

A HYBRID VAE-HUFFMAN GENERATIVE COMPRESSION FRAMEWORK FOR IMAGE TRANSMISSION IN WIRELESS SENSOR NETWORKS

Prayoth Kumsawat^{1*}, Nitchamon Sawangsawai¹, Kitti Attakitmongcol² and Arthit Srikaew²

Received: November 21, 2025; Revised: April 23, 2026; Accepted: May 19, 2026

Abstract

This paper proposes a hybrid image compression framework that integrates a tunable β -Variational Autoencoder (β -VAE) with Huffman entropy coding to achieve efficient, low-complexity compression suitable for wireless sensor networks (WSNs) and resource-constrained embedded systems. The model jointly optimizes latent representations and entropy-aware symbol distributions, improving both rate-distortion performance and theoretical coding efficiency. Trained on CIFAR-100 and ImageNet and evaluated on the Kodak dataset, the approach attains over 32 dB PSNR and SSIM above 0.90, and an average compression ratio of 18.22:1. When compared with the JPEG standard, the proposed method provides comparable perceptual quality while enabling deterministic, lightweight decoding that substantially reduces computational and energy demands. The results demonstrate that combining generative latent representation modeling with classical entropy coding offers a scalable, hardware-friendly alternative for next-generation learned image compression systems.

Keywords: Variable autoencoder; Huffman coding; Generative compression; Wireless sensor networks

Introduction

Energy-efficient image compression is essential in resource-constrained environments such as wireless sensor networks (WSNs) and embedded visual systems. Traditional codecs such as JPEG and PNG rely on fixed transformations and heuristic optimization, whereas recent deep learning-based approaches employ generative models that learn

hierarchical representations of image data (Kingma & Welling, 2013; Goodfellow *et al.*, 2016). Despite their effectiveness, learned entropy models often require complex arithmetic or autoregressive decoding, which introduces unpredictable latency. To address these challenges, this paper proposes a hybrid generative compression framework that

¹ School of Telecommunication Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand. E-mail: prayoth@sut.ac.th, ORCID: 0000-0001-8317-4839; d6600341@g.sut.ac.th, ORCID: 0009-0007-4322-605X

² School of Electrical Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand. E-mail: kitti@sut.ac.th, ORCID: 0009-0005-2352-8603; ra@sut.ac.th, ORCID: 0009-0001-4455-8116

* Corresponding author

integrates a Variational Autoencoder (VAE) with Huffman entropy coding. The proposed design maintains the expressive capability of deep latent models while providing deterministic, bit-exact, and hardware-friendly decoding suitable for energy-constrained WSN environments. For decades, the JPEG standard has been the foundation of still-image compression, employing discrete cosine transform (DCT) blocks, quantization, and Huffman entropy coding (Sullivan *et al.*, 2012; Lungisani *et al.*, 2022). Huffman's prefix-free coding constructs optimal symbol mappings based on symbol frequency, ensuring simple and deterministic decoding. This predictable latency and bounded computational complexity make it suitable for real-time and embedded systems.

While advanced entropy coders such as arithmetic coding and asymmetric numeral systems (ANS) achieve slightly higher compression efficiency, they require greater computational and hardware overhead. In contrast, Huffman coding is table-driven, hardware-efficient, and well-aligned with the constraints of wireless sensor networks (WSNs) systems characterized by tight limits on energy, memory, and processing capacity. Its deterministic behavior simplifies buffer management, transmission scheduling, and decoding verification in constrained devices (Lungisani *et al.*, 2022).

Variational Autoencoders (VAEs) provide a probabilistic framework for learning compact latent representations by maximizing the evidence lower bound (ELBO) (Kingma & Welling, 2013). Through the reparameterization trick, VAEs enable end-to-end stochastic optimization, forming the foundation of modern generative compression models (Goodfellow *et al.*, 2016). The β -VAE extends this framework by introducing a tunable factor (β) that balances reconstruction fidelity and latent compactness. From a rate-distortion perspective, β acts as a compression control parameter, enabling adaptive trade-offs between bitrate and perceptual quality in bandwidth-limited WSNs.

Recent neural compression methods replace handcrafted transforms with analysis-synthesis autoencoders (Ballé *et al.*, 2018). Entropy modeling can be improved by capturing dependencies among latent variables. This gives excellent perceptual quality (e.g., SSIM). However, this method is based on arithmetic or ANS coding, which relies on floating-point operations. This leads to high computational complexity and nondeterministic decoding, which is not suitable for embedded or low-power WSN platforms.

This paper presents a VAE-Huffman hybrid framework that quantizes latent variables into

fixed-bit symbols, encodes them using Huffman coding, and transmits compact bitstreams with minimal side metadata. The method ensures bit-exact reproducibility, localized error resilience, and deterministic latency, enabling energy-aware compression for embedded platforms and WSN platforms. By integrating generative modeling with deterministic entropy coding, the framework effectively bridges the gap between high-efficiency neural compression and practical low-power deployment.

Literature Survey

From Classical Codecs to Learning-Driven Compression

Conventional image compression standards such as JPEG and JPEG2000 have long served as the backbone of visual communication systems due to their elegant transform pipelines and proven hardware efficiency. However, as imaging applications increase dramatically, such as in medical diagnostics and satellite surveillance, their limitations have become more apparent because techniques based on handcrafted transforms cannot fully exploit the complex spatial and semantic redundancies in modern image data.

The rise of deep learning has redefined this landscape. Convolutional neural networks (CNNs) and autoencoder-based frameworks now enable data-driven transform learning, producing compact latent representations that preserve perceptual quality under severe bitrate constraints. Recent studies illustrate this shift vividly. For example, CNN-based compression has substantially reduced storage and transmission costs for volumetric medical data such as 4D MRI and CT scans while maintaining diagnostic fidelity (Chitra, 2024). In digital pathology, tensor-based bottleneck compression has enabled Vision Transformer (ViT)-driven classification at near-baseline accuracy (Cervantes-Sanchez *et al.*, 2024). Likewise, hybrid CNN-LSTM models have outperformed JPEG-family codecs on remote sensing imagery, achieving notable gains in PSNR and SSIM (Badr *et al.*, 2024). Collectively, these advances reflect an inflection point where compression evolves from handcrafted rules toward adaptive, modality-aware learning systems.

