electronics

MDPI

Article

Stable Diffusion-Driven Semantic Coding Method for Image Transmission Under Low SNR Conditions

Sili Liu, Rong Lv, Zhixi Yang, Junxiang Qin and Yonggang Zhu *

The Sixty-Third Research Institute, National University of Defense Technology, Nanjing 210007, China; lsl2013@nudt.edu.cn (S.L.); lvrong17@nudt.edu.cn (R.L.); nudtyzx@163.com (Z.Y.); qinjx@163.com (J.Q.)
* Correspondence: zhuyonggang@nudt.edu.cn

Abstract

With the advancement of wireless communication technologies, especially the emergence of mobile communication technologies such as satellite internet and sensor networks, the rapid proliferation of communication facilities has given rise to challenges such as the scarcity of spectrum bandwidth resources, heightened channel interference, and increased noise. Consequently, traditional image source coding technologies urgently require further improvements in their compression ratio and anti-interference capability. Targeting image transmission scenarios characterized by low signal-to-noise ratios and constrained channel bandwidths, this paper proposes an image semantic coding method based on the pre-trained Stable Diffusion model, producing a zero-shot universal image compressor. This compressor leverages the denoising network of the Stable Diffusion model, with feedback from channel SNR, to further enhance the adaptability of transmitted data to channel interference. Additionally, by designing quantization and entropy coding methods for feature tensors in the semantic space, the compression ratio of the image coding process is further improved. Simulation results demonstrate that the proposed method not only achieves superior compression performance but also ensures relatively high similarity between the decoded reconstructed image and the original. Notably, it delivers a significant improvement in the perceptual similarity of human visual quality. Furthermore, the method can adapt to Gaussian noise channels, Rician fading channels, and Rayleigh fading channels with low SNR, exhibiting broad application prospects in the field of wireless communication coding methods, where the electromagnetic environment is growing increasingly complex.

Keywords: wireless semantic communication; low SNR; stable diffusion model; image coding method

Check for updates

Academic Editor: Athanasios D. Panagopoulos

Received: 20 March 2026
Revised: 10 May 2026
Accepted: 23 May 2026
Published: 4 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In recent years, satellite communication networks have witnessed rapid development, emerging as a vital component of civil communication infrastructure and playing an indispensable role in wide-area communications under complex electromagnetic environments [1]. However, due to their long transmission distance, current satellite communication bandwidths are constrained, especially for newly developed satellite internet services, which struggle to meet users' real-time transmission requirements, among which image transmission demands are the most urgent. Traditional image coding algorithms like JPEG have limited compression capability for data volume under certain distortion constraints. On the other hand, due to the open nature of wireless channels and the gradual scarcity of spectrum resources, interference and noise in transmission channels can

no longer be ignored, and malicious interference in adversarial channels is unavoidable. Under harsh channels with low signal-to-noise ratio (SNR), traditional coding methods exhibit a "cliff effect" [2], where image quality deteriorates sharply after decompression. Furthermore, channel feedback enables adaptive coding and related strategies to enhance channel transmission efficiency and achievable rate [3]. With the development of artificial intelligence technology, semantic communication technology based on deep neural networks provides a new approach. Traditional wireless anti-interference technologies rely on Shannon's formula, sacrificing communication bandwidth as a cost, while semantic communication converts images into semantic information via neural networks [4]. This semantic information contains key image details of human concern; further coding of such information removes redundant and non-critical semantic elements, thus compressing the bit data volume for transmission, reducing bandwidth requirements, and potentially improving interference tolerance beyond the limits of Shannon's formula.

Based on the above requirements, there is a need to explore a coding technology that meets distortion constraints and exhibits better compression performance under low SNR conditions. With the development of deep learning algorithms and large models, semantic communication has become a research hotspot. Shi Guangming et al. [5] proposed a new semantic communication approach from the perspective of intelligent perception and discussed a semantic coding mechanism. Kalfa et al. [6] introduced a semantic signal-processing framework adaptable to different communication tasks at the receiver. Niu Kai et al. [7] and Zhang Ping et al. [4] explored the measurement of semantic information, proposed an intelligent and efficient semantic communication system architecture, and established a mathematical theory of semantic communication based on synonymous mapping [8]. For image source transmission, Bourtsoulatze et al. [9] proposed Joint Source-Channel Coding based on a convolutional neural network to perform image transmission over wireless channels, optimizing semantic codecs to enhance transmission performance. For image retrieval tasks, Jankowski et al. [9] developed an edge–cloud collaborative semantic communication method, significantly improving task performance. The emergence of large generative AI models has provided new means for developing ultra-low-rate high-fidelity semantic communication systems. Visual generative AI models such as Sora [10], Lumiere [11], and DALL·E [12], pre-trained on massive data, have acquired cognitive foundations about image distributions and can generate high-quality images from text prompts. Current generative models, such as GPT-4, are increasingly applied to image compression, achieving significantly higher compression ratios than traditional algorithms [13].

Current semantic coding approaches still face several critical limitations. Firstly, small models trained on specific datasets lack sufficient generalization capability, performing well only on their designated test sets and proving inadequate as universal compressors. Secondly, semantic compression based on large language models like GPT-4 presents inherent constraints in performance evaluation; typically measured using entire datasets, these methods utilize encoders that output fixed-dimensional vectors. Consequently, adjustments are possible only in input data volume, not the dimensionality of the encoded tensor representation. This renders compression that is solely reliant on large language models impractical at low data volumes [14]. Thirdly, research predominantly focuses on wide-bandwidth scenarios such as 5G mobile communications, largely neglecting low SNR scenarios characterized by significant channel interference.

