Enhancement of Image Transmission through ML Based Joint Source-Channel Coding

Beza Mezmur Hawaz[1] Anand Anbalagan (PhD)[2] Yosef Kefyelaw Enku (PhD)[3*]
[1,2,3]Department of Electrical and Electronics Engineering, Faculty of Electical Electronics and Information Communication Technology, FDRE Technical and Vocational Training Institute, Addis Ababa, Ethiopia.
*Corresponding author Email: enkuyosef2000@gmail.com

ABSTRACT

Reliable wireless image transmission under low signal-to-noise ratio (SNR) and limited bandwidth remains a significant challenge because conventional communication systems based on Shannon's separation theorem often suffer from performance degradation and the cliff effect under adverse channel conditions. Although Deep Joint Source–Channel Coding (DeepJSCC) has shown promising performance, further improvements are needed to enhance image reconstruction quality and transmission robustness. This study proposes an Enhanced Deep Joint Source–Channel Coding (EDeepJSCC) framework based on convolutional neural networks (CNNs) for end-to-end wireless image transmission over Additive White Gaussian Noise (AWGN) and Rayleigh fading channels. The proposed framework jointly optimizes source and channel coding to improve reconstruction quality without requiring explicit channel estimation. Experiments were conducted using the CIFAR-10 dataset, consisting of 50,000 training images and 10,000 testing images, and implemented in TensorFlow using the Adam optimizer under multiple SNR conditions and bandwidth ratios. Performance was evaluated using Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and SNR. Experimental results demonstrate that the proposed EDeepJSCC framework achieves improved image reconstruction quality and more graceful degradation than conventional JPEG/BPG with LDPC-based transmission, particularly under low-SNR and bandwidth-constrained conditions. These findings indicate that EDeepJSCC is a promising solution for robust wireless image transmission in next-generation communication systems.

KEYWORDS
Deep Joint Source–Channel Coding; Enhanced Deep Joint Source–Channel Coding; Convolutional Neural Networks; Image Reconstruction

1. Introduction

The rapid growth of multimedia services, Internet of Things (IoT) applications, autonomous systems, remote sensing, and intelligent healthcare has significantly increased the demand for reliable wireless image transmission. These applications require high-quality image delivery with low latency and high reliability despite limited bandwidth and varying channel conditions. Shannon's Separation Theorem, introduced by Claude E. Shannon, is one of the fundamental principles of digital communication theory. The theorem states that, under ideal conditions with infinitely long codewords and sufficiently complex coding schemes, source coding (data compression) and channel coding (error correction) can be designed and optimized independently without sacrificing overall communication performance. This principle forms the theoretical foundation of most conventional digital communication systems. The lowest

possible source coding rate needs to be less than the channel capacity for a source to be reliably transmitted across a channel; according to Shannon's separation theorem, reliable communication is possible when the source coding rate is lower than the channel capacity. Based on this principle, conventional wireless image transmission systems employ separate source coding techniques, such as JPEG or Better Portable Graphics (BPG), followed by channel coding methods such as Low-Density Parity-Check (LDPC) codes. Without compromising efficacy, independently design the source and channel coding. This well-known theorem provides modularity, which is the basis for the remarkable success of real-world communication systems. A source-independent channel code (LDPC) is used to encode the bit channel in mature coincident wireless picture transmission systems, and a source coding method (BPG/JPEG) is used to compress the image (Yang et al., 2022).

Improvements in high-performance computer facilities have led to a rise in the popularity of deep-learning methods that make use of deep neural networks. Deep learning has enormous strength and versatility since it can work with unstructured data and assess a vast number of attributes. Layers of data processing are used in deep learning, with each layer gradually extracting features. Low-level features are captured by the first layers, and these features are combined by the next layers to create a comprehensive representation (Mathew et al., 2021).

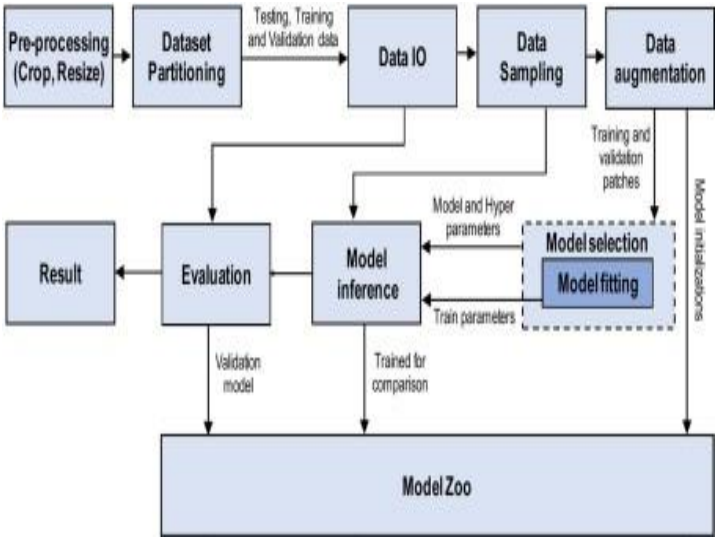

Fig.1. Deep learning

An Artificial Neural Network (ANN) with a deep feed-forward architecture is called a Convolutional Neural Network (CNN), or ConvNet. It works very well for learning and generalizing abstract properties from spatial input, such as photographs. Multiple processing layers that collect characteristics at different degrees of abstraction make up a deep CNN model. While the deeper layers concentrate on high-level features (low abstraction), the first layers extract low-level features (high abstraction). Figure 1 shows the general architecture of a CNN, which includes the different types of layers. These are covered in more detail in the sections that follow. In many computer vision applications today, such as image classification, object detection, face detection, speech recognition, vehicle recognition, facial expression recognition, and more, CNNs are a top method for getting impressive results.

Existing approaches still face challenges in improving reconstruction quality, robustness, and progressive image transmission under varying channel conditions. Therefore, there is a need to develop an enhanced DeepJSCC framework capable of providing more reliable and efficient wireless image transmission of 60,000 color images that make up the CIFAR-10 and CIFAR-100 datasets, which were first presented by Krizhevsky & Hinton in 2009. Ten thousand of them are set aside for testing (validation), and fifty thousand are used for training. The photos are classified into 10 and 100 classes, respectively, and a top-1 error is used to evaluate the classification performance.

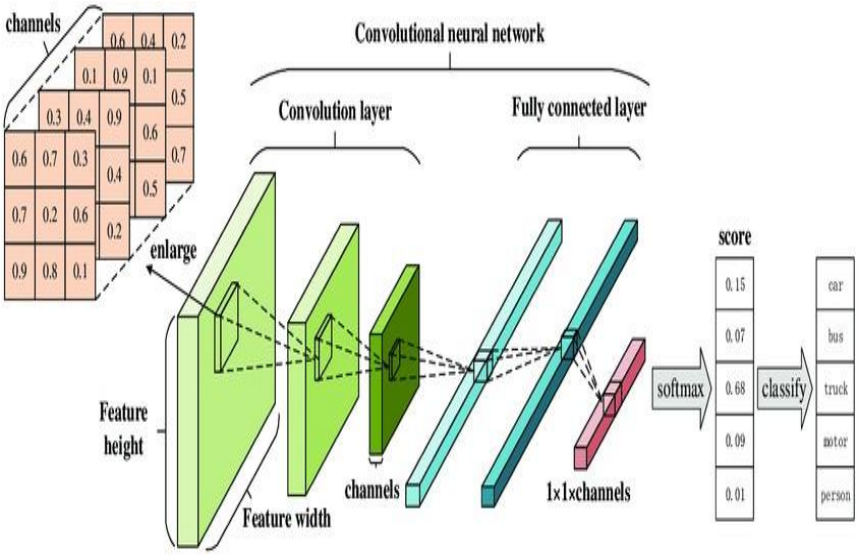

Fig.2. Concept Model Of CNN

Fig.3. image Of Cifar10 Dataset

Finding an encoder-decoder pair that reliably and efficiently transmits a compressible source across a noisy channel is the goal of joint source-channel coding (JSCC). By using deep neural networks as autoencoders, deep learning has improved JSCC for photos. Neural networks make up the encoder-decoder pair in deep JSCC. The transmitted codeword is a lower-dimensional code that is created by compressing the source input by the encoder network (Lee et al., 2023). Source and channel coding are two crucial elements in the current data transmission process. The former retains important information needed to reconstruct the signal with a given degree of fidelity by lowering redundancy in the source signal. For example, popular picture compression techniques like JPEG AND BPG enable a reduction in communication load with little to no compromise in reconstruction quality. Channel coding provides structural redundancy to enable dependable decoding in the presence of channel defects. An example of a communication system diagram with distinct source and channel codes (Tung et al., 2022).