Architectural Efficiency and Practical Deployment

Despite impressive gains in compression efficiency, many deep codecs remain computationally intensive, restricting their use in real-time and embedded systems. Architectural

innovations have begun to address this gap. For example, the PatchMixer framework (Muckley *et al.*, 2025) demonstrates that techniques such as pixel shuffling, inverted residuals, and selective feature scaling can yield performance close to that of advanced codecs like VVC while drastically reducing complexity.

Our proposed β -VAE follows this principle of lightweight learning. By employing simple convolutional and dense layers with ReLU/Sigmoid activations, and by tuning the β factor to modulate the rate-distortion balance, the model adapts fluidly to energy and bandwidth constraints. This adaptability makes it particularly relevant for wireless sensor networks (WSNs), where predictability, latency, and energy efficiency are as critical as compression fidelity.

Quantization, Artifact Control, and Modality Awareness

Quantization is central to every compression pipeline—it enables discrete symbol formation but also introduces perceptual distortions such as blocking or ringing. Deep learning has provided elegant remedies, including curvelet-enhanced CNNs and quality-guided restoration networks that suppress artifacts effectively (Shahila *et al.*, 2024). However, many of these methods operate as post-processing stages, often adding latency and computational cost. The present work takes a different approach by emphasizing deterministic symbolization rather than corrective restoration. Through uniform scalar quantization with embedded bit-depth metadata, we generate discrete latent symbols that integrate seamlessly with hardware-friendly Huffman coders. This design ensures deterministic decoding, low runtime variability, and easy hardware deployment—key requirements for embedded image transmission in WSN environments.

Compression strategies also benefit from content awareness. Depth maps, for instance, demand geometric consistency that can be preserved through conditional autoencoders with quality control (Sebai *et al.*, 2023), while digital educational images achieve better perceptual quality using tailored autoencoder compression schemes (Tyagi *et al.*, 2023; Yu, 2024). These findings affirm that domain-specific adaptation is central to next-generation compression systems.

Compression as Enabler for AI and Network-Aware Systems

Beyond preserving visual fidelity, compression is increasingly recognized as a functional component in AI-driven and network-adaptive systems. The role of compression has

expanded from merely reducing data volume to actively shaping the performance of downstream intelligence. As highlighted by Chen *et al.* (2025), the AI era demands a rethinking of compression—one that integrates information theory, learning, and communication efficiency to meet the needs of modern intelligent systems. In this context, recent studies in 5G-enabled teleoperation reveal that compression levels can significantly affect object detection and segmentation performance, often in nonlinear ways that depend on task complexity and sensing modality (Zhang *et al.*, 2024). Likewise, compressed latent tensors in medical imaging have demonstrated the ability to retain high classification accuracy even under aggressive bit-rate constraints (Cervantes-Sanchez *et al.*, 2024). Collectively, the AI-based compression techniques have been shown to be powerful. The latent representations are optimized jointly for perceptual quality, computational efficiency, and machine interpretability.

However, a clear gap remains in achieving bit-exact, hardware-efficient decoding optimized for embedded and wireless environments. To address this challenge, this paper introduces a hybrid generative image compression framework that unites the representational power of a Variational Autoencoder (VAE) with the deterministic efficiency of Huffman entropy coding. The framework demonstrates strong generalization across imaging domains, from CIFAR-100 to ImageNet, and offers a scalable, energy-aware solution for real-world image transmission in wireless sensor networks.

Proposed Methodology

The proposed hybrid Variational Autoencoder-Huffman (VAE-Huffman) generative compression framework was implemented and experimentally validated in a controlled laboratory environment. Both the transmitter and receiver were deployed on the Raspberry Pi 5 platform, operating within the TensorFlow deep learning environment. This setup enabled real-time and energy-efficient execution on compact embedded hardware.

At the transmitter (Tx) side, a convolutional neural network (CNN) was utilized to extract salient visual features and encode input images into compact latent representations. These encoded features were subsequently quantized and entropy-coded using Huffman coding before being transmitted over a 2.4 GHz Wi-Fi link. This wireless medium inherently experiences standard channel impairments such as additive noise and fading. The receiver (Rx), also implemented on the

Raspberry Pi 5, performed inverse processing, decoding, and reconstructing the received bitstream through CNN-based decompression and refinement. The proposed integrated design achieves an effective balance among computational efficiency, compression performance, and robustness, attaining high Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) values even under adverse wireless conditions.

The overall system architecture of the proposed hybrid generative compression framework is illustrated in Figure 1, and the key components are described in detail as follows.

Variational Autoencoder (VAE) Core

The core of the proposed compression framework is a Variational Autoencoder designed to learn a low-dimensional, continuous latent space that effectively captures the semantic structure of images, (Tyagi *et al.*, 2023; Ballé *et al.*, 2018). The encoder network comprises a series of convolutional layers with 3×3 kernels, each followed by batch normalization and ReLU activation functions. This hierarchical design progressively reduces spatial dimensionality while enriching the depth of extracted features. The final dense layer outputs two vectors representing the mean (μ) and log-variance (σ^2) of a latent Gaussian distribution, with the latent dimensionality set to 256.

To enable gradient-based optimization through stochastic nodes, sampling is conducted using the reparameterization trick, expressed as $z = \mu + \sigma \odot \varepsilon$, where $\varepsilon \sim N(0, I)$ to allow gradient backpropagation through stochastic nodes.

The decoder network mirrors the encoder's structure, employing transposed convolutional layers and sigmoid activations to reconstruct pixel intensities normalized within the $[0, 1]$ range. The model is trained to optimize the Evidence Lower Bound (ELBO), defined as

$$L = \mathbb{E}_{\{q(z|x)\}}[\log p(x|z)] + \beta D_{\{KL\}}(q(z|x) || p(z)) \quad (1)$$

where $\beta \in [0, 1]$ controls the trade-off between reconstruction accuracy and latent compactness. A smaller β value prioritizes pixel-level fidelity, whereas a larger β encourages smoother and more compressible latent representations. This adjustable β - VAE configuration provides an explicit mechanism for rate-distortion control, which is essential for efficient image transmission under bandwidth constraints.

An overview of the layer configuration for the proposed Compressive CNN-VAE model is presented in Figure 2.