The main contributions of this paper are outlined as follows.

First, a zero-shot image semantic coding framework is constructed based on the pre-trained Stable Diffusion model (SD_Semantic). The proposed architecture supports general image compression without dataset-specific fine-tuning, overcoming the poor generalization of conventional lightweight deep learning-based coding models.

Second, a channel-aware adaptive semantic optimization mechanism is proposed. By embedding SNR feedback into the denoising network of Stable Diffusion, the developed scheme adapts to adversarial channels with low SNR and strong malicious interference, mitigates the cliff effect of traditional coding, and enhances the transmission's robustness against interference.

Third, latent-space entropy coding is tailored for semantic feature tensors in diffusion models. It further compresses redundant bit information and achieves high-fidelity image transmission at high compression ratios under low-SNR channel conditions.

Fourth, a multi-dimensional evaluation benchmark is established to compare the proposed scheme with traditional image coding methods and existing semantic coding methods. Experimental results in terms of PSNR, SSIM, LPIPS, bit rate and channel robustness demonstrate that the proposed method achieves superior overall performance over traditional coding schemes in low-SNR scenarios.

Compared with conventional coding and decoding methods, traditional schemes only eliminate redundancy in the pixel domain without exploiting high-level image semantic priors and lack channel adaptation capability. They suffer from sharp image quality degradation under low-SNR channels with malicious interference. By leveraging the powerful visual semantic prior and generative modeling capability of Stable Diffusion, the proposed method removes redundancy at the semantic level. Meanwhile, the introduced channel state awareness mechanism enables the scheme to adapt to complex electromagnetic and interference channels, breaking through the performance bottleneck of traditional coding approaches. Existing CNN-based joint source-channel coding and lightweight semantic models are trained on task-specific datasets, leading to limited generalization and poor universality. Most of them are designed for ideal broadband communication scenarios while ignoring low-SNR conditions and malicious interference. In contrast, this paper adopts a zero-shot architecture based on a pre-trained large generative model, which does not rely on scenario-specific datasets. It inherently adapts to harsh channel environments, thereby possessing wider applicability and stronger robustness.

2. Pre-Trained Model-Based Semantic Coding Methods

This thesis proposes an image semantic coding method, SD_Semantic, which is based on pre-trained generative diffusion models. The following sections will elaborate on the semantic communication model, image semantic coding workflow, and neural network architecture.

2.1. Semantic Communication Model

Source coding is a process of data compression aimed at eliminating redundancy from the source data as much as possible, whereas channel coding introduces redundant information to enhance transmission reliability. The design of both source and channel coding often requires joint consideration. Figure 1 illustrates a semantic communication architecture based on pre-trained generative diffusion models for image coding. The input source is denoted by $x \in \mathrm{R}^m$. The semantic spatial feature tensor z obtained by the semantic extraction network is derived through the semantic extraction function $z \in f_{\mathrm{se}}(x; \varphi_{\mathrm{se}})$, where φ_{se} denotes the learnable parameters of the semantic extraction network. The tensor z is mapped to the channel input symbol vector y via quantization and the function $y \in f_{\mathrm{e}}(z; \varphi_{\mathrm{e}})$, where φ_{e} represents the parameters for quantization and entropy coding. The symbol vector y passes through the wireless channel and arrives at the receiver, with the wireless channel modeled as $\hat{y} \in H(y, \nu)$, where ν denotes the wireless channel parameters. In this paper, we mainly consider the impacts of channel noise types and signal-to-noise ratio (SNR) on the transmitted symbols. At the receiver, an estimate $\hat{z} = g_{\mathrm{d}}(\hat{y}; \varphi_{\mathrm{d}})$ of

the semantic tensor z is obtained through decoding. The estimate $\hat{x} = g_{\mathrm{sd}}(\hat{z}; \varphi_{\mathrm{sd}})$ of the original source is recovered via semantic reconstruction, where φ_{d} denotes the parameters for entropy decoding, and φ_{sd} represents the parameters of the semantic decoding and reconstruction network.

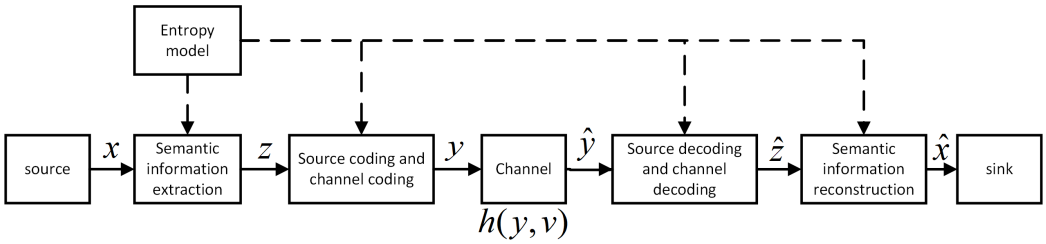


Figure 1. Semantic encoding and decoding communication workflow.

