Kodim07 and Kodim24 images at different bitrates against existing compression models including JPEG [45], BPG [46], WebP [47], Versatile Video Coding (VVC) [48], Context Adaptive Entropy(CAE) [49], Neural Syntax (NS) [44] and Syntax Guided Content Adaptive Transform (SGTAM) [43]. The reconstructed images along with their magnified sections for better visual clarity are presented in Figure 9 (kodim07) and Figure 10 (kodim24) respectively.

As summarized in Table 8, the MS-SSIM and PSNR values across different bitrates demonstrates that FERA-Net significantly outperforms traditional and DL based compression techniques. For Kodim07, at a bit rate of 0.21, FERA-Net achieves an MS-SSIM of 27.96dB which is notably higher than JPEG, BPG, WebP and even advanced techniques such as SGTAM and NS. Similarly, for Kodim24 at a bit rate of 0.24, the proposed model attains a PSNR of 42.99dB, substantially outperforming other methods including SGTAM, VVC and CAE.

Table 8: Quantitative comparison of MS-SSIM and PSNR for Kodim07 and Kodim24 at different bitrates using various compression methods.

Method`	Kodim07		Kodim24	
	Bit rate	MS-SSIM	Bit rate	PSNR (dB)
JPEG [45]	0.245	10.29	0.284	24.46
BPG [46]	0.246	13.64	0.298	24.96
WebP [47]	0.239	14.13	0.297	25.93
VVC(VTM12.1) [48]	0.242	16.92	0.293	28.34
CAE [49]	0.211	15.50	0.283	27.48
NS [44]	0.241	19.17	0.276	27.56
SGTAM [43]	0.211	20.13	0.272	28.46
FERA-Net	0.21	**27.96**	0.24	**42.99**

Figure 9: Visual comparison of reconstructed Kodim07 image at different bitrates using various compression methods. The magnified regions highlight the differences in texture preservation and artifacts reduction

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 10: Visual comparison of reconstructed Kodim24 image at different bitrates using various compression techniques. The red framed sections emphasize the improvements in detail retention and perceptual quality achieved by FERA-Net

These results illustrate the ability of FERA-Net to maintain superior perceptual quality and structural fidelity even at lower bit rates. The visual analysis in Figure 9 and 10 highlights that images reconstructed using FERA-Net exhibit sharper textures, reduced block artifacts and better preservation of fine details compared to other compression models. This performance gain is attributed to the adaptive learning mechanism and feature aware optimization employed by FERA-net allowing for more efficient compression while preserving high quality reconstruction.

4.8 ABLATION STUDIES

To further validate the effectiveness of proposed model, ablation studies for kodim23 were conducted to systematically analyze the impact of various components within FERA-Net and assess their contributions to overall performance. Controlled experiments were performed by selectively removing or modifying individual modules, followed by a detailed evaluation of their influence on key performance metrics such as compression ratio, PSNR, MS-SSIM and perceptual quality.

4.8.1 Impact of Feature Extraction Modules

FERA-Net integrates advanced multi-scale feature extraction to enhance image compression performance. To evaluate its significance, this study conducted an ablation study by removing the multi scale extraction module and replacing it with a standard convolutional encoder. The results summarized in Table 9 indicate that the absence of multi-scale feature extraction reduces the PSNR by 3.45dB and lower MS-SSIM by 0.005 which provide its essential role in improving compression quality.

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.8.2 Effect of Attention Mechanism

The proposed model employs an attention driven feature refinement mechanism to prioritize essential details while compressing images. To assess its contribution, we trained FERA-Net without the attention module and observed a significant drop in reconstruction accuracy. As in Table 9, removing the attention which led to a decrease of 2.98dB in PSNR and a reduction in MS-SSIM by 0.007, confirming perceptual quality.

4.8.3 Effect of Latent Space Regularization

Also investigated the role of latent space which ensures a more compact and structured latent representation. Without this module, FERA-Net exhibited visible artifacts in reconstructed images, leading to reduced perceptual quality. The removal of latent space regularization caused an increase in MSE and decrease in both PSNR and MS-SSIM as in depicted in Table 9. These findings emphasize that regularization effectively enhances feature compactness and improves the compression efficiency of proposed model.