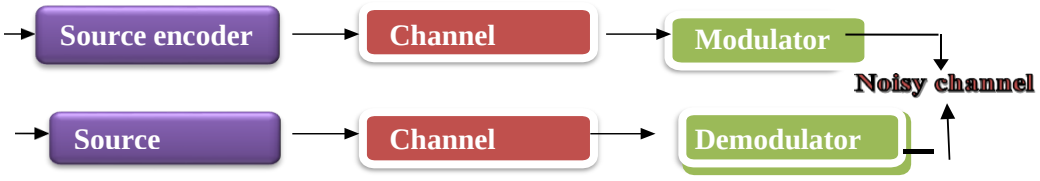

FiG.4. Diagram of a typical separation-based digital image communication system

A systematic two-step encoding method is used by modern communication systems to efficiently transfer images. The image data is then subjected to source coding, which uses algorithms to eliminate duplication and minimize the amount of data that needs to be sent. To provide effective transmission over communication channels and optimize bandwidth use, this compression is essential. Error-correcting coding is the second step applied to the compressed bitstream. Redundancy in the data is systematically added via error-correcting codes. Receivers can detect and correct errors caused by noise, interference, or other channel impairments during transmission thanks to this redundancy. The system's ability to incorporate error-correcting capabilities improves data transmission robustness and dependability, which is especially important for preserving image quality and integrity under a variety of communication scenarios.

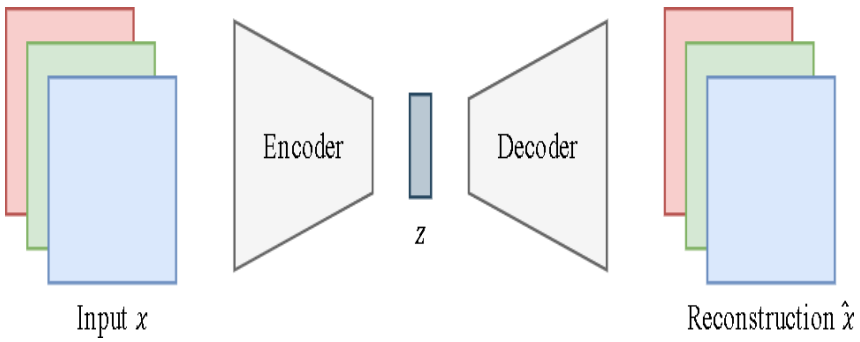

Fig.5. General structures on an autoencoder

This method is theoretically supported by Shannon's separation theorem. It suggests that the best course of action for source and channel blocks that are infinitely long is to divide the coding tasks into separate tasks. This way, each coding technique can be optimized separately, resulting in the best possible overall system performance in terms of error resilience and data compression efficiency (Bourtsoulatze et al., 2019). Deep learning, particularly through autoencoder architectures, has revolutionized image compression. Encoder-decoder pipelines based on autoencoders offer efficient compression. Advanced techniques, like generative adversarial training, have achieved remarkably low compression rates, enhancing image quality and transmission efficiency significantly in recent years (Cheng et al., 2019).

Autoencoders serve as a robust tool in unsupervised Learning, focusing on learning latent representations by minimizing reconstruction loss. This approach supports tasks like sparse and robust representation learning, where noisy data can be effectively modeled and reconstructed to preserve original information fidelity (Jiang et al., 2019). Learning a meaningful latent representation that captures relevant features for the given task, without needing to directly encode every pixel's grey value. This ability makes autoencoders versatile for tasks ranging from data compression to feature extraction in various domains (Michelucci, 2022).

By adding a pilot signal as an extra input to the decoder, the JSCC technique can be extended to multi-user settings with varying SNRs and adjusted to changing SNRs. A known pilot signal can be sent from the transmitter to the receiver during real wireless transmission. Then, to help with decoding, the receiver calculates the SNR using this pilot signal (Ding et al., 2021). With deep neural network (DNN) optimization, source and channel coding are simultaneously optimized to reconstruct the source signal and image at the receiver with the least amount of distortion. However, in some situations, this method might lead to a noticeable decrease in perceptual quality (Yang & Kim, 2022).

1.1 Statement of the Problem

Wireless image transmission has become an essential component of modern communication systems; however, achieving high-quality image reconstruction under noisy wireless channels remains challenging. Conventional communication systems based on separate source and channel coding exhibit significant performance degradation in low-SNR and bandwidth-limited environments. Although DeepJSCC has improved transmission efficiency by jointly optimizing source and channel coding, existing methods still experience limitations in reconstruction quality, robustness against channel variations, and progressive image transmission. These limitations motivate the development of an enhanced DeepJSCC framework that can provide improved image reconstruction accuracy, greater robustness under different channel conditions, and more reliable wireless image transmission.

1.2. Research Objectives

The objectives of the study are to:

(1) design an end-to-end EDeepJSCC architecture using convolutional neural networks;

(2) evaluate the proposed framework under AWGN and Rayleigh fading channels;

(3) compare the proposed method with conventional JPEG/BPG and LDPC-based communication systems, and

(4) evaluate system performance using PSNR, MSE, and SNR.

2. Literature Review

Recent advances in deep learning and wireless communication have significantly transformed image compression and transmission systems, leading to the development of advanced Joint Source–Channel Coding (JSCC) approaches that overcome the limitations of traditional separated source and channel coding techniques. The following studies highlight major developments in deep image compression, neural network-based JSCC, progressive image transmission, and channel coding strategies, providing the foundation for identifying existing challenges and research gaps.

The state-of-the-art approach introduces several advancements over previous studies that rely on single models. First, the recurrent architecture has been enhanced to improve spatial information propagation within the network's hidden states (Johnston et al., 2018). The algorithm optimizes bit allocation across visually complex regions of an image, thereby improving encoding efficiency. Additionally, during training, a pixel-wise loss function weighted by Structural Similarity Index (SSIM) metrics is employed. This weighting strategy enhances reconstruction quality by aligning the training process with perceptual image quality assessments.

A unique approach that incorporates a high-level prior into an end-to-end trainable model for image compression using Variational Autoencoders (VAEs) has also been introduced. To achieve effective compression, this prior improves the model's ability to capture spatial dependencies within latent representations. Furthermore, the model demonstrates superior rate–distortion performance compared with existing artificial neural network (ANN)-based techniques when evaluated using both conventional metrics, such as Peak Signal-to-Noise Ratio (PSNR), and perceptual quality metrics. This dual emphasis on visual fidelity and quantitative distortion measures demonstrates the potential of VAEs to effectively balance compression efficiency and image reconstruction quality (Ballé et al., 2018).

The separation technique serves as the foundation for most contemporary communication systems. This preference is attributed to the modularity of separately designed source and channel coding components and the absence of a Joint Source–Channel Coding (JSCC) architecture that achieves comparable efficiency with acceptable coding and decoding complexity (Yang et al., 2022).