After designing the semantic extraction network architecture, quantization, and entropy coding methods, this study optimizes the parameters φ_e and φ_d to minimize the distortion between the source image and the reconstructed image at the receiver, as formalized in Equation (1).

$$(\varphi_{\mathrm{se}}{}^*, \varphi_{\mathrm{e}}{}^*, \varphi_{\mathrm{d}}{}^*, \varphi_{\mathrm{sd}}{}^*) = \arg \min_{\varphi_{\mathrm{se}}, \varphi_{\mathrm{e}}, \varphi_{\mathrm{d}}, \varphi_{\mathrm{d}}} \mathrm{E}_{x \sim p_x} \mathrm{E}_{\hat{x} \sim p_{\hat{x}|x}} [d(x, \hat{x})] \tag{1}$$

Here, $\varphi_e{}^*$ and $\varphi_d{}^*$ denote the optimal trainable parameters of the encoding and decoding neural networks, while $d(x, \hat{x})$ represents the distortion between the source data and the decoded reconstruction after transmission. In the field of image processing, the similarity between two images is typically measured by the minimum mean squared error (MSE), the peak signal-to-noise ratio (PSNR), and the structural similarity index (SSIM). This paper adopts PSNR, SSIM, and the Learned Perceptual Image Patch Similarity (LPIPS) as distortion metrics, where LPIPS is a perceptual similarity metric based on deep feature representations. PSNR is one of the metrics for measuring image distortion [15]. Its calculation is based on the definition of MSE, as formalized in Equations (2) and (3), given a clean image I of size $m \times n$ and its noisy version K.

$$\mathrm{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \tag{2}$$

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}} \right) \tag{3}$$

Here, MAX_I denotes the maximum possible pixel value of the image. If each pixel is quantized with 8 bits, then MAX_I is 255. This paper denotes the bit depth of pixel quantization as B, then $\mathrm{MAX}_I = 2^B - 1$. Typically, for uint8 data, the maximum pixel value is 255, whereas for floating-point data, it is 1. The SSIM serves as a metric for quantifying structural similarity between two images [15]. It computes the mean luminance values of both images and utilizes these means as comparative parameters for luminance alignment, with the calculation method formalized in Equation (4).

$$\mathrm{SSIM} = \frac{(2\mu_x \mu_y + c_1)(2\sigma_x \sigma_y + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4}$$

Here, μ_x and μ_y denote the mean values of the luminance components for image x and image y, respectively, σ_x and σ_y represent the variances of image x and image y, while c_1 and c_2 are constants introduced for numerical stability.

2.2. Semantic Encoding and Decoding Workflow

The source data in this study comprises image tensors denoted as $H * W * C$, where H, W, and C represent the height, width, and channel dimensions, respectively. This thesis proposes a semantic image coding method based on pre-trained generative diffusion models for low SNR scenarios. As illustrated in Figure 2, the communication pipeline initiates with natural-scene images as the source signal; these undergo semantic feature extraction to transform and compress raw pixel data into latent-space semantic feature tensors, significantly reducing bitrate requirements. Subsequent quantization of these tensors further enhances compression efficiency, followed by entropy coding and modulation for channel transmission. After demodulation, the received bitstream is decoded and dequantized to reconstruct semantic feature tensors, which are then processed by an auto-decoder to recover image content. Simultaneously, Deepseek-VL generates textual descriptions of the source images, which are transmitted to the receiver to guide the diffusion model's reconstruction via semantic alignment. Notably, the Low-Density Parity-Check Code (LDPC) coding is adaptively activated based on real-time channel SNR measurements to optimize error correction without necessarily compromising throughput, wherein both its deployment status and code rate are dynamically configurable.

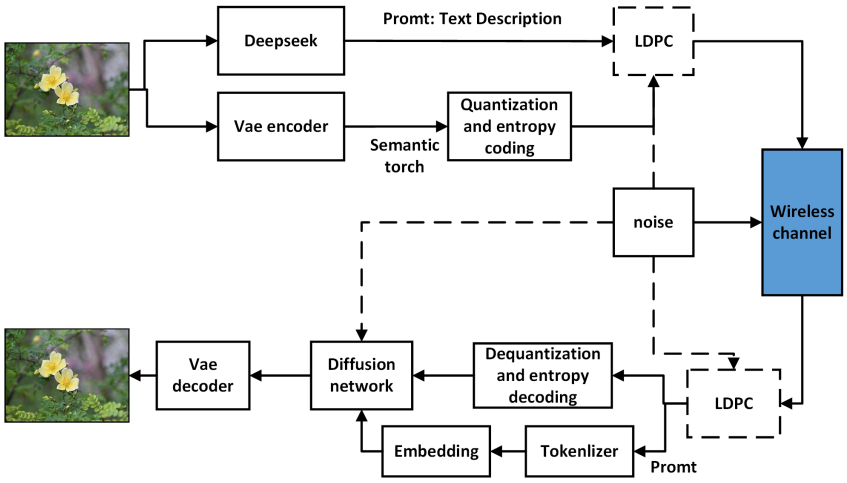


Figure 2. Semantic coding workflow based on stable pre-trained diffusion models.