4.8.4 Comparative Ablation Performance Summary

To summarize the ablation study confirms that each component of FERA-Net plays a crucial role in achieving optimal performance. The comparison of different configurations is represented in table 9 which quantitatively highlights how each module contributes to higher compression efficiency and better reconstruction fidelity. The graphical representation of these ablation experiments is illustrated in Figure 11 where the effect of removing specific components for the sample image Kodim23 is visually depicted for better understanding.

Table 9: Performance Analysis of Ablation Studies on FERA-Net

Model Configuration	Compression Ratio	PSNR (dB)	MS-SSIM
Full FERA-Net	44.26:1	42.37	0.999
Without Multi scale feature extraction	40.5:1	40.75	0.994
Without Attention Mechanism	39.8:1	41.12	0.992
Without Latent Space Regularization	37.5:1	41.87	0.991

Figure 11: Visual Representation of Ablation Study Results

Received: August 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.9 DISCUSSION

This research demonstrates the effectiveness of FERA-Net in achieving superior image compression and reconstruction performance compared to traditional and DL based methods. Through extensive evaluations on standard datasets, including Kodak and CLIC, the proposed model consistently outperforms existing approaches in terms of compression ratio, PSNR and MS-SSIM, ensuring high quality image preservation. The integration of advanced DL techniques, optimized feature extraction and adaptive encoding strategies contribute to the model's robustness and efficiency. Additionally, visual comparison highlight the perceptual quality improvements achieved by FERA-Net, particularly in preserving fine details and structural information across varying bit rates. However, despite its promising performance, the model exhibits increased computational complexity during the training phase which could be optimized further for real time applications.

5. Conclusion

This research proposed FERA-Net an advanced DL based image compression model that significantly improves compression efficiency and reconstruction quality. Extensive experiments conducted on Kodak and CLIC datasets demonstrate that FERA-Net achieves superior performance compared to traditional and DL based compression techniques. The proposed model attains a compression ratio of 44.26:1 with high reconstruction fidelity achieving PSNR of 42.38dB and MS-SSIM of 0.9991. Comparative analysis with existing methods such as JPEG, JPEG2000, BPG, WebP, VVC and DL based models confirms that proposed model outperforms these approaches across various bitrates. Additionally, the visual comparison illustrate the model's ability to retain finer details while minimizing artifacts, making it an effective solution for high quality image compression.

Despite its strong performance, FERA-Net exhibits a higher computational cost during training which may limit its application in resource constrained environments. Future research will focus on optimizing the model, reducing the computational complexity and extending its capabilities to video compression and multi-modal data compression. Additionally, incorporating adaptive learning mechanisms to dynamically adjust compression parameters based on image content can further enhance its efficiency and generalization.

References

[1] Ijaz, U., Anwar, M. F., Ijaz, A., Kharal, M. H., Afzal, A., Iqbal, A., & Gillani, F. Lossy Image Compression Unveiled: A Comprehensive Evaluation of DCT, Wavelet Transform, and Vector Quantization.

[2] Ungureanu, V. I., Negirla, P., & Korodi, A. (2024). Image-Compression Techniques: Classical and "Region-of-Interest-Based" Approaches Presented in Recent Papers. *Sensors*, *24*(3), 791.

[3] Zhang, S., & Metaxas, D. (2024). On the challenges and perspectives of foundation models for medical image analysis. *Medical image analysis*, *91*, 102996.

[4] Joshua, B. A. (2024). *Development of an optimized multimedia compression model for improved quality of service in virtual learning environments* (Doctoral dissertation, Kampala International University).

[5] Xue, X., Marappan, R., Raju, S. K., Raghavan, R., Rajan, R., Khalaf, O. I., & Abdulsahib, G. M. (2023). Modelling and analysis of hybrid transformation for lossless big medical image compression. *Bioengineering*, *10*(3), 333.

[6] Liao, H., & Li, Y. (2024). LC-TMNet: learned lossless medical image compression with tunable multi-scale network. *PeerJ Computer Science*, *10*, e2511.