TurboAE is a fully end-to-end neural encoder–decoder framework that achieves significant improvements in channel coding. The approach utilizes neural networks to learn efficient representations of transmitted data, achieving lower word error rates compared with conventional methods. This advancement demonstrates the potential of deep learning for enhancing communication reliability and efficiency in transmitting structured data over noisy channels (Jiang et al., 2019). Similarly, an adaptive TDBC protocol with optimal mode selection, relay selection, and power allocation under a data-rate fairness constraint was proposed for two-way multi-relay transmission systems. Furthermore, a low-complexity

JSCC technique was introduced to improve the transmission of progressive error-resilient JPEG bitstreams through this system (Bi & Liang, 2017).

The proposed JSCC technique for wireless image transmission utilizes convolutional neural networks (CNNs) to directly map image pixels into complex-valued channel symbols. Notably, it avoids the cliff effect and provides graceful performance degradation under varying signal-to-noise ratio (SNR) conditions. The approach demonstrates robustness across different channel environments, including slow Rayleigh fading, by learning noise-resilient representations. It consistently outperforms traditional digital communication systems based on separation principles across various SNR levels and bandwidth constraints (Bourtsoulatze et al., 2019).

Various strategies and models for progressive refinement in image transmission have been explored, including multiple-decoder architectures, residual image transmission, and single-encoder approaches. These techniques demonstrate strong performance under low SNR conditions and limited bandwidth, while maintaining graceful degradation across different testing environments. Neural networks facilitate effective progressive transmission by preserving high performance under challenging communication conditions. Joint training of a single encoder with multiple receivers achieved optimal results; however, alternative approaches, such as residual transmission and single-decoder architectures, also demonstrated comparable effectiveness, providing flexibility according to system requirements. Overall, these approaches highlight the adaptability and robustness of neural network-based methods for achieving efficient image transmission in complex communication environments (Kurka & Gunduz, 2019).

In another study, an advanced deep image compression algorithm was developed to present a novel digital strategy for image retrieval over wireless channels. The approach incorporated capacity-achieving channel codes and introduced a JSCC technique to reliably transmit feature vectors over extremely limited channel bandwidths. The findings demonstrated that the proposed autoencoder-based JSCC scheme outperformed both conventional digital schemes and JSCC schemes without feature decoding. These results highlight the importance of deep neural network (DNN)-based JSCC techniques for addressing the stringent latency and bandwidth requirements of wireless image retrieval applications (Jankowski et al., 2020; Kurka & Gunduz, 2020).

A novel autoencoder-based JSCC scheme for two-user image transmission over noisy channels integrates noiseless output feedback to improve reconstruction quality at the receiver, particularly for wireless communication applications. This approach represents an advancement in deep JSCC by adapting to diverse SNR conditions and extending scalability to two-user scenarios, thereby improving transmission reliability under varying channel conditions (Ding et al., 2021). Furthermore, a policy network was incorporated to optimize the rate–quality trade-off, enhancing transmission efficiency and signal fidelity across different communication environments (Yang & Kim, 2022). Similarly, MLSC-Image demonstrates strong performance under high compression ratios and low SNR conditions by emphasizing comprehensive image feature extraction. This approach highlights the potential of semantic feature extraction for improving image transmission quality under challenging

communication conditions (Zhang et al., 2022).

The performance of image transmission has also been evaluated using an LDPC-coded massive MIMO with an OFDM system. This model improves simulation results in terms of Bit Error Rate (BER), and LDPC coding has demonstrated effectiveness in image recovery. Performance metrics, including PSNR and Root Mean Square Error (RMSE), indicate superior performance, particularly under low SNR conditions (Hasan & Jassim, 2020). LDPC codes are characterized by sparse parity-check matrices that enable efficient encoding and decoding processes. They are widely recognized for their strong error-correction capability and decoding efficiency, achieving performance close to the Shannon limit, especially with long code lengths (Wang, 2021).

The reviewed related works help identify the research gaps, design considerations, and methodological approaches relevant to the proposed thesis. Numerous investigations have been conducted to improve system performance and evaluate robustness under different conditions. However, challenges remain, including the effects of mismatched channel conditions, such as variations in SNR and the number of multipath components used during JSCC training, as well as the vulnerability of JSCC systems to signal clipping. Despite these challenges, deep joint source–channel coding (DJSCC) has enabled various applications for enhanced image transmission, including low-latency, high-resolution, and low-complexity imaging systems. DJSCC has therefore emerged as a promising approach for addressing advanced image transmission challenges in wireless communication environments.

2.1.Comparative Analysis of Existing Studies

The reviewed literature demonstrates that Deep Joint Source–Channel Coding (DeepJSCC) has significantly improved wireless image transmission by jointly optimizing source and channel coding. Previous studies have reported improved image reconstruction quality and graceful degradation under low signal-to-noise ratio (SNR) conditions compared with conventional separation-based communication systems. However, most existing approaches primarily focus on reconstruction performance under specific channel conditions without comprehensively evaluating transmission robustness across different bandwidth constraints and fading environments. Furthermore, limited attention has been given to enhancing progressive image transmission while maintaining computational efficiency. These limitations indicate the need for an enhanced DeepJSCC framework capable of providing improved robustness, reconstruction quality, and transmission reliability under diverse wireless communication scenarios.

Conventional separation-based communication systems under identical experimental settings. Therefore, there is a need to develop an enhanced Deep Joint Source–Channel Coding (EDeepJSCC) framework that improves reconstruction quality, transmission reliability, and robustness for wireless image transmission under diverse channel conditions.

Table 1. Literature summary table

No. Ref.	Year / Author	Algorithm	SNR	PSNR	Dataset	Bandwidth
1	2018 Mikolaj, Deniz & Krystian	SGD optimization, learning rate = 0.001 and batch size = 32, 64; compression = JPEG2000	SNRs (-6-6 dB)	PSNR (0.0-0.5 dB)	ImageNet	B = 64 or B = 512
2	2019 David & Deniz	Training images 32 x 32; learning rate = 0.001 and batch size = 64; compression = JPEG	SNRs (1-20 dB)	PSNR (22-25 dB)	CIFAR-10	B = 64 or B = 512
3	2020 David & Deniz	Training images 32 x 32; learning rate = 0.0001 and batch size = 64; compression = JPEG	SNRs (1-25 dB)	PSNR (22-25 dB)	CIFAR-10	K = 512 K = 1024
4	2021 Ding, Li, Ma & Fan	Training images 32 x 32; learning rate = 0.0001 and batch size = 128; compression = JPEG	SNRs (1-20 dB)	PSNR (22-25 dB)	CIFAR-10	K = 512 K = 1024
5	2022 Kurka & Gunduz	Training images 32 x 32; learning rate = 0.002 and batch size = 128, 256; compression = JPEG and BGP	SNRs (-2-20 dB)	PSNR (24-34 dB)	CIFAR-10 MNIST	Bandwidth ratio k/n = 0.5

2.2.Research Gap

Although several studies have demonstrated the effectiveness of Deep Joint Source–Channel Coding (DeepJSCC) for wireless image transmission, significant challenges remain. Existing methods primarily focus on improving reconstruction quality under specific channel conditions, while limited attention has been given to enhancing transmission robustness across varying SNR levels and bandwidth-constrained environments. Furthermore, many studies do not comprehensively evaluate progressive image reconstruction or compare proposed methods with conventional separation-based communication systems under consistent experimental settings. Therefore, there is a need to develop an enhanced DeepJSCC framework that improves image reconstruction quality, transmission reliability, and robustness under diverse wireless channel conditions while maintaining efficient bandwidth utilization.