2.3. Semantic Coding Neural Network Model

The image semantic extraction and reconstruction modules employ the Variational Autoencoder (VAE) from the stable diffusion model. VAE utilizes pre-trained models to transform an input image into a semantic latent space tensor via a deep neural network. As illustrated in Figure 3, the VAE encoder decomposes the original image into $n * n$ independent Gaussian distributions, sampling from which yields an $n * n$ feature tensor encapsulating the semantic characteristics of the source image [10]. The VAE encoder and decoder share a symmetric network architecture, trained jointly during optimization.

The following introduces the principles of the VAE network [16]. First, we assume there is a dataset $X = \{x^{(i)}\}_{i=1}^{N}$, where samples are independent and identically distributed (i.i.d). We assume that each sample in this dataset is generated from a stochastic process, which is described as follows: First, a semantic space tensor $z^{(i)}$ is sampled from the semantic space variable distribution $p_\theta(z)$. We assume that the semantic space distribution $p_\theta(z)$ here is a continuous distribution. Then, based on the latent space tensor $z^{(i)}$, a data sample $x^{(i)}$ is generated, which follows the conditional distribution $p_\theta(x|z = z^{(i)})$.

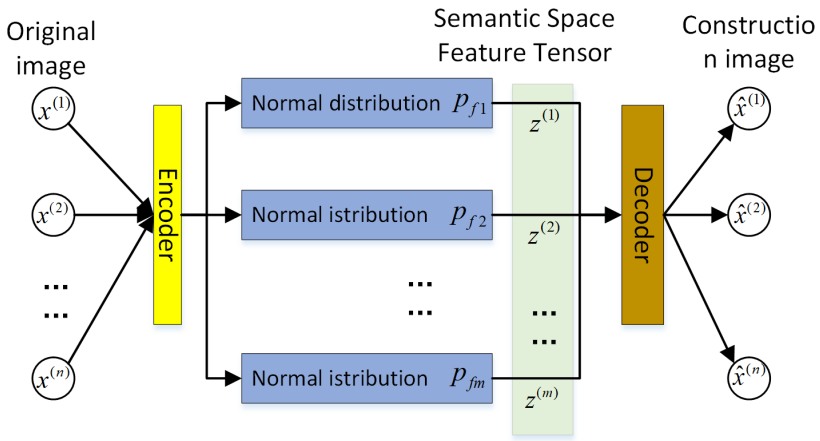


Figure 3. Schematic architecture of VAE.

According to the above assumptions, the probability of each sample in the dataset can be calculated as $p(x^{(i)}) = \int p_\theta(x^{(i)}|z)p_\theta(z)\mathrm{d}z$. If each term in the formula has an analytical expression, we can solve the model parameters θ through maximum likelihood estimation, and the objective function of maximum likelihood is shown in Equation (5).

$$\begin{aligned}\theta^* &= \arg\max_\theta \sum_{i=1}^{N} \log p_\theta(x^{(i)}) \\ &= \arg\max_\theta \sum_{i=1}^{N} \log \int p_\theta(x^{(i)}|z)p_\theta(z)\mathrm{d}z\end{aligned} \tag{5}$$

However, since $p(x^{(i)}) = \int p_\theta(x^{(i)}|z)p_\theta(z)\mathrm{d}z$ is non-computable, the maximum likelihood function cannot be used to solve for θ. VAE introduces a distribution $q_\phi(z|x)$, which serves as an approximation of the true distribution $p_\theta(z|x)$. This z represents the latent space distribution given the sample x, and we employ machine learning via the VAE network to solve for the parameters ϕ. In VAE, $q_\phi(z|x)$ is the encoder network that maps samples to latent space tensors, while $p_\theta(x|z)$ is the decoder network that maps latent space tensors back to sample data. Since the distribution $q_\phi(z|x)$ is an approximation of the true distribution $p_\theta(z|x)$, the KL divergence of this distribution with respect to the true distribution on the sample $x^{(i)}$ is as expressed in Equation (6).

$$\begin{aligned}&D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z|x^{(i)})) \\ &= \int q_\phi(z|x^{(i)}) \log \frac{q_\phi(z|x^{(i)})}{p_\theta(z|x^{(i)})}\mathrm{d}z \\ &= \int q_\phi(z|x^{(i)}) \log \frac{q_\phi(z|x^{(i)})}{p_\theta(z,x^{(i)})}\mathrm{d}z + \log p_\theta(x^{(i)})\end{aligned} \tag{6}$$

After transposing terms, we can obtain Equation (7).

$$\begin{aligned}\log p_\theta(x^{(i)}) &= D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z|x^{(i)})) \\ &- \int q_\phi(z|x^{(i)}) \log \frac{q_\phi(z|x^{(i)})}{p_\theta(z|x^{(i)})}\mathrm{d}z\end{aligned} \tag{7}$$

The second term is denoted as $\eta(\theta, \phi, x^{(i)})$, as shown in Equation (8).

$$\begin{aligned}&\eta(\theta, \phi, x^{(i)}) = \int q_\phi(z|x^{(i)}) \log \frac{p_\theta(z|x^{(i)})}{q_\phi(z|x^{(i)})}\mathrm{d}z \\ &= \mathbb{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(z, x^{(i)}) - \log q_\phi(z|x^{(i)})]\end{aligned} \tag{8}$$

Equation (7) can be rewritten as Equation (9).

$$\begin{aligned} &\log p_\theta(x^{(i)}) \\ &= D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z|x^{(i)})) + \eta(\theta, \phi, x^{(i)}) \end{aligned} \tag{9}$$