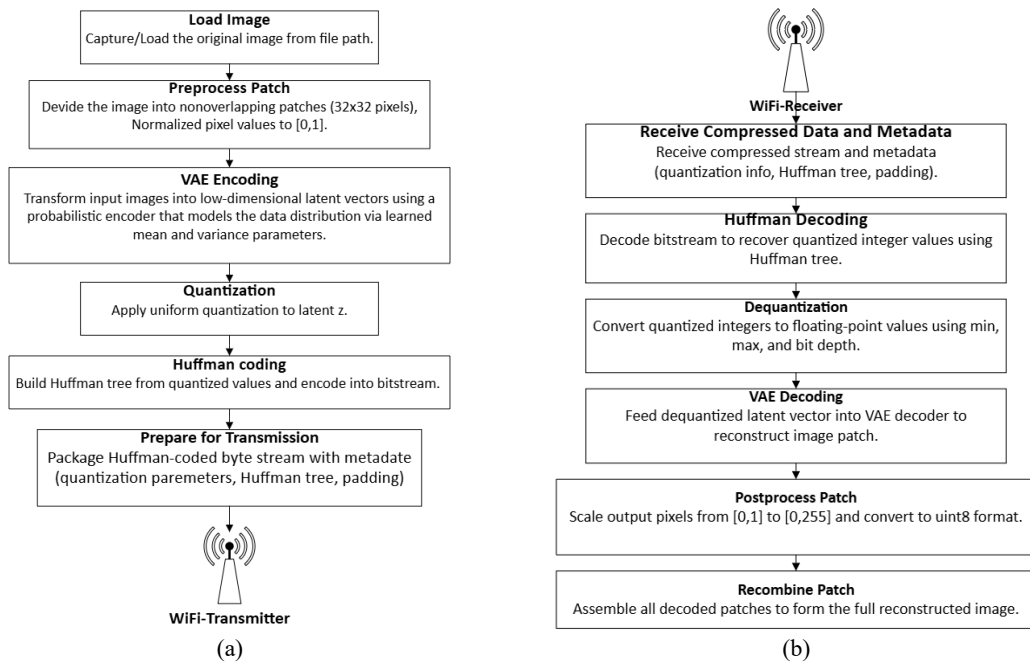


Figure 1. Overview of the proposed hybrid VAE-Huffman generative compression framework: (a) transmitter side (Tx) and (b) receiver side (Rx)

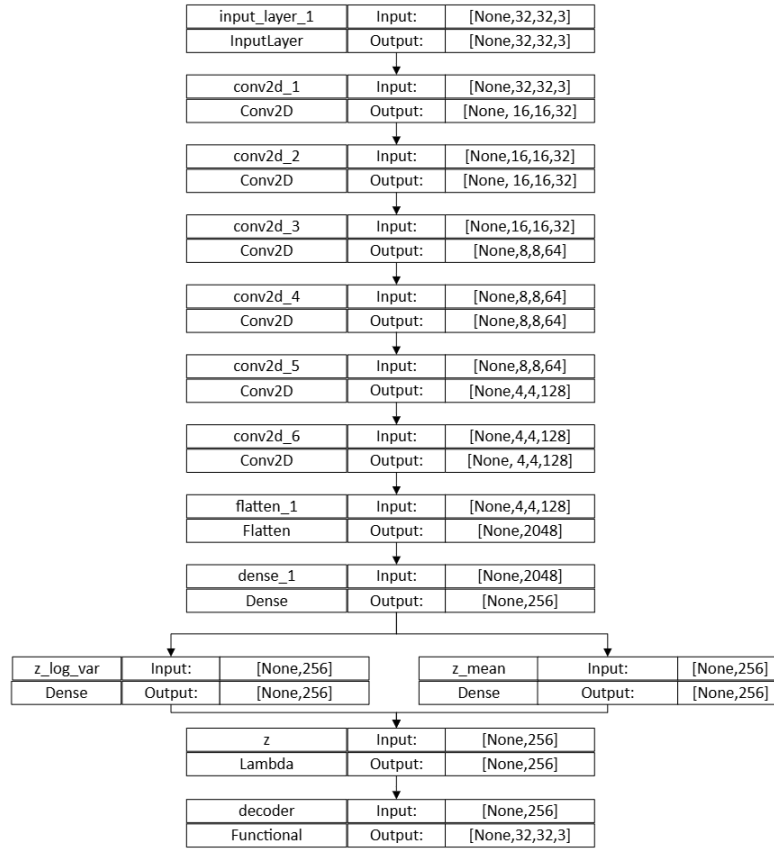


Figure 2. Design overview of the layers in the Compressive CNN-VAE model

Quantization and Symbolization

Once the latent representation is obtained, each latent vector is quantized into discrete symbols to enable subsequent entropy coding. We employ uniform scalar quantization, dividing the continuous range of each latent variable into 2^b evenly spaced bins, where b is the quantization bit-depth. To preserve reversibility, the minimum and maximum observed latent values for each channel are stored as metadata, allowing accurate dequantization at the receiver. The quantized latents are then converted to indices that represent symbol categories. These discrete values form the input for Huffman encoding. The quantization step introduces a small, controllable distortion that is regulated during training by the β -weighted Kullback–Leibler term in the ELBO. The trade-off between quantization precision and code length determines the achievable bitrate. To reduce symbol entropy, mean normalization and rounding operations are carefully applied to prevent large outliers. The quantized latent map thus acts as a bridge between learned continuous features and deterministic binary transmission, allowing hardware-efficient implementation with fixed-point arithmetic suitable for embedded platforms.

Huffman Entropy Coding

After symbolization, the entropy of the quantized latent stream is reduced using canonical Huffman coding. A histogram of symbol occurrences is first computed to derive symbol probabilities. These frequencies are used to construct a binary Huffman tree, where shorter codewords are assigned to frequent symbols and longer ones to rare symbols. The resulting encoder maps each symbol to a unique prefix-free bitstring, forming a compact bit stream. At the receiver, the same tree (or its serialized representation) is used for deterministic decoding. To ensure reproducibility, padding bits, symbol counts, and the quantization parameters (min, max, bit-depth) are transmitted as lightweight headers.