[7] Hershcovitch, M., Choshen, L., Wood, A., Enmouri, I., Chin, P., Sundararaman, S., & Harnik, D. (2024). Lossless and Near-Lossless Compression for Foundation Models. *arXiv preprint arXiv:2404.15198*.

[8] Pawłowski, P., & Piniarski, K. (2024). Efficient Lossy Compression of Video Sequences of Automotive High-Dynamic Range Image Sensors for Advanced Driver-Assistance Systems and Autonomous Vehicles. *Electronics*, *13*(18), 3651.

[9] Urbaniak, I. A. (2024). Using Compressed JPEG and JPEG2000 Medical Images in Deep Learning: A Review. *Applied Sciences*, *14*(22), 10524.

[10] Jamil, S., Piran, M. J., Rahman, M., & Kwon, O. J. (2023). Learning-driven lossy image compression: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, *123*, 106361.

[11] FANGFANG, L. DESIGN AND ANALYSIS OF EFFICIENT METHODS FOR PROVIDING A DESIRED QUALITY IN IMAGE LOSSY COMPRESSION.

Received: August 06, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[12] Jamil, S. (2024). Review of Image Quality Assessment Methods for Compressed Images. *Journal of Imaging*, *10*(5), 113.
[13] Ijaz, U., Ijaz, A., Iqbal, A., Gillani, F., & Hayat, M. (2023). Comparative Analysis of Lossless Image Compression Algorithms.
[14] Dantas, P. V., Sabino da Silva Jr, W., Cordeiro, L. C., & Carvalho, C. B. (2024). A comprehensive review of model compression techniques in machine learning. *Applied Intelligence*, *54*(22), 11804-11844.
[15] Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., & Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthcare Analytics*, 100216.
[16] Trigka, M., & Dritsas, E. (2025). A Comprehensive Survey of Deep Learning Approaches in Image Processing. *Sensors*, *25*(2), 531.
[17] Relic, L., Azevedo, R., Gross, M., & Schroers, C. (2025). Lossy image compression with foundation diffusion models. In *European Conference on Computer Vision* (pp. 303-319). Springer, Cham.
[18] Cai, S., Liang, X., Cao, S., Yan, L., Zhong, S., Chen, L., & Zou, X. (2024). Powerful Lossy Compression for Noisy Images. *arXiv preprint arXiv:2403.14135*.
[19] Li, X., Naman, A., & Taubman, D. (2024). Neural Network Assisted Lifting Steps For Improved Fully Scalable Lossy Image Compression in JPEG 2000. *arXiv preprint arXiv:2403.01647*.
[20] Duan, Z., Lu, M., Ma, Z., & Zhu, F. (2023). Lossy image compression with quantized hierarchical vaes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 198-207).
[21] Duan, Z., Lu, M., Ma, J., Huang, Y., Ma, Z., & Zhu, F. (2023). Qarv: Quantization-aware resnet vae for lossy image compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
[22] Lee, J., Cho, S., & Kim, M. (2024). An end-to-end joint learning scheme of image compression and quality enhancement with improved entropy minimization. *ETRI Journal*, *46*(6), 935-949.
[23] Tung, B. J., & Yang, S. H. Learned Image Compression with Cross Scale Attention Transformer and Asymmetric Cnn. *Available at SSRN 5050662*.
[24] Liu, J., Sun, H., & Katto, J. (2023). Learned image compression with mixed transformer-cnn architectures. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14388-14397).
[25] Kodak, E. (1993). Kodak lossless true color image suite (photocd pcd0992). *URL http://r0k. us/graphics/kodak*, *6*, 2.
[26] Fraihat, S., & Al-Betar, M. A. (2023). A novel lossy image compression algorithm using multi-models stacked AutoEncoders. *Array*, *19*, 100314.
[27] McAteer, I., Ibrahim, A., Zheng, G., Yang, W., & Valli, C. (2019). Integration of biometrics and steganography: a comprehensive review. *Technologies*, *7*(2), 34.
[28] Kabeen, K., & Gent, P. Image compression and discrete cosine transform. *College of Redwoods*.
[29] Ho, Y. H., Chan, C. C., Peng, W. H., Hang, H. M., & Domański, M. (2021). Anfic: Image compression using augmented normalizing flows. *IEEE Open Journal of Circuits and Systems*, *2*, 613-626.
[30] Yamauchi, S., & Kawamura, M. (2024). A Neural-Network-Based Watermarking Method Approximating JPEG Quantization. *Journal of Imaging*, *10*(6), 138.
[31] Jiang, Y., Cui, R., & Liu, F. (2023). Multi-resolutional human visual perception optimized pathology image progressive coding based on JPEG2000. *Signal Processing: Image Communication*, *115*, 116960
[32] Naumenko, V., Kovalenko, B., & Lukin, V. (2023). BPG-based compression analysis of Poisson-noisy medical images. *Radioelectronic and Computer Systems*, (3), 91-100
[33] Xie, Y., Cheng, K. L., & Chen, Q. (2021, October). Enhanced invertible encoding for learned image compression. In *Proceedings of the 29th ACM international conference on multimedia* (pp. 162-170).
[34] Qi, Y., Song, Y., Jia, Z., Jia, Z., Wang, Y., Zhang, L., ... & Zheng, H. (2025). Investigation of Image Compression Based on Semantic Network and Deep Residual Variational Auto-Encoder
[35] Akbari, M., Liang, J., & Han, J. (2019, May). DSSLIC: Deep semantic segmentation-based layered image compression. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2042-2046). IEEE
[36] Hoang, T. M., Zhou, J., & Fan, Y. (2020). Image compression with encoder-decoder matched semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 160-161)