Since the KL divergence is non-negative, $\eta(\theta, \phi, x^{(i)})$ serves as a lower bound for $\log p_\theta(x^{(i)})$. Therefore, the maximum likelihood objective $\sum_{i=1}^{N} \log p_\theta(x^{(i)})$ is transformed into maximizing $\sum_{i=1}^{N} \eta(\theta, \phi, x^{(i)})$, which is the ELBO (Evidence Lower Bound). This can be further decomposed as shown in Equation (10).

$$\begin{aligned} &\eta(\theta, \phi, x^{(i)}) = \int q_\phi(z|x^{(i)}) \log \frac{p_\theta(z, x^{(i)})}{q_\phi(z|x^{(i)})} \mathrm{d}z \\ &= -D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z)) \\ &+ \mathbb{E}_{z\sim q_\phi(z|x^{(i)})} \log p_\theta(x^{(i)}|z) \end{aligned} \tag{10}$$

To solve the expectation $_{z\sim q_\phi(z|x^{(i)})} \log p_\theta(x^{(i)}|z)$ containing parameters, it is necessary to compute the gradient with respect to the parameters and update them. For convenience, we denote $f(z) = \log p_\theta(x^{(i)}|z)$. Then, we compute the gradient of $\mathbb{E}_{z\sim q_\phi(z|x^{(i)})} \log p_\theta(x^{(i)}|z)$ with respect to the parameter ϕ, as shown in Equation (11).

$$\begin{aligned} &\nabla_\phi \mathrm{E}_{z\sim q_\phi(z|x^{(i)})} f(z) \\ &= \mathrm{E}_{z\sim q_\phi(z|x^{(i)})} f(z) \nabla_\phi \log q_\phi(z|x^{(i)}) \end{aligned} \tag{11}$$

In the VAE network, a reparameterization trick is proposed. A distribution $p(\varepsilon)$ and a function $g_\phi(\varepsilon, x^{(i)})$ dependent on parameters $\phi, \varepsilon, x^{(i)}$ are constructed, satisfying Equation (12).

$$\mathbb{E}_{z\sim q_\phi(z|x^{(i)})}[f(z)] = \mathbb{E}_{\varepsilon\sim p(\varepsilon)}[f(g_\phi(\varepsilon, x^{(i)}))] \tag{12}$$

The above expectation is calculated using the Monte Carlo method to obtain Equation (13).

$$\begin{aligned} &\mathbb{E}_{z\sim q_\phi(z|x^{(i)})}[f(z)] = \mathbb{E}_{\varepsilon\sim p(\varepsilon)}[f(g_\phi(\varepsilon, x^{(i)}))] \\ &\simeq \frac{1}{L} \sum_{l=1}^{L} f(g_\phi(\varepsilon^{(l)}, x^{(i)})) \end{aligned} \tag{13}$$

where $\varepsilon^{(l)} \sim p(\varepsilon)$, substituting Equation (13) into Equation (8) gives Equation (14).

$$\begin{aligned} &\eta(\theta, \phi, x^{(i)}) \\ &= \frac{1}{L} \sum_{l=1}^{L} [\log p_\theta(z^{(i,l)}, x^{(i)}) - \log q_\phi(z^{(i,l)}|x^{(i)})] \end{aligned} \tag{14}$$

where $z_{i,l} = g_\phi(\varepsilon_l, x_i), \varepsilon_l \sim p(\varepsilon)$, substituting Equation (13) into Equation (10) gives Equation (15).

$$\begin{aligned} &\eta(\theta, \phi, x^{(i)}) = -D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z)) \\ &+ \frac{1}{L} \sum_{l=1}^{L} \log p_\theta(x^{(i)}|z^{(i,l)}) \end{aligned} \tag{15}$$

where $z^{(i,l)} = g_\phi(\varepsilon^{(l)}, x^{(i)}), \varepsilon^{(l)} \sim p(\varepsilon)$, both Equations (14) and (15) can be used to estimate $\eta(\theta, \phi, x^{(i)})$ and the gradients of parameters θ and ϕ. The difference between the two formulas is that Equation (15) requires calculating the KL divergence. When we assume that both $q_\phi(z|x^{(i)})$ and $p_\theta(z)$ are Gaussian distributions, the KL divergence can be directly computed instead of estimated, so the gradient variance calculated by the latter is smaller. The VAE network adopts the approach of calculating the KL divergence.

The noise injection network progressively adds Gaussian noise to the latent space vectors output by the VAE encoder. This process drives the semantic feature tensor toward a stochastic noise distribution. Subsequently, the denoising network predicts and removes the injected noise to recover the original semantic feature tensor. The denoising intensity parameter (denoted as λ) controls the noise magnitude: when $\lambda = 0$, no noise is added; when $\lambda = 1$, maximum noise is applied. In this study, the initial denoising intensity is set to $\lambda = 0.2$. Both the noise injection and denoising networks utilize pre-trained stable diffusion models, with their operational workflows illustrated in Figure 4 [10].