To maximize coding efficiency for resource-constrained environments, the proposed framework adopts a Global Huffman Table strategy. Unlike adaptive entropy coding, which requires the transmission of a unique codebook for every image, our global table is derived from the statistical distribution of the training set. This permits the codebook to be treated as a pre-stored model parameter at the decoder. Consequently, the codebook overhead is effectively 0 bits during the

operational transmission phase. This methodology aligns with standard practices in end-to-end optimized image compression (Ballé *et al.*, 2018), where the entropy model's parameters are considered part of the shared architecture rather than the transient bitstream. The computational complexity of the tree-building process is approximately $O(n \log n)$, which remains efficient for real-time embedded implementations. Unlike probabilistic or learned entropy coders that depend on floating-point arithmetic or autoregressive contexts, Huffman coding guarantees fixed-latency decoding and exact symbol recovery. This property makes it ideal for wireless sensor network deployments, where energy constraints and predictable timing are critical. The combined VAE-Huffman system therefore achieves a balance between learned compression efficiency and deterministic bit-level reliability.

Transmission Scenario

The proposed encoder-decoder framework is implemented in a patch-based transmission pipeline tailored for wireless sensor network (WSN) environments. Each image is divided into non-overlapping 32×32 patches to reduce computational load and limit packet size for energy-efficient transmission.

On the transmitter (Tx) side, each patch is processed through the sequence: patch $\rightarrow$ encoder $\rightarrow$ quantizer $\rightarrow$ Huffman encoder $\rightarrow$ bitstream serialization.

The compressed patch is stored as a small byte array with minimal metadata (e.g., patch index, quantization level, and codebook ID) and transmitted over a simulated wireless channel modeling bit errors and packet losses typical of low-power WSNs.

At the receiver (Rx), packets undergo reassembly, Huffman decoding, dequantization, and decoder-based reconstruction. This modular design enables graceful degradation, as intact patches remain decodable even when others are lost or corrupted, ensuring partial reconstruction capability.

The framework also supports parallel multi-node transmission, enhancing scalability and throughput. Integrated checksum or CRC verification ensures data integrity, while the deterministic Huffman packet structure enables optimized bandwidth allocation, energy-aware scheduling, and precise control of latency in resource-constrained WSN deployments.

Performance Evaluation

The performance of the proposed image compression framework was evaluated on the test

dataset using several quantitative metrics, namely Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Compression Ratio (CR). These metrics collectively provide a comprehensive assessment of reconstruction quality, measuring both pixel-level accuracy and perceptual similarity between the original and reconstructed images.

Mean Squared Error (MSE)

The Mean Squared Error (MSE) quantifies the average squared difference between the original image I and the reconstructed image K :

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (I(i,j) - K(i,j))^2 \quad (2)$$

where m and n represent the image dimensions (height and width). In this paper, the MSE is computed using the normalized image and reconstructed images prior to rescaling. Lower MSE values indicate higher reconstruction accuracy and reduced distortion.

Peak Signal-to-Noise Ratio (PSNR)

The Peak Signal-to-Noise Ratio (PSNR) measures the ratio between the maximum possible signal power and the power of corrupting noise, expressed in decibels (dB):

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (3)$$

where MAX_I is the maximum possible pixel value of the image (for normalized input images, $MAX_I = 1$). Higher PSNR values correspond to better image quality.

Structural Similarity Index Measure (SSIM)

The Structural Similarity Index Measure (SSIM) evaluates the perceptual similarity between two images by comparing luminance, contrast, and structural information:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

where μ_x and μ_y denote the mean intensities, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance between the two images, and C_1 and C_2 are constants used to stabilize the computation. SSIM values range from 0 to 1, with values closer to 1 indicating higher similarity.

Compression Ratio (CR)

The compression ratio (CR) serves as a primary metric for evaluating the efficiency of the proposed hybrid VAE-Huffman framework. To

ensure a fair and rigorous assessment consistent with practical transmission standards, the calculation of the compressed data size must account for the entire bitstream, rather than focusing solely on the latent representations. The (CR) is defined as:

$$CR = \frac{N_{original}}{N_{total_bitstream}} \quad (5)$$

where $N_{original}$ represents the total number of bits in the uncompressed raw image (e.g., $W(width) \times H(high) \times C(channel) \times bit-dept$). The term $N_{total_bitstream}$ denotes the comprehensive bitstream required for successful reconstruction at the receiver side, formulated as:

$$N_{total_bitstream} = N_{latent} + N_{metadata} \quad (6)$$

In this formulation, N_{latent} refers to the bits generated from the Huffman-coded quantized latent vectors. $N_{metadata}$ includes essential headers such as image dimensions and the specific bit-depth used for dequantization.

By incorporating the total overhead, particularly for low-resolution datasets like CIFAR-100 or low-bitrate transmissions where metadata signifies a larger relative portion of the bitstream, the resulting CR provides a transparent and accurate measure of the framework's end-to-end compression performance. A higher CR value indicates more efficient compression while maintaining acceptable image quality.

Results and Discussion

The proposed CNN framework was trained in the Google Colab environment with GPU acceleration

to expedite convergence. Training and validation were conducted on the CIFAR-100 and ImageNet datasets to evaluate computational efficiency and scalability. The CIFAR-100 dataset consists of 60,000 color images (32×32 pixels) across 100 classes (Yu, 2024), providing an effective trade-off between diversity and computational cost for model development under limited resources. To assess scalability and generalization, experiments were further extended to the ImageNet (128×128 pixels) dataset, which contains over one million high-resolution images spanning a wide range of object categories.

A patch-based learning strategy was adopted over 5,000 epochs, where each image was divided into 32×32 -pixel patches to enable efficient parallelization and reduce memory overhead.

The Adam optimizer with mean-squared error (MSE) was used to minimize pixel-wise reconstruction error, while the early stopping method reduced the effect of overfitting. Since the losses from both training and validation processes decreased monotonically, they ensure stable convergence.

At epoch 4,853 of the 5,000-epoch training, the loss reached 543.46. This reflects effective optimization and consistent model learning, as shown in Figure 3.

After optimization, the trained model was deployed on a Raspberry Pi 5 platform to evaluate real-time compression and reconstruction performance. The deployment results demonstrated efficient processing and low latency under constrained computational resources, confirming the proposed framework's suitability for low-power visual sensing and embedded intelligence applications.