3. Materials and Methods

This study adopts an end-to-end deep learning approach to improve wireless image transmission using the proposed Enhanced Deep Joint Source–Channel Coding (EDeepJSCC) framework. The methodology consists of dataset preparation, encoder design, channel modeling, decoder implementation, model training, and performance evaluation. The proposed framework jointly optimizes source and channel coding to improve image reconstruction quality under noisy wireless communication channels. The overall workflow

of the proposed methodology is illustrated in Figure 6. JSCC makes use of CNNs. As a non-trainable layer, the channel is incorporated, which guarantees differentiability by adding random values at each realization. Together, NN components receive training to maximize system performance in end-to-end communications. This constitutes the core elements of the DeepJSCC architecture, which uses neural encoders and decoders. Models with one encoder and two decoders are examples of variations; these models investigate different approaches for peak performance. Throughout this paper, **x** denotes the input image, **z** represents the encoded channel symbols, **ẑ** denotes the received noisy symbols, **θ** and **ϕ** represent the encoder and decoder parameters, respectively, **k** denotes the channel bandwidth, and **μ** represents the channel signal-to-noise ratio (SNR). These symbols are used consistently throughout the mathematical formulation of the proposed EDeepJSCC framework. The encoder is a CNN and is represented by the deterministic function $\mathbf{f}^{\theta}$ parameterized by vector **θ**. It receives an input from the source image **x**. and outputs at once all the channel symbols z. i.e. have $\mathbf{Z} = \mathbf{f}^{\theta}(x)$ with $\mathbf{Z} \in C^{k}$. Where k is the total bandwidth. $k = \sum^{L}$

channel input is **Z** = (**Z**j‸ **Z**L) where **Zj** is transmitted over the Ĉ*j*-th channel consider that, for each valid subset ℓ of channel output vectors received, a different decoder is

employed to transform the noisy symbols into reconstruction $\bar{x}_{\ell}$

Thus, the decoder is a CNN

represented by $g\ell\ \varphi\ell$ where $\Phi\ell$ is the learned parameter vector. the concatenation of all channel outputs for subset ℓ by **ẑ**ℓ, i.e... **ẑ**ℓ=Uj∈ℓ**ẑ**j. The corresponding reconstruction is given by:

$$\hat{x}_{\ell} = g_{\{\varphi\ell\}}(\hat{z}_{\ell}) \qquad (3.1)$$

Optimize all the parameters jointly to minimize the average distortion between the input image

x and a partial reconstruction $\bar{x}_{\ell}$:

$$(\theta^*, \varphi^*) = \arg\min_{\{\theta, \varphi\ell\}} E_{\{p(x, \hat{x}_{\ell})\}} [\, d(x, \hat{x}_{\ell}) \,] \qquad (3.2)$$

Where *Φ*is the collection of all decoder parameter vectors. d(**x,** $\bar{x}$ ℓ) is a given distortion measure,

and p(**x,** $\bar{x}$)ℓ is the joint probability distribution of the original and reconstructed images,

Which depends on the channel and input image statistics, as well on the encoder and decoder parameters. Note that this is a multi-objective problem, as multiple reconstructions are considered and the parameter vector **θ** is common in the optimization of every reconstruction.

The system modeled as an AE with 1 encoder and L decoders

$$(1 / L\,N) \sum_{\ell=1}^{L} \sum_{j=1}^{N} d(x_j, \hat{x}_{\{\ell,j\}}) \qquad (3.3)$$

where N is the number of training samples and d(x, $\bar{x}$)is the MSE distortion.

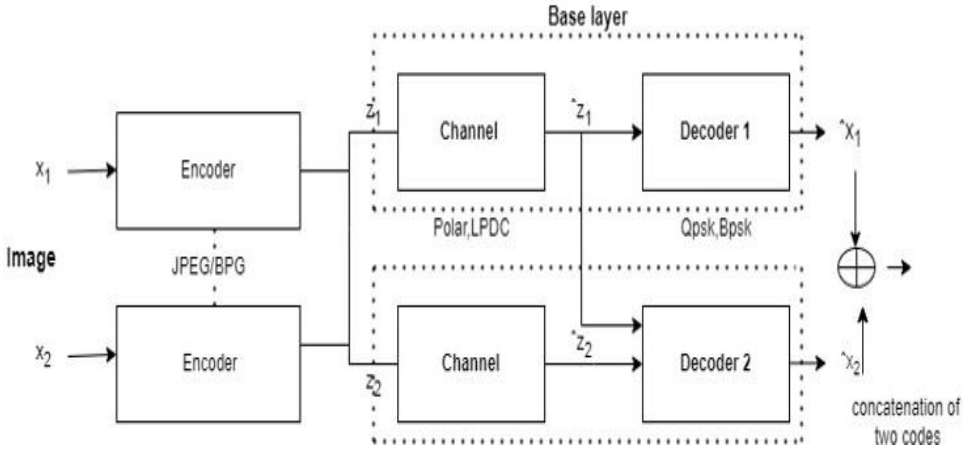

Fig.6. Block diagram of EdeepJSCC

3.1.Channel model

Consider a block diagram of EdeepJSCC shown in Fig. 6. Channel SNR is known both at the joint source-channel encoder and the joint source-channel decoder. Wireless transmission of an image, where the input image of size height x weigh x Channel is represented by a vector $\mathbf{x}\in R^n$. Where n = HxWxC and $\mathbb{R}$ denotes the set of real numbers. The joint source-channel encoder encodes **x** and the SNR μ, and the encoding function $\mathbf{f}\theta:\mathbb{R}^n x\mathbb{R}\ C^k$ leads to a vector of complex-valued channel input symbols $\mathbf{Z}\in C^k$ The encoding process can be expressed as:

An average power constraint is imposed on the transmitted signal at every channel. i.e,

$$(1 / K\_i)\ E[Z\_i^* Z\_i] \leq P, \quad \forall i \in [L] \tag{3.4}$$

$$Z = f\_\theta(x, \mu) \in C\text{^}k \tag{3.5}$$

k is the size of channel input symbols. θ is the parameter set of the joint source-channel encoder, μ $\in\mathbb{R}$ is the channel decoder and feedback to the joint source-channel encoder, and C denotes the set of complex numbers. The encoder maps the n-dimensional vector of complex-valued image **x** to a k-dimensional vector of verage power constraint at THE JSCE, ^{1}kis also imposed, where $\mathbf{Z}^*$ denotes the complex conjugate of **Z.** The encoded symbols **Z** are transmitted over a noisy channel represented by the function η: C^kC^k AWGN channel is considered. The channel output symbols $\hat{\mathbf{z}}\in C^k$ receivedby the joint source-channel decoder are expressed as:

$$\hat{z} = \eta(Z) = Z + \omega \tag{3.6}$$

The vector $\boldsymbol{\omega}\in C^k$ consists of independent and identically distributed (i.i.d) samples with the distribution CN $(0, \sigma^2\boldsymbol{I})\sigma^2$ is the noise power, and CN $(\cdot,\cdot)$ denotes a circularly symmetric complex Gaussian distribution. The proposed method can be applied to other differentiable channels; consider the following fading channel:

$$\hat{z} = hZ + \omega \tag{3.7}$$

where $h\in C$ is the channel gain. Applying equalization at the receiver, the model can be represented as Eq. while the noise has a different distribution. The EdeepJSCC function.