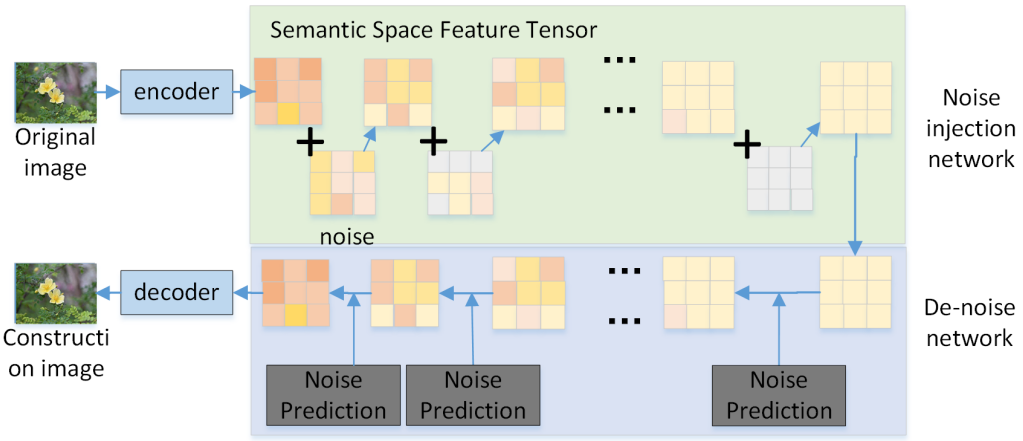


Figure 4. Schematic architecture of the noise injection network and the denoising network.

In this study, the prompt not only serves as a generic input for the Stable Diffusion model but also includes textual descriptions of the source image generated by the DeepSeek large model. Both the textual descriptions and the semantic feature vectors of the image are jointly fed as inputs for decoding and reconstruction. As channel noise increases and the SNR progressively decreases, this study linearly increases the denoising intensity from 0.04 to 0.2. Guided by the prompt, the noise injection and denoising networks adaptively perform their operations. Under extremely low SNR conditions where the semantic feature tensor is severely corrupted by noise, this thesis assumes the channel can only transmit textual information. In such cases, the text-to-image capability of the stable diffusion model is activated, directly synthesizing images solely from the prompt.

The noise addition network processes the image semantic space tensor z by progressively introducing Gaussian noise, ensuring that the distribution of the noisy data gradually converges to a Gaussian distribution associated with the input data [16]. Let the noise-free data be denoted as z_0, which in this paper refers to the feature tensor obtained by superimposing quantization noise n_q and channel noise n_c onto the semantic space feature tensor output by the autoencoder, i.e., $z_0 = z + n_q + n_c$. Here, $z_0 \sim q(z_0)$, and $q(z_0)$ represents the original noise-free data distribution, and the state transition from time $t-1$ instant to t is characterized by Equation (16).

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t} \cdot z_{t-1}, \beta_t \cdot \mathrm{I}) \tag{16}$$

where $t \in \{0, 1, \cdots, T\}$, $\mathcal{N}$ denotes a Gaussian distribution, β_t is a noise scaling factor associated with time instant t, and I is an identity matrix of the same dimension as the initial state z_0. Given the input z_0, the joint distribution of $z_1, z_2, \cdots, z_T$ can be expressed as Equation (17).

$$q(z_1, z_2, \cdots, z_T|z_0) = \prod_{t=1}^{T} q(z_t|z_{t-1}) \tag{17}$$

According to the properties of Markov processes, the state at time t given the input z_0 can be expressed as Equation (18).

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t} \cdot z_{t-1}, \beta_t \cdot \mathrm{I}) \tag{18}$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \Pi_{s=0}^{t} \alpha_s$. Based on Equation (16), the relationship between z_t and z_{t-1} is shown in Equation (19).

$$z_t = \sqrt{\alpha_t} \cdot z_{t-1} + \sqrt{1-\alpha_t} \cdot \mu_{t-1} \tag{19}$$

where $\mu_{t-1} \sim \mathcal{N}(0, \mathrm{I})$, the relationship between z_t and z_0 can be obtained by recursion as shown in Equation (20).

$$\begin{aligned} z_t &= \sqrt{\alpha_t} \cdot z_{t-1} + \sqrt{1-\alpha_t} \cdot \epsilon_{t-1} \\ &= \sqrt{\alpha_t \alpha_{t-1}} \cdot z_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}} \cdot \bar{\epsilon}_{t-2} \\ &= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}} \cdot z_{t-3} + \sqrt{1-\alpha_t \alpha_{t-1} \alpha_{t-2}} \cdot \bar{\epsilon}_{t-3} \\ &\cdots \\ &= \sqrt{\bar{\alpha}_t} \cdot z_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon \end{aligned} \tag{20}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\bar{\epsilon}_{t-2}$ is the distribution obtained by summing two Gaussian distributions. According to the properties of Gaussian noise, for two Gaussian distributions with different variances $\mathcal{N}(\mathbf{0}, \sigma_1^2 \cdot \mathbf{I})$ and $\mathcal{N}(\mathbf{0}, \sigma_2^2 \cdot \mathbf{I})$, their summed Gaussian distribution is $\mathcal{N}(\mathbf{0}, (\sigma_1^2 + \sigma_2^2) \cdot \mathbf{I})$. Therefore, Equation (20) can be rewritten as Equation (21).