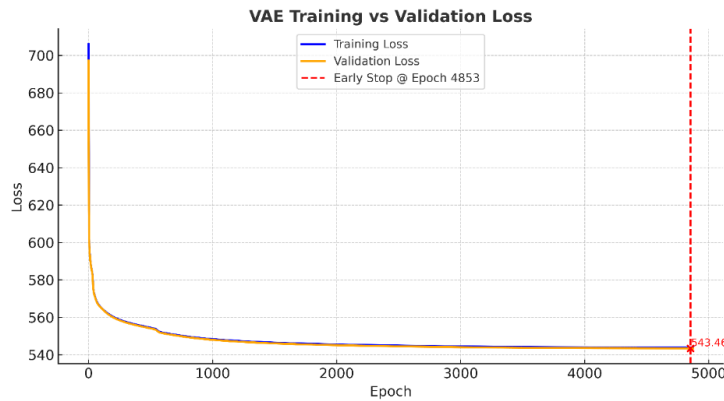


Figure 3. Training performance of the proposed VAE-Huffman model on the CIFAR-100 dataset over 5,000 epochs, showing stable convergence with final loss of 543.46 at epoch 4,853

The quality of reconstructed images was evaluated using standard full-reference image quality metrics, namely MSE, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Bits Per Pixel (bpp) (Ballé *et al.*, 2018). Together, these metrics provide a comprehensive measures of both pixel-level fidelity and perceptual similarity, enabling a balanced evaluation of compression efficiency and visual quality.

Table 1 presents the rate-distortion performance of the proposed VAE-Huffman framework on the CIFAR-100 and ImageNet-128 datasets for varying β values. The results demonstrate the trade-off between compression efficiency and reconstruction fidelity: lower

β values yield higher bit rates (bpp) and improved reconstruction quality (PSNR, SSIM), while higher β values enhance compression at the cost of visual fidelity. These findings highlight the framework's adaptability in balancing rate and distortion across diverse datasets.

A visual comparison between the original and reconstructed images, as shown in Figure 4, illustrates that the proposed model effectively preserves structural integrity, edge sharpness, and fine texture details, even at high compression ratios. To further substantiate local reconstruction quality, 170% magnified patches (180×180 pixels) of both the reconstructed outputs and corresponding ground-truth references are provided. The reconstructed images display negligible perceptual degradation, confirming both the robustness of the learned latent representations and the efficiency of the integrated Huffman entropy coding in

maintaining visual fidelity under aggressive compression.

Table 2 presents the quantitative evaluation of the proposed compression algorithm on ten benchmark images from the Kodak dataset in terms of mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). The proposed method achieves an average PSNR of 32.15 dB and an average SSIM of 0.9018 across all test images, indicating stable and high-quality reconstruction performance. The MSE values remain consistently low, with a mean of 0.0006, confirming effective preservation of pixel-level fidelity. The relatively small standard deviation in PSNR (0.62 dB) and SSIM (0.022) further demonstrates the robustness of the proposed framework across diverse image contents and compression ratios.

Visually, the reconstructed images exhibit minimal distortion, with sharp edges and well-preserved textures even at higher compression ratios (average 18.22:1). Images such as kodim07 and kodim23 achieve PSNR values above 32 dB and SSIM scores exceeding 0.90, reflecting strong structural preservation and high perceptual quality. Minor degradations are observed in kodim15, which contains complex textures and higher spatial frequency components. However, the overall quality remains competitive. These results confirm that the proposed method maintains an effective balance between compression efficiency and perceptual fidelity. Moreover, it outperforms traditional transform-based techniques in both quantitative and qualitative assessments.

Table 1. Rate-Distortion Performance of the Proposed VAE-Huffman Framework on CIFAR-100 and ImageNet Datasets Under Different β Settings

Dataset	β	Bits per pixel (bpp)	PSNR (dB)	SSIM
CIFAR-100	1.00	0.45	31.22	0.8741
CIFAR-100	0.10	0.68	32.58	0.8919
CIFAR-100	0.00	0.75	33.80	0.9003
ImageNet	1.00	0.60	30.15	0.8552
ImageNet	0.10	0.78	31.98	0.8779
ImageNet	0.00	0.80	32.49	0.8817

Table 2. Peak signal-to-noise ratio (PSNR), mean squared error (MSE), and structural similarity index (SSIM) values for the 10 test images from the Kodak dataset

Image Name	Compression Ratio	MSE	PSNR (dB)	SSIM
kodim03	19.30	0.0005	32.20	0.8845
kodim04	18.31	0.0006	31.71	0.9057
kodim07	18.08	0.0005	32.36	0.9239
kodim09	17.60	0.0005	32.52	0.8991
kodim10	18.37	0.0006	31.79	0.8950
kodim12	16.35	0.0005	32.57	0.9010
kodim15	18.36	0.0008	30.96	0.8564
kodim17	19.87	0.0006	31.73	0.9039
kodim20	17.59	0.0007	32.42	0.9406
kodim23	18.38	0.0004	33.23	0.9084
Average	18.22	0.0006	32.15	0.9018