Φ: $C^k x\ \mathbb{R}\ \ \mathbb{R}^n$ to map $\hat{\mathbf{z}}$ and μ as follows:

$$\hat{x} = g\_\Phi(\hat{z}, \mu) = g\_\Phi(\eta(f\_\theta(x, \mu)), \mu) \tag{3.8}$$

Where $\bar{x} \in R^n$ is the original image **x,** Φ is the parameter set of the EdeepJSCC. The distortion

$\bar{x}$

between the original image **x** and the reconstructed image is performance is evaluated by the PSNR between the input image x and a reconstruction x j. The MSE is inversely proportional to the Peak signal noise ratio. The PSNR metric measures the ratio between the maximum possible power of the signal and the power of the noise that corrupts the signal. The PSNR is defined as follows:

$$d(x, \hat{x}) = (1 / n) \sum\_\{i=1\}^\{n\} (x\_i - \hat{x}\_i)^2 \quad (3.9)$$

$$PSNR\_\ell = 10 \log_{10}(MAX^2 / MSE\_\ell) \quad (3.10)$$

Where the average SNR is given by: $\|x - \hat{x}\_\ell\|^2$ (3.11)

$$SNR = 10 \log_{10}(P / \sigma^2) \text{ dB} \quad (3.12)$$

4. Simulation and Result

Enhancing image transmission through deep learning involves integrating techniques from both compression and error correction, which enhances the reliability and efficiency of image transmission over a communication channel. Here's a step-by-step simulation and result process:

Step 1: EDJSCC Model Design

1.1. Design a deep learning architecture: Combine two components into a single neural network.

1.2. Source coding: Use techniques like autoencoders or convolutional neural networks (CNNs) to compress images.

1.3. Channel coding: Incorporate error-correction coding schemes LDPC.

1.4. Joint optimization: Train the model end-to-end to jointly optimize JSCC objectives.

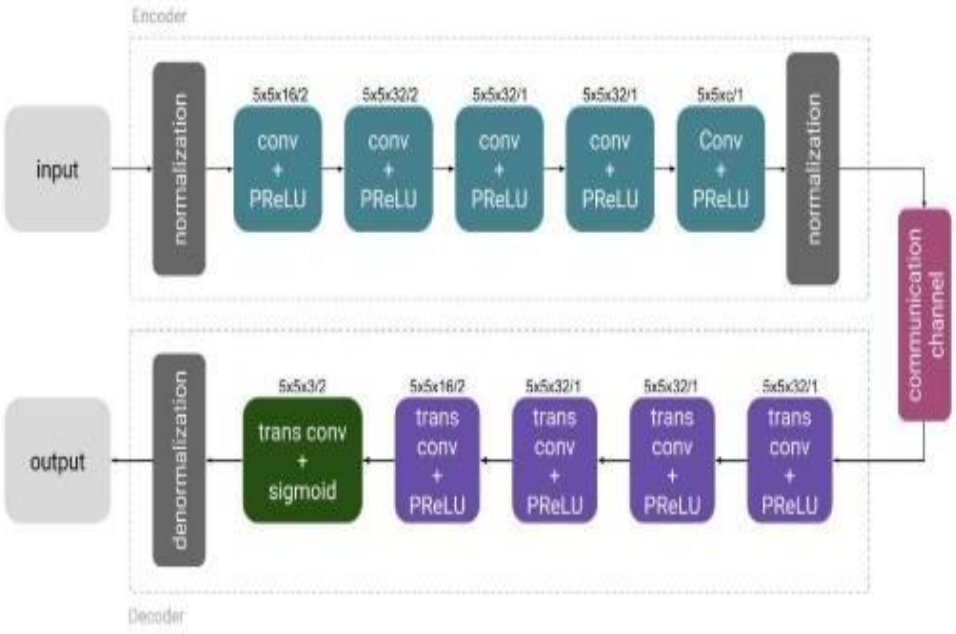

Fig.7. The architecture of the encoder and decoder for the CNN scheme

The CNN architectures are used for the encoder and decoder components. The encoder and decoder are symmetric, containing the same number of convolutional layers and trainable weights. The convolutional layers are responsible for feature extraction and downsampling

(at the encoder) or upsampling (at the decoder) through the stride and varying the depth of the output space. The last layer of the encoder is parameterized by depth c. which defines the total bandwidth ratio of all the layers combined: k/n = (H/4 x W/4 x c)/(HxWx3) = c/48, where H and W are the height and width of an image with 3 color channels. After each convolution, it uses the parametric ReLU (PReLU) activation function, or a sigmoid in the last block of the decoder, to produce outputs in the range [0,1]. Normalization at the beginning of the encoder and denormalization at the end of the decoder convert values from the range [0, 255] to [0, 1] (and vice versa). At the end of the encoder, the output of the last convolution layer is normalized as in Eq. (1) so the average power of each layer is constrained to P = 1. Note that the total channel bandwidth k used by DeepJSCC depends on the input dimension n; that is, the same model can output different channel bandwidths for different input sizes, keeping the bandwidth ratio constant. Thus, will consider the bandwidth ratio (k/n) when presenting and comparing results. The Adam technique for stochastic gradient descent with a learning rate of 0.0001 and a batch size of 256 images for training was used in the TensorFlow simulations. With dimensions of n= 332×32×3, the CIFAR-10 dataset contains 10,000 test images in addition to 50,000 training images. This dataset is used to train and assess the fully convolutional model. The average PSNR calculated for the full CIFAR-10 test dataset is used to report the results. Ten distinct realizations of the noisy channel were used to transmit each image to evaluate its robustness and performance under various circumstances.

The proposed EDeepJSCC framework employs a fully convolutional encoder–decoder architecture. The encoder consists of four convolutional layers with 3 × 3 convolution kernels and progressively increasing feature channels to extract high-level image representations. The decoder employs four symmetric convolutional layers to reconstruct the transmitted image from the received channel symbols. Parametric Rectified Linear Unit (PReLU) activation functions are used after each convolutional layer, while a sigmoid activation function is applied to the final decoder layer to generate normalized pixel values. The network parameters are optimized using the Adam optimizer with a learning rate of 0.0001.

Training Configuration

The proposed model was implemented using TensorFlow and trained on the CIFAR-10 dataset. The Adam optimization algorithm was employed with a learning rate of 0.0001 and a batch size of 256. The model was trained until convergence using multiple SNR values (1, 4, 7, 13, and 19 dB) under bandwidth ratios of 1/6 and 1/12. During training, the encoder and decoder networks were jointly optimized using the mean squared error (MSE) reconstruction loss.

Dataset Description

The original CIFAR-10 dataset contains RGB images with a resolution of 32 × 32 × 3 pixels. For compatibility with the proposed CNN implementation and training framework, the images were resized to 256 × 256 × 3 during the preprocessing stage before being used as network inputs. This preprocessing step does not alter the dataset itself but standardizes the input dimensions for model training.

Table 2. Training and testing algorithm table

Input original image x and the SNR μ
Transmitter Perform the encoder process $x \in R^n$
Perform the channel encoder process $Z=f\theta(x, \mu) \in C^k$
Receiver Perform channel decoding $\bar{x}$ Perform decoding function using deep joint source-channel decoder=$g\Phi(\hat{z}, \mu) = {}^{g}\Phi\ (\eta(f\theta(x,\mu)), \mu)$ $\bar{x} \quad - \quad \bar{x}$ The distortion between the x and the image is expressed as $d(x, \ ) = \frac{1}{n} \sum \cdot \ ,(x_i - \ i)^2$ $i=1$

Step 2: Evaluation

Length of dataset	**781**
Dataset image shape	256, 256, 3

Step 2.1. Test the trained model: Evaluate the trained model on the test dataset.Cifa10 dataset classification