$$\begin{aligned} z_t &= \sqrt{\alpha_t} \cdot z_{t-1} + \sqrt{1-\alpha_t} \cdot \epsilon_{t-1} \\ &= \sqrt{\alpha_t} \cdot \left(\sqrt{\alpha_{t-1}} \cdot z_{t-2} + \sqrt{1-\alpha_{t-1}} \cdot \epsilon_{t-2} \right) + \sqrt{1-\alpha_t} \cdot \epsilon_{t-1} \\ &= \sqrt{\alpha_t \alpha_{t-1}} \cdot z_{t-2} + \sqrt{\alpha_t (1-\alpha_{t-1})} \cdot \epsilon_{t-2} + \sqrt{1-\alpha_t} \cdot \epsilon_{t-1} \\ &= \sqrt{\alpha_t \alpha_{t-1}} \cdot z_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}} \cdot \bar{\epsilon}_{t-2} \end{aligned} \tag{21}$$

The standard deviation of the sum of two Gaussian distributions is given by Equation (22).

$$\sqrt{\alpha_t (1-\alpha_{t-1}) + (1-\alpha_t)} = \sqrt{1-\alpha_t \alpha_{t-1}} \tag{22}$$

In the noise addition network, since the noise added at each step is identically distributed Gaussian noise, the noisy state z_T at time T can be directly derived from the input z_0 When $\bar{\alpha}_T \approx 0, T$, the distribution of z_T at time T is nearly a Gaussian distribution, which can be defined as Equation (23).

$$q(z_T) := \int q(z_T \mid x_0) q(z_0) dz_0 \approx \mathcal{N}(z_T; \mathbf{0}, \mathbf{I}) \tag{23}$$

The denoising network estimates the noise distribution by learning from the existing states, further obtains the state at the previous time instant, and gradually constructs real data from the Gaussian distribution. Based on the forward diffusion results, it can be considered that the posterior distribution of the noisy state z_T at time T satisfies $p(z_t) \sim \mathcal{N}(z_t; \mathbf{0}, \mathbf{I})$, and the joint distribution $p_\theta(z_0, z_1, \ldots, z_T)$ is also a Markov chain, which is defined as Equation (24).

$$p_\theta(z_0, z_1, \ldots, z_T) := p(z_T) \prod_{t=1}^{T} p_\theta(z_{t-1} \mid z_t) \tag{24}$$

The state z_{t-1} at time $t-1$ can be obtained from the state z_t at the previous time step t, and its conditional distribution is expressed as Equation (25).

$$p_\theta(z_{t-1} \mid z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t), \Sigma_\theta(z_t, t)) \tag{25}$$

Here $\mu_\theta(z_t, t)$ and $\Sigma_\theta(z_t, t))$ denote the noise mean and variance obtained by the noise estimation network at time t, respectively, with θ being the parameters of the noise estimation network. In this case, given the input z_0, the true conditional distribution between the state z_t at time t and the previous state z_{t-1} at time $t-1$ is expressed as Equation (26).

$$q(z_{t-1} \mid z_t, z_0) = \mathcal{N}\left(z_{t-1}; \tilde{\mu}_t(z_t, z_0), \tilde{\beta}_t \cdot \mathbf{I}\right) \tag{26}$$

where the parameters of the noise posterior distribution, $\widetilde{\mu}_t$ and $\widetilde{\beta}_t$, are given by Equation (27).

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(z_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \epsilon_t\right), \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t \tag{27}$$

Here $\sum_\theta(z_t, t) = \sigma_t^2 \cdot \mathbf{I}$, that is, $\sigma_t^2 = \tilde{\beta}_t$, so the predicted posterior conditional distribution is shown in Equation (28).

$$\mu_\theta(z_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(z_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \epsilon_\theta(z_t, t)\right) \tag{28}$$

Based on the known formula, the state z_t at time t satisfies $z_t = \sqrt{\bar{\alpha}_t} \cdot z_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon$. Therefore, the optimization objective of the denoising network is to make the estimated noise distribution close to the real noise distribution, as shown in Equation (29).

$$\begin{aligned} L_{LDM} = \mathbb{E}_{z_0, t, \epsilon_t \sim \mathcal{N}(0, \mathbf{I})} \\ \left[\left\| \epsilon_t - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t} \cdot z_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon, t\right)\right\|_2^2\right] \end{aligned} \tag{29}$$

The state z_{t-1} at time $t-1$ can be expressed as Equation (30).

$$\begin{aligned} z_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{z_t - \sqrt{1-\bar{\alpha}_t} \cdot \epsilon_\theta(z_t, t)}{\sqrt{\bar{\alpha}_t}}\right) \\ + \sqrt{1-\bar{\alpha}_{t-1}} \cdot \epsilon_\theta(z_t, t) \end{aligned} \tag{30}$$

where $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the real data distribution can be gradually obtained through reverse sampling based on the noise distribution estimated by the noise estimation network at different time instants, as per Equation (30). In the image restoration task of this paper, a conditional diffusion model must be employed to generate the expected restored image. Specifically, the semantic space feature tensor with quantization errors, channel errors, and noise is used as the initial input image, and the text description of the image is introduced as a condition into the noise estimation network to estimate the conditional noise distribution. The conditional diffusion model used in this paper shares an identical forward diffusion process with the classical diffusion model. The only difference lies in whether the image text description is introduced as a prompt during the reverse sampling process [17]. The text description m is processed by an encoder τ_ω to obtain the corresponding conditional embedding tensor $\tau_\omega(m)$, which is fused with the input semantic space feature tensor z_t via cross-attention mechanism to guide image restoration, as shown in Equations (31) and (32).