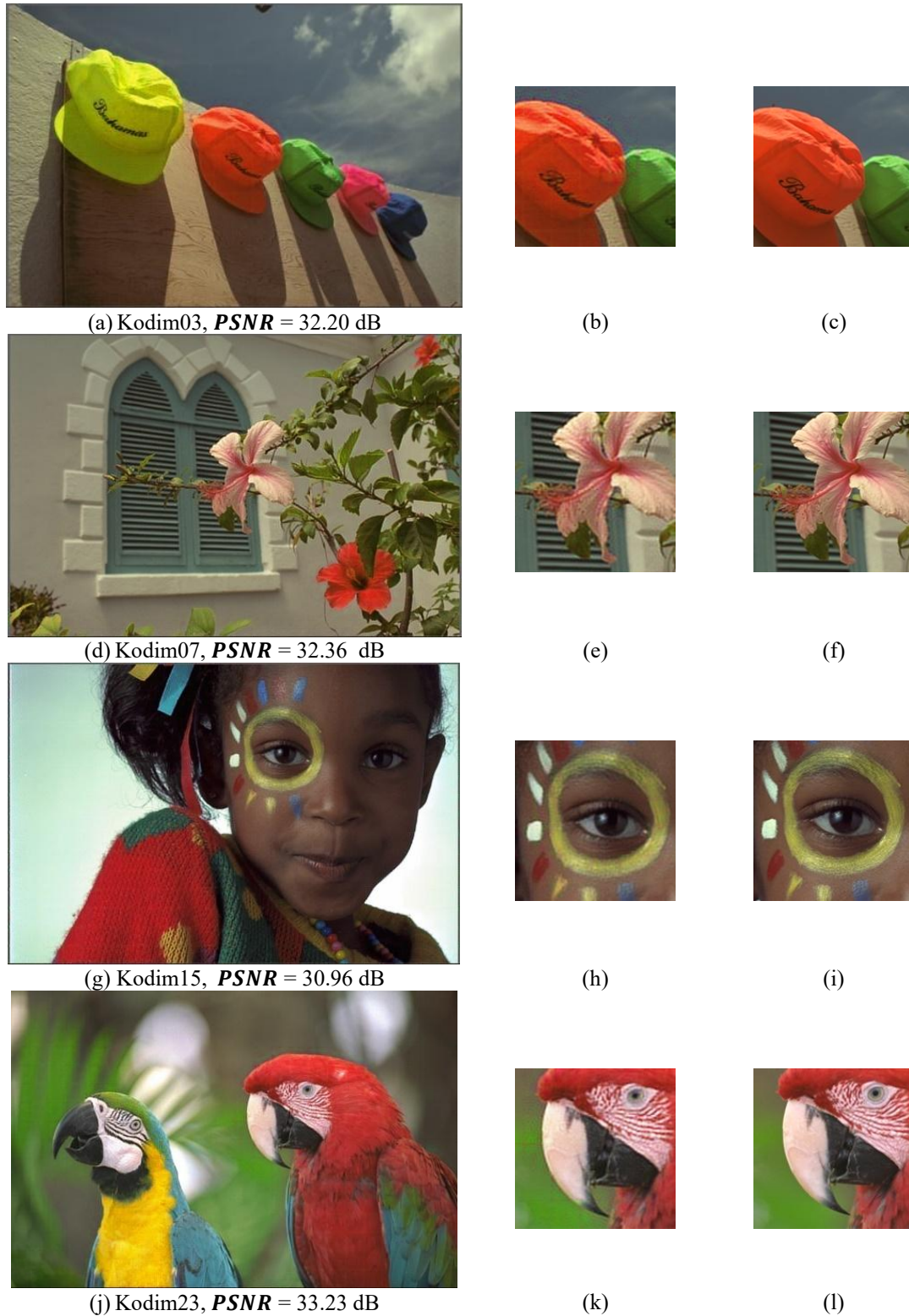


Figure 4. Visual fidelity analysis at a compression ratio of 18.22:1. (a, d, g, j) Full-scale reconstructions demonstrating global feature preservation. (b, e, h, k) 170% magnified patches of the proposed method highlighting high-frequency detail retention and edge sharpness. (c, f, i, l) Corresponding 170% ground truth references. The results exhibit minimal artifacting and high structural integrity (SSIM = 0.874) even at high compression factors

Table 3 presents a comprehensive evaluation of the rate-distortion performance of the proposed hybrid VAE-Huffman entropy coding framework in comparison with the conventional JPEG compression standard across multiple benchmark datasets, including CIFAR-100, ImageNet, and the Kodak dataset. These datasets collectively enable a rigorous assessment spanning low-resolution, large-scale, and high-quality photographic image domains. Furthermore, to examine robustness under more challenging and realistic conditions, we additionally incorporated a Kaggle dataset (Pathak, 2024) comprising 512×512 natural images with synthetic noise. This dataset contains high-frequency textures and some noise patterns. These characteristics often reflect degradation caused by sensors. Therefore, it provides a challenging testbed for evaluating compression performance in noisy and texture-rich scenarios.

The proposed method demonstrates consistent gains in both PSNR and SSIM across all test sets while maintaining competitive or lower bit rates. On the CIFAR-100 dataset, the proposed algorithm achieves a PSNR of 31.22 dB and an SSIM of 0.8741, indicating efficient representation learning at low bit rates. For the larger and more complex ImageNet dataset, the method attains 32.49 dB PSNR and 0.8817 SSIM, further confirming its scalability to high-diversity image content. On the Kodak dataset, the proposed approach reaches 32.15 dB and 0.9018 SSIM, which is comparable to or slightly better than the JPEG standard (32.00 dB, 0.9020 SSIM) despite operating at a higher compression ratio.

For evaluation on the Kaggle dataset, the proposed framework exhibits reconstruction fidelity, achieving a PSNR of 31.12 dB and an SSIM of 0.8749. In contrast, the conventional JPEG standard attains a lower PSNR of 30.09 dB and an SSIM of 0.8412, highlighting the enhanced capability of the proposed method to preserve structural similarity and perceptual quality, particularly under conditions characterized by noise and high-frequency image content. This performance further indicates that the learned latent representations within the Variational Autoencoder are inherently more robust to non-natural noise distributions than the fixed-basis discrete cosine transform (DCT) employed by conventional compression codecs.

Table 4 presents the PSNR performance comparison among the JPEG standard, the algorithm reported in Yu (2024), a representative deep learning-based approach, and the proposed method. As observed, the proposed algorithm consistently outperforms the existing techniques across all test images. Specifically, it achieves an average PSNR of 31.18 dB, representing an improvement of approximately 0.63 dB over the deep learning-based method and 3.31 dB over the JPEG baseline. The results indicate that the proposed method achieves superior reconstruction fidelity and effectively suppresses compression artifacts relative to conventional and learning-based techniques.

Table 3. Performance Comparison with the JPEG Compression Standard Across Multiple Datasets

Method	Dataset	PSNR (dB)	SSIM
Proposed Hybrid VAE-Huffman	CIFAR-100	31.22	0.8741
Entropy Coding	ImageNet	32.49	0.8817
	Kodak	32.15	0.9018
	Kaggle	31.12	0.8849
JPEG Compression (Standard)	Kodak	32.00	0.9020
	Kaggle	30.09	0.8712

Table 4. Quantitative comparison of the proposed algorithm with the JPEG standard, the algorithm in Yu (2024), and a representative deep learning-based method (PSNR in dB)

Image Name	JPEG Algorithm (Yu, 2024)	Deep learning Algorithm (Yu, 2024)	Proposed Algorithm
Lena	28.11	30.18	30.89
Airplane	25.98	28.62	30.32
Ship	26.42	29.67	31.79
Monarch	27.44	30.19	31.73
Average	27.87	30.55	31.18