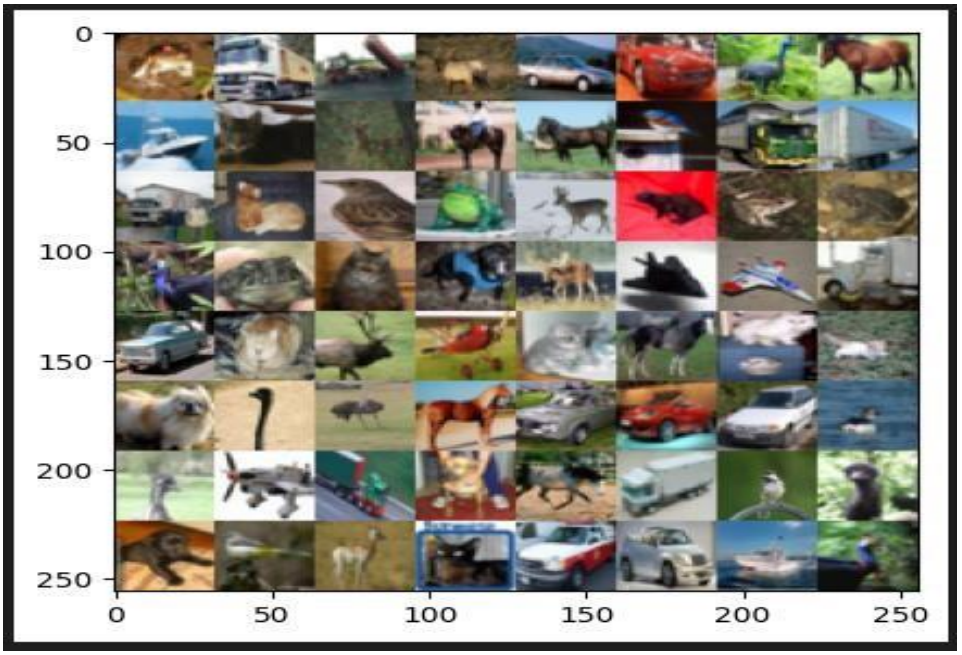

Fig.8. Simple CNN model classifier

Since each layer can be specified using a model in only one line of code, the Keras Library makes the process of generating models relatively intuitive. Add a feature. After carrying out the function indicated in Table 4.1.2, the code utilized for the entire process is self-explanatory and describes the CNN model employing four two-dimensional layers. First, a sequential function is used to establish the model. This enables the construction of a linear

stack of layers, which are viewed as a stack of objects, with each layer passing data to the one below it. Following the initialization of 32 convolution filters of 3x3 size by the first two Conv2D(32, (3, 3)‖ scripts, two more Conv2D(64, (3, 3)‖ scripts employ an increased number of filters.

Table 3. CNN autoencoder model

```
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 32, 32, 3)         0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 32, 32, 64)        1792
_________________________________________________________________
batch_normalization_3 (Batch (None, 32, 32, 64)        256
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 16, 16, 64)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 16, 16, 16)        9232
_________________________________________________________________
batch_normalization_4 (Batch (None, 16, 16, 16)        64
```

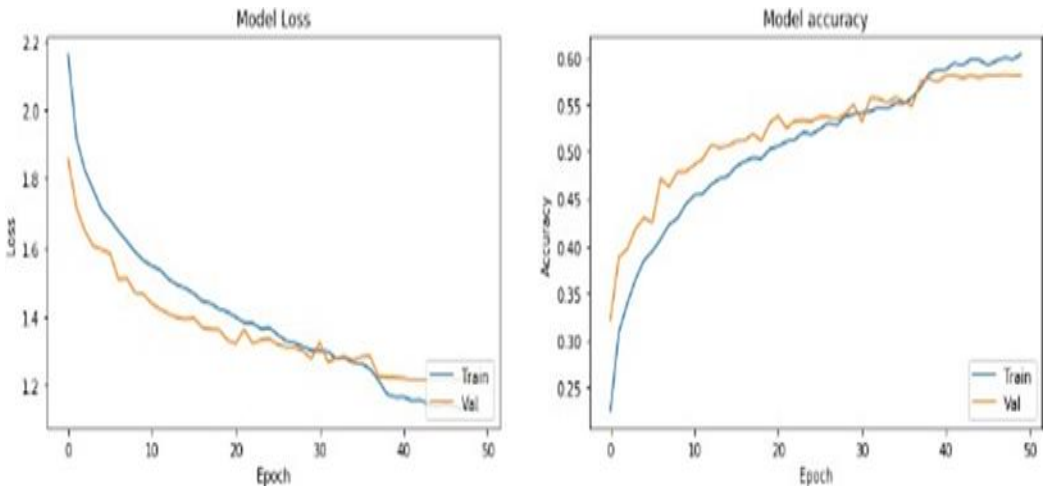

Fig.9. model accuracy and loss result

```
showOrigAndReconstructedImages(X_test, decoded_test_imgs_conv_ae, 32, 32, 3)
```

Fig.10. original and reconstructed image

Table 4. Classification report

Classification Report:

	precision	recall	f1-score	support
Airplane	0.64	0.63	0.63	1000
Automobile	0.68	0.68	0.68	1000
Bird	0.53	0.39	0.45	1000
Cat	0.39	0.44	0.41	1000
Deer	0.53	0.49	0.51	1000
Dog	0.50	0.43	0.46	1000
Frog	0.58	0.69	0.63	1000
Horse	0.62	0.66	0.64	1000
Ship	0.68	0.71	0.69	1000
Truck	0.61	0.66	0.64	1000
accuracy			0.58	10000
macro avg	0.58	0.58	0.57	10000
weighted avg	0.58	0.58	0.57	10000

Step 3: Measure performance: Calculate PSNR, SNR, and MSE to assess image quality and reliability.

Performance Evaluation

Peak SNR, or available channel uses per image pixel, for the transmission of images from the CIFAR10 dataset over an AWGN channel with a contraction rate of 1/3. The effectiveness of the DeepJSCC architecture using LDPC channel coding after BPG codec concatenation. By assuming a capacity-achieving channel code, we also achieve the performance reached by the BPG codec. Observe that this corresponds to the optimal performance that can be obtained by attaining capacity and by a separation-based strategy that combines BPG compression with any channel coding scheme, including ones that take advantage of the channel output feedback. This is because feedback does not enhance the capacity. We find that a capacity-achieving channel code followed by BPG outperforms DeepJSCC; however, when channel output feedback is used. LDPC used the IEEE 802.11 WiFi standard parameters in our studies.

Table 5. LDPC experiment parameters

LDPC code block length	648
Submatrix size	27
Decoding algorithm	Belief propagation
Iteration	10

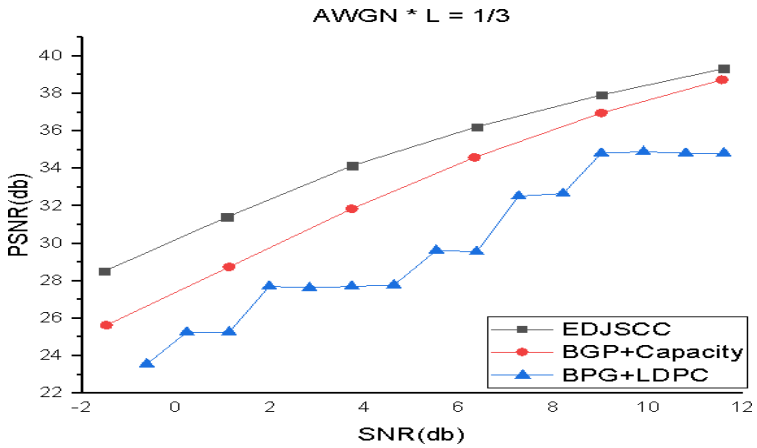

Fig.11. performance of different SNR values

CIFAR-10 Performance Evaluation: There are 10,000 test images and 50,000 training images in the CIFAR-10 dataset for training and testing. All of the image sizes are 32 x 32. Using the Adam optimizer, we trained the models with SNRs of 1, 4, 7, 13, and 19 dB, respectively, under bandwidth compression ratios of 1/6 and 1/12 for the model. We set the batch size to 128 and the learning rate to 10−4. The picture displays the PSNR performance of Deep-JSCC (EDJSCC) for R = 1 6 and R = 1 12. The dashed line represents baselines, whereas the solid line represents EDJSCC. The outcomes of training on SNRs with intervals ranging from 0 to 12 are displayed by the red solid line. The following is a summary of these findings.