Received: August 06, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[37] Luo, S., Yang, Y., Yin, Y., Shen, C., Zhao, Y., & Song, M. (2018). DeepSIC: Deep semantic image compression. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part I 25* (pp. 96-106). Springer International Publishing

[38] Lei, Z., Hong, X., Shi, J., Su, M., Lin, C., & Xia, W. (2023). Quantization-Based Adaptive Deep Image Compression Using Semantic Information. *IEEE Access*, *11*, 118061-118077.

[39] Yang, F., Herranz, L., Van De Weijer, J., Guitián, J. A. I., López, A. M., & Mozerov, M. G. (2020). Variable rate deep image compression with modulated autoencoder. IEEE Signal Processing Letters, 27, 331-335.

[40] Hu, Y., Yang, W., Ma, Z., & Liu, J. (2021). Learning end-to-end lossy image compression: A benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(8), 4194-4211

[41] He, D., Yang, Z., Peng, W., Ma, R., Qin, H., & Wang, Y. (2022). Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5718-5727).

[42] Tong, K., Wu, Y., Li, Y., Zhang, K., Zhang, L., & Jin, X. (2023, October). Qvrf: A quantization-error-aware variable rate framework for learned image compression. In *2023 IEEE International Conference on Image Processing (ICIP)* (pp. 1310-1314). IEEE

[43] Shi, Y., Ye, L., Wang, J., Wang, L., Hu, H., Yin, B., & Ling, N. (2024). Syntax-Guided Content-Adaptive Transform for Image Compression. *Sensors*, *24*(16), 5439

[44] Wang, D., Yang, W., Hu, Y., & Liu, J. (2022). Neural data-dependent transform for learned image compression. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition* (pp. 17379-17388)

[45] Wallace, G. K. (1992). The JPEG still picture compression standard. *IEEE transactions on consumer electronics*, *38*(1), xviii-xxxiv

[46] Yee, D., Soltaninejad, S., Hazarika, D., Mbuyi, G., Barnwal, R., & Basu, A. (2017, October). Medical image compression based on region of interest using better portable graphics (BPG). In *2017 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 216-221). IEEE

[47] Ginesu, G., Pintus, M., & Giusto, D. D. (2012). Objective assessment of the WebP image coding algorithm. *Signal processing: image communication*, *27*(8), 867-874

[48] Bross, B., Wang, Y. K., Ye, Y., Liu, S., Chen, J., Sullivan, G. J., & Ohm, J. R. (2021). Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(10), 3736-3764.

[49] Lee, J., Cho, S., & Beack, S. K. (2018). Context-adaptive entropy model for end-to-end optimized image compression. *arXiv preprint arXiv:1809.10452*

Received: August 06, 2025