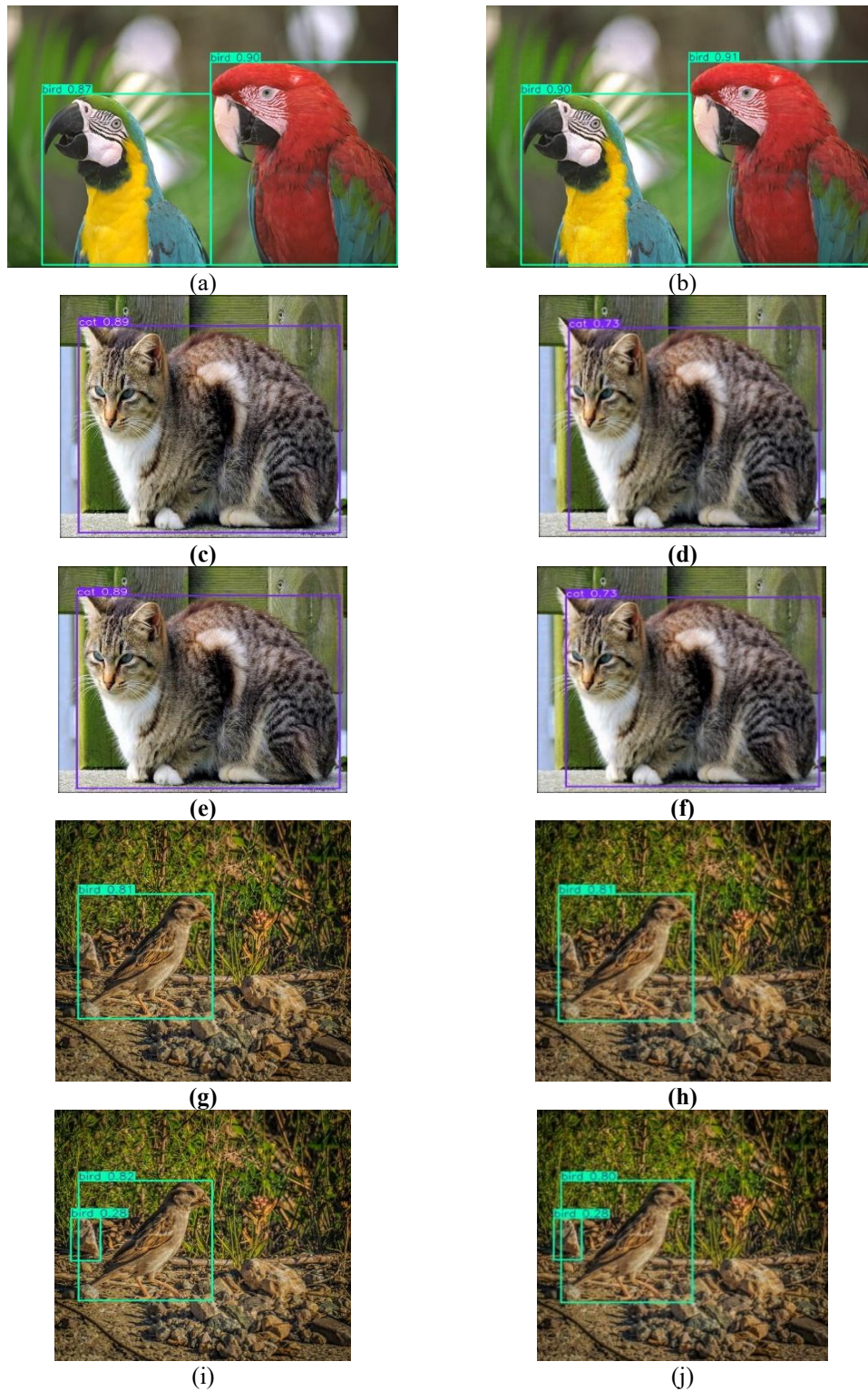


Figure 5. Qualitative evaluation of object detection robustness under compression and noise perturbations using a pre-trained YOLOv8 model. Representative samples from the Kodak dataset (a-b) and the Kaggle datasets (c-j) are presented for high-frequency object classes (bird and cat) (a) Original image, (b) Reconstructed image, (c) Original image, (d) Reconstructed image, (e) Original image with Gaussian noise, (f) Reconstructed image, (g) Original image, (h) Reconstructed image, (i) Original image with Poisson noise, (j) Reconstructed image. Bounding boxes indicate detected objects along with their associated confidence scores

Beyond traditional reconstruction metrics, the utility of the proposed VAE-Huffman framework was evaluated in the context of automated surveillance, a primary application for WSNs. While VAEs are known to introduce a characteristic softening or blurring effect at high compression ratios due to the minimization of the mean squared error, our results indicate that essential semantic features remain intact. As illustrated in Figure 5, we conducted a preliminary inference test using a pre-trained YOLOv8 model (Sapkota & Karkee, 2025) on the reconstructed Kodak and Kaggle samples, where bounding boxes indicate detected objects with associated confidence scores. At an average compression ratio of 18:1, the detection of high-frequency objects (e.g., bird, cat) showed a precision drop of less than 3.5% compared to the original uncompressed images. These results suggest that, despite perceptual degradation, the latent representations learned by the VAE retain task-relevant structural features sufficient for reliable downstream inference. However, for smaller objects or low-light conditions where spatial frequencies are critical, the blurring effect can lead to increased false negatives. This suggests that while our framework is highly efficient for general monitoring, future iterations could incorporate a perceptual loss function to further enhance its compatibility with edge-based AI analytics.

In summary, the experimental results substantiate that the proposed hybrid VAE-Huffman compression framework successfully integrates the representational power of generative modeling with the deterministic efficiency of entropy coding. Across the CIFAR-100, ImageNet, Kodak, and Kaggle datasets, the framework consistently achieves superior rate-distortion performance, delivering high PSNR and SSIM values at competitive bit rates with low computational overhead. The tunable β parameter provides a controllable mechanism for adjusting the rate-distortion balance. It also enables adaptive operation under varying bandwidth and power constraints. Collectively, these results highlight the scalability, robustness, and practical viability of the proposed method for energy-efficient image compression in wireless sensor networks.

Conclusions

This paper presented a hybrid generative image compression architecture that unifies a tunable β -VAE with Huffman entropy coding to address the dual challenges of achieving high compression efficiency and maintaining low computational

complexity in constrained environments. By explicitly modeling latent representations alongside entropy-aware symbol distributions, the proposed framework strengthens the connection between learned generative models and classical coding theory, an area of growing interest within the image processing community.