Table 6. Comparison of SNRtest and PSNR

Dataset	**Cifar10**
Bandwidth Ratio	1 / 6
Channel	Relay
Plot SNR	Dataset, Ratio, Channel, Alpha=0.5

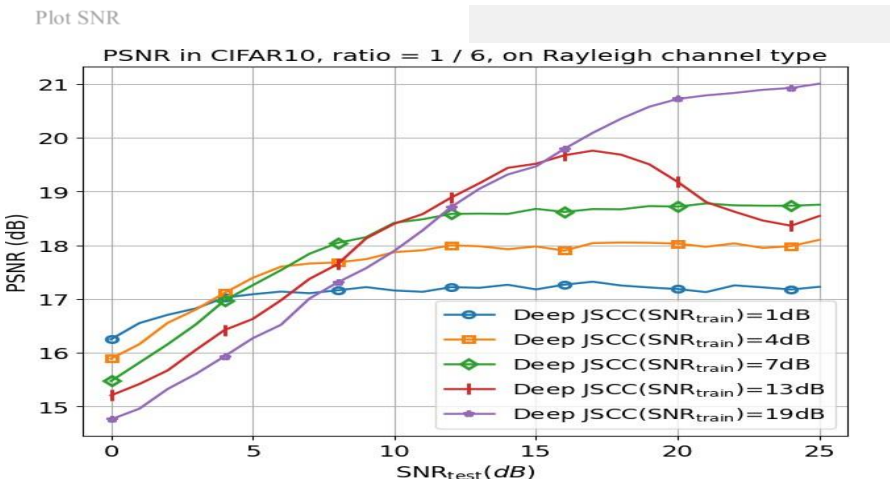

Fig.12. Performance comparisons are conducted across (SNRs) over a Rayleigh channel. In scenario (a), bandwidth ratio 1/12, (b), 1/6.

Table 7. Comparison of SNRtest and PSNR

Dataset	**Cifar10**
Bandwidth Ratio	1 / 6
Channel	AWGN
Plot SNR	Dataset, Ratio, And Channel

Table 8. Ccomparison of SNRtest AND PSNR

Dataset	**Cifar10**
Bandwidth Ratio	1 / 12
Channel	AWGN
Plot SNR	Dataset, Ratio, And Channel

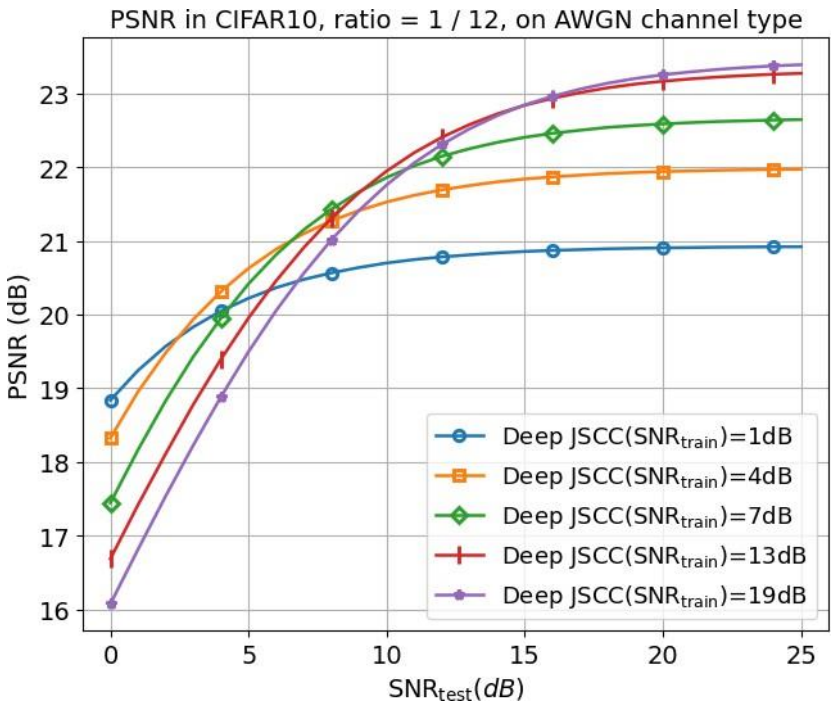

Fig.13. Performance comparisons of (SNRs) AWGN) channel. In scenario (a), the bandwidth ratio 1/12, (b), 1/6.

Figure 13 shows CIFAR10's performance. Every EDeep-JSCC curve is trained at a particular SNR. Using 0–25 intervals of 5 SNR, the EDeepJSCC curves were trained at 1 dB, 4 dB, 7 dB, 13 dB, and 19 dB. (a) EDeepJSCC reconstruction distortion on Cifar10 R = 1/6. (b) EDeepJSCC reconstruction distortion on Cifar10, R = 1/12

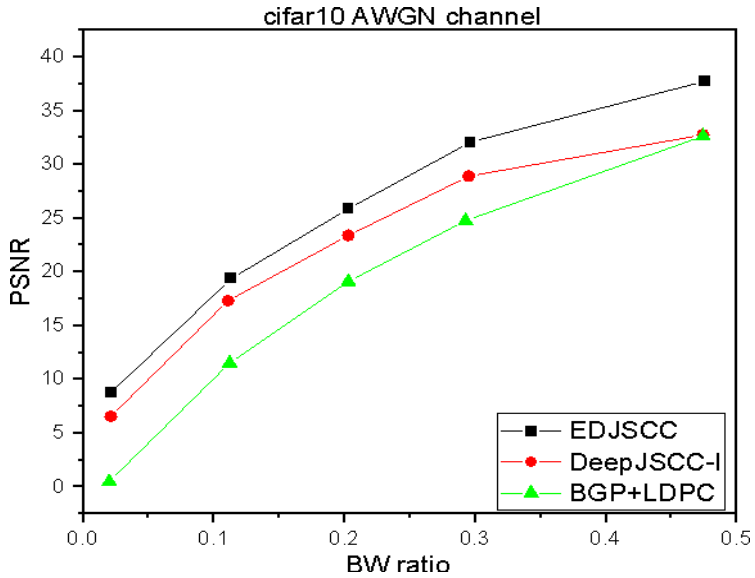

Fig.14. Comparison of deepJSCC and EDJSCC

Edeep JSCC, when the SNR testing is lower than the SNR training, achieves graceful degradation. Figure 14 shows that the separation-based BPG + LDPC transmission scheme suffers a significant performance drop, known as the cliff effect.

Step 4. Test the trained model: Evaluate the trained model on the test dataset.

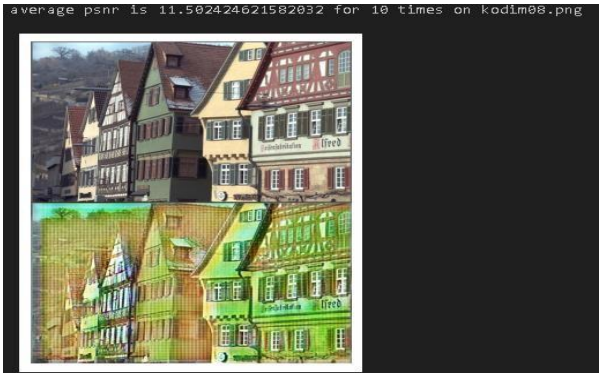

Fig.15. Test the trained model

Fig.16. SNR test and PSNR value

5. Discussion

The experimental results demonstrate that the proposed Enhanced Deep Joint Source–Channel Coding (EDeepJSCC) framework effectively improves wireless image transmission under noisy channel conditions. The proposed model consistently achieves higher Peak Signal-to-Noise Ratio (PSNR) values while maintaining lower Mean Squared Error (MSE) compared with conventional separation-based communication systems employing JPEG/BPG compression and Low-Density Parity-Check (LDPC) channel coding. These improvements indicate that jointly optimizing source and channel coding enables the network to learn more robust feature representations for reliable image reconstruction. Furthermore, the proposed EDeepJSCC framework exhibits graceful degradation as the signal-to-noise ratio decreases. Unlike conventional communication systems that suffer from the cliff effect, the proposed framework maintains stable reconstruction performance under varying channel conditions. This characteristic makes the proposed approach particularly suitable for bandwidth-constrained wireless communication environments where channel quality changes dynamically. The experimental findings are consistent with previous studies on DeepJSCC reported by Kurka and Gündüz, which demonstrated the superiority of end-to-end learning-based communication systems over conventional separation-based methods under low-SNR conditions. However, the proposed EDeepJSCC framework further enhances transmission robustness by improving reconstruction quality across multiple SNR levels while maintaining efficient bandwidth utilization. Although the proposed framework demonstrates promising performance, its evaluation is currently limited to the CIFAR-10 dataset and simulated AWGN and Rayleigh fading channels. The reconstructed images show noise removed at 23.4dB. Future work should investigate larger and more complex image datasets, additional wireless channel models, and real-time hardware implementation to further validate the effectiveness of the proposed approach.

6. Conclusion

A convolutional neural network (CNN) or a variational autoencoder is used to encode the image). This network is trained to map the input image directly to a representation suitable for transmission over the channel. The training loop incorporates the AWGN channel model. By reducing the reconstruction error, the encoder and decoder networks are tuned to function well in the simulated channel conditions. The models are trained for 19,043 steps with a batch size of 256 using Adam optimizers with $\beta1 = 0.9$ and $\beta2 = 0.999$. The learning rate is set to 0.0001. These settings are critical for achieving good performance as they provide more stable and accurate estimates of the decoder, smoothing and speeding up convergence. The

bias-free denoisers for CIFAR-10 are trained using codewords of the CIFAR-10 training data. Controlling the intricacy of the decoding procedure and guaranteeing the effective execution of matrix operations. Using a block length of 648 and a submatrix size of 27, we employed the IEEE 802.11 WiFi standard in all of our LDPC studies on the separate source-channel coding method. Using ten iterations of the belief propagation technique, the LDPC codes are decoded. Before entering the network, every pixel in the image is converted from its original integer values [0, 255] to floating-point values [0, 1]. We directly assess PSNR and SNR during testing using the decoder's output in the [0, 1] range.

7. Recommendations

Based on the findings of this study, several recommendations can be made for future research. First, the proposed EDeepJSCC framework should be evaluated using larger and higher-resolution image datasets, such as ImageNet and Kodak, to assess its scalability and generalization capability. Second, future studies should investigate the performance of the proposed framework under more realistic wireless communication scenarios, including fast fading channels, multiple-input multiple-output (MIMO) systems, and 5G/6G communication networks. Third, integrating semantic communication and transformer-based architectures may further improve transmission efficiency and image reconstruction quality. Finally, implementing the proposed framework on real-time hardware platforms would provide valuable insights into its practical deployment in next-generation wireless communication systems.

Limitations of the study and implications for further research

Although the proposed EDeepJSCC framework demonstrates promising performance for wireless image transmission, several limitations should be acknowledged. The experimental evaluation was conducted using only the CIFAR-10 dataset and simulated AWGN and Rayleigh fading channels, which may not fully represent practical communication environments. In addition, the proposed framework has not been validated using real-time hardware implementation or large-scale image datasets. Furthermore, computational complexity and inference latency were not investigated in this study. These limitations provide opportunities for future research to improve the applicability and robustness of the proposed framework.

Declaration of competing interest:

The authors declare no conflict of interest.

Disclosure of Artificial Intelligence (AI) Usage:

We used AI for paraphrasing and editing grammatical issues.

Funding Statement:

This research received no external funding.

References

Ballé, J., Minnen, D., Singh, S., Hwang, S. J., & Johnston, N. (2018). Variational image compression with a scale hyperprior. *International Conference on Learning Representations (ICLR).* https://doi.org/10.48550/arXiv.1802.01436

Bi, H., & Liang, Y.-C. (2017). Adaptive transmission of progressive JPEG images over two-way multi-relay channels. *IEEE Transactions on Wireless Communications, 16*(11), 7463–7476. https://doi.org/10.1109/TWC.2017.2737596

Bourtsoulatze, E., Kurka, D. B., & Gündüz, D. (2019). Deep joint source-channel coding for wireless image transmission. *IEEE Transactions on Cognitive Communications and Networking, 5*(3), 567–579. https://doi.org/10.1109/TCCN.2019.2919300

Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2019). Deep convolutional autoencoder-based lossy image compression. *IEEE International Conference on Image Processing (ICIP)*, 304–308. https://doi.org/10.1109/ICIP.2019.8803508

Ding, Z., Li, Y., Ma, X., & Fan, P. (2021). Feedback enabled deep joint source-channel coding for wireless image transmission. *IEEE Transactions on Wireless Communications, 20*(11), 7378–7392. https://doi.org/10.1109/TWC.2021.3083015

Jankowski, M., Gündüz, D., & Mikolajczyk, K. (2021). Wireless image retrieval at the edge. *IEEE Journal on Selected Areas in Communications, 39*(1), 89–100. https://doi.org/10.1109/JSAC.2020.3036955

Jiang, Y., Kim, H., Asnani, H., Kannan, S., Oh, S., & Viswanath, P. (2019). Turbo autoencoder: Deep learning based channel codes for point-to-point communication channels. *Advances in Neural Information Processing Systems, 32.*

Johnston, N., Vincent, D., Minnen, D., Covell, M., Singh, S., Chinen, T., Jin Hwang, S., Shor, J., & Toderici, G. (2018). Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4385–4393. https://doi.org/10.1109/CVPR.2018.00461

Kurka, D. B., & Gündüz, D. (2019). Successive refinement of images with deep joint source-channel coding. *2019 IEEE International Symposium on Information Theory (ISIT)*, 2059–2063. https://doi.org/10.1109/ISIT.2019.8849580

Kurka, D. B., & Gündüz, D. (2020). DeepJSCC-f: Deep joint source-channel coding of images with feedback. *IEEE Journal on Selected Areas in Communications, 38*(11), 2590–2600. https://doi.org/10.1109/JSAC.2020.3000815

Lee, C., Hu, X., & Kim, H.-S. (2023). Deep joint source-channel coding with iterative source error correction. *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS 2023), 206*, 3879–3902.

Michelucci, U. (2022). An introduction to autoencoders. *arXiv.* https://doi.org/10.48550/arXiv.2201.03898

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Tung, T.-Y., & Gündüz, D. (2022). DeepWiVe: Deep-learning-aided wireless video transmission. *IEEE Journal on Selected Areas in Communications, 40*(9), 2570–2583. https://doi.org/10.1109/JSAC.2022.3191354

Yang, M., Bian, C., & Kim, H.-S. (2022). OFDM-guided deep joint source channel coding for wireless multipath fading channels. *IEEE Transactions on Cognitive Communications and Networking, 8*(2), 584–599. https://doi.org/10.1109/TCCN.2022.3151935

Yang, M., & Kim, H.-S. (2022). Deep joint source-channel coding for wireless image transmission with adaptive rate control. In *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5193–5197). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746675

Zhang, Z., Yang, Q., He, S., Sun, M., & Chen, J. (2022). Wireless transmission of images with the assistance of multi-level semantic information. *IEEE Wireless Communications Letters, 11*(12), 2535–2539. https://doi.org/10.1109/LWC.2022.3196371