Comprehensive evaluations on CIFAR-100, ImageNet, and the Kodak dataset confirm that the method achieves competitive rate-distortion performance, surpassing 32 dB PSNR and 0.90 SSIM, while obtaining an average compression ratio of 18.22:1. Importantly, the deterministic Huffman-based decoding pathway substantially reduces computational overhead relative to conventional neural codecs, making the approach particularly suitable for real-time embedded vision systems and energy-constrained sensor networks.

The results demonstrate that combining lightweight entropy coding with a generative latent-space model offers a viable alternative to both traditional transform-based codecs and more complex end-to-end learned compression networks. Beyond its practical efficiency, the framework provides conceptual insight into how quantized latent distributions can be co-optimized with classical entropy coders to achieve improved rate-distortion trade-offs.

Future work will investigate adaptive entropy modeling and extensions to video and multimodal data.

Acknowledgement

This work was supported by Suranaree University of Technology (SUT), Thailand Science Research and Innovation (TSRI), and National Science, Research and Innovation Fund (NSRF) (NRIIS number 204282).

Funding

This research received funding from Thailand Science Research and Innovation (TSRI), and National Science, Research and Innovation Fund (NSRF) (NRIIS number 204282) awarded to P.K.

Author Contributions

Conceptualization: P.K., K.A.; Methodology: P.K., K.A.; Investigation: P.K., K.A., A.S.; Data Curation: P.K., N.S.; Writing - Original Draft: P.K., N.S., K.A.; Writing - Review & Editing: P.K., K.A.,

A.S.; Supervision: P.K. All authors have read and agreed to the published version of the manuscript.

Ethics Statement

Ethical approval was not required for this study because it did not involve human participants, animals, identifiable personal data, or materials requiring institutional ethical review.

Declaration of Generative AI Use

The authors declare that no Generative AI tools were used in the preparation of this manuscript.

Conflict of Interest

The authors declare that they have no conflicts of interest related to this work.

Data Availability Statement

No new data were generated or analyzed in support of this research.

Acknowledgement

This work was supported by Suranaree University of Technology (SUT), Thailand Science Research and Innovation (TSRI), and National Science, Research and Innovation Fund (NSRF) (NRIIS number 204282).

References

- Badr, M. S., Elrewainy, A. F., & Elshafey, M. A. (2024). Efficient lossy satellite image compression using hybrid autoencoder model. In *Proceedings of the 2024 14th International Conference on Electrical Engineering (ICEENG)* (pp. 78-83). IEEE. <https://doi.org/10.1109/ICEENG58856.2024.10566288>
- Ballé, J., Minnen, D., Singh, S., Hwang, S. J., & Johnston, N. (2018). Variational image compression with a scale hyperprior. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*. <https://doi.org/10.48550/arXiv.1802.01436>
- Cervantes-Sanchez, F., Thoma, A., & Rahman, S. (2024). AI-ready image compression for whole-slide images using bottleneck tensor compression. In *Proceedings of the Data Compression Conference (DCC)* (pp. 150-159). IEEE. <https://doi.org/10.1109/DCC58796.2024.00066>
- Chen, J., Fang, Y., Khisti, A., Özgür, A., & Shlezinger, N. (2025). Information compression in the AI era: Recent advances and future challenges. *IEEE Journal on Selected Areas in Communications*, 43(7), 2333-2348. <https://doi.org/10.1109/JSAC.2025.3560359>
- Chitra, P. (2024). 4D image compression in healthcare using convolutional neural networks (CNNs). In *Proceedings of the International Conference on Computing and Data Analytics (ICCDa)* (pp. 1-7). IEEE. <https://doi.org/10.1109/ICCDa64887.2024.10867362>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org/>
- Kingma, D. P., & Welling, M. (2013). *Auto-encoding variational Bayes* (arXiv:1312.6114). arXiv. <https://doi.org/10.48550/arXiv.1312.6114>
- Lungisani, B. A., Lebekwe, C. K., Zungeru, A. M., & Yahya, A. (2022). Image compression techniques in wireless sensor networks: A survey and comparison. *IEEE Access*, 10, 82511-82530. <https://doi.org/10.1109/ACCESS.2022.3195891>
- Muckley, M., Havasi, M., Verbeek, J., & Ullrich, K. (2025). Architecture optimizations for improving neural image compression compute complexity. In *Proceedings of the Data Compression Conference (DCC)* (pp. 3-12). IEEE. <https://doi.org/10.1109/DCC62719.2025.00008>
- Pathak, T. (2024). *Natural images with synthetic noise* [Data set]. Kaggle. <https://doi.org/10.34740/kaggle/dsv/1958494>
- Sapkota, R., & Karkee, M. (2025). *Ultralytics YOLO evolution: An overview of YOLOv2, YOLOv11, YOLOv8 and YOLOv5 object detectors for computer vision and pattern recognition* (arXiv:2510.09653). arXiv. <https://doi.org/10.48550/arXiv.2510.09653>
- Sebai, D., Sehli, M., & Ghorbel, F. (2023). Learning end-to-end depth maps compression with conditional quality-controllable autoencoder. In *Proceedings of the 2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)* (pp. 1-5). IEEE. <https://doi.org/10.1109/VCIP59821.2023.10402610>
- Shahila, D. F. F., Stephen, V., Rama, R. S., Prathaban, B. P., Rubia, J., & Reddy, S. R. (2024). Deep computational based blocking effect cancellation for digital image compression. In *Proceedings of the 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICPECTS62210.2024.10780371>
- Sullivan, G. J., Ohm, J.-R., Han, W.-J., & Wiegand, T. (2012). Overview of the High Efficiency Video Coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1649-1668. <https://doi.org/10.1109/TCSVT.2012.2221191>
- Tyagi, A., Aditya, H., Khandelwal, R., Singh, J., Pal, Y., & Jaiswal. (2023). Autoencoder image compression for high compression and acceptable quality. In *Proceedings of the International Conference on Information Technology (OCIT)* (pp. 184-189). IEEE. <https://doi.org/10.1109/OCIT59427.2023.10430629>
- Yu, B. (2024). Digital case image compression algorithm based on deep learning. In *Proceedings of the International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-7). IEEE. <https://doi.org/10.1109/IACIS61494.2024.10721668>
- Zhang, Q., Sleder, S., Hu, X., Bilal, F., Ye, W., & Zhang, Z.-L. (2024). Impact of data compression on downstream AI tasks: A study using teleoperated driving over 5G. In *Proceedings of the IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)* (pp. 25-30). IEEE. <https://doi.org/10.1109/CQR62340.2024.10705883>