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d}}) \cdot V \tag{31}$$

$$\begin{aligned} Q &= W_Q^{(i)} \cdot \varphi_i(z_t) \\ K &= W_K^{(i)} \cdot \tau_\omega(m) \\ V &= W_V^{(i)} \cdot \tau_\omega(m) \end{aligned} \tag{32}$$

Here, $\varphi_i(z_t)$ denotes the intermediate layer representation of the denoising network. Then the objective function under this control condition can be expressed as Equation (33).

$$L_{LDM} = \mathbb{E}_{z_0,t,m,\epsilon_t \sim \mathcal{N}(0,\mathbf{I})}[||\epsilon - \epsilon_\theta(z_t, t, \tau_\omega(m))||_2^2] \tag{33}$$

3. Simulation Experiment and Results Analysis

3.1. Simulation Parameter Settings

The simulation employs the Kodak benchmark dataset—a standard test set in image compression research comprising 24 RGB-format images with diverse visual styles and content. For communication channel modeling, 4-QAM modulation is implemented across Gaussian, Rayleigh, and Rician fading channels, with SNR varying from 0 to 50 dB. Detailed simulation parameters are listed in Table 1.

Table 1. Simulation parameter.

Configuration Item	Value/Specification
Dataset	Kodak
Modulation Scheme	4-QAM
Channel Type	Gaussian, Rayleigh, Rician
SNR	0–50 dB
LDPC code rate	1/2

Futhermore, we adopt 8-bit uniform quantization combined with Huffman entropy coding. The denoising strength is controlled by the number of denoising steps, which is calculated as a linear function of the SNR. A higher SNR corresponds to fewer denoising steps. In the test, we use the pre-trained weights of stable diffusion v1.4, and all network parameters are frozen during the experiment. In terms of compression ratio configuration, we set the compression ratios of comparative experiments according to that of the proposed model. This is realized by adjusting the quality factors of JPEG and Webp in the Python 3.8 code. For the Deep-JSCC scheme [2], the compression ratio is adjusted by configuring its bandwidth parameter.

3.2. Experimental Procedure and Results Analysis

3.2.1. Comparison of Source Comprssion Capabilities

To compare the source compression capabilities of the SD_Semantic method with traditional coding methods, we set a high SNR of 100 dB to benchmark the compression ratio and transmission distortion between the proposed semantic compression method SD_Semantic and traditional source-channel coding schemes. Pixel-level distortion was quantified using conventional metrics: PSNR and SSIM. For perceptual quality and semantic fidelity evaluation, the LPIPS metric was employed. LPIPS leverages pre-trained convolutional neural networks (e.g., AlexNet) to extract image features, aligning with human judgments of visual similarity [18]. Since networks like AlexNet are widely adopted in downstream vision tasks (e.g., object detection, image classification, and semantic segmentation), LPIPS effectively captures human-perceived visual quality and semantic distortion. Experimental results are summarized in Table 2. The baseline schemes for comparison in this paper include JPEG, WebP, and Deep-JSCC [2]. Specifically, WebP is a modern image format developed by Google. It is designed to deliver image quality comparable to

that of JPEG and PNG while achieving a significant reduction in file size. WebP supports both lossy and lossless compression, and provides a set of advanced features such as transparency, animation, and metadata processing. It mainly adopts intra-frame prediction combined with discrete cosine transform and entropy coding to reduce the file size while preserving image quality as much as possible. Compared with JPEG, WebP introduces an intra-frame prediction algorithm, which reduces storage data volume by predicting the color values of each pixel block. Each pixel block of size 16×16 or 4×4 can be predicted using the surrounding pixel values. Only the residuals between the actual pixel values and the predicted values are stored after prediction, thereby effectively reducing the amount of data.

Table 2. Compression ratio and image distortion at SNR = 100 dB.

Evaluation Metric	SD_Semantic	JPEG	WebP	Deep-JSCC
Compression ratio	0.004831	0.005203	0.005504	0.33
PSNR	23.588698	25.061536	27.516189	25.0750
SSIM	0.806716	0.739952	0.859174	0.5999
LPIPS	0.101648	0.291946	0.255175	0.4470

As indicated in Table 2, the proposed semantic coding method based on pre-trained generative diffusion models achieves superior compression ratio performance compared with traditional JPEG and WebP. Regarding distortion metrics: In terms of PSNR, Webp outperforms JPEG, while JPEG surpasses the proposed SD_Semantic. In terms of SSIM, Webp exceeds SD_Semantic, and SD_Semantic is superior to JPEG. In terms of LPIPS, SD_Semantic significantly surpasses Webp, while Webp exceeds JPEG. Compared with the Deep-JSCC method, our proposed approach achieves better performance in terms of compression ratio, SSIM and LPIPS. Under high SNR conditions, the core advantages of the proposed algorithm lie merely in the compression ratio and LPIPS perceptual performance.

To visually validate these effects, we randomly selected five images from the dataset for qualitative comparison, as shown in Figure 5. The SD_Semantic coding method can better preserve image details compared to traditional JPEG and Webp, such as the details of the doorknob in the first image, the details of trees and rivers in the second image, the texture details of the window in the third image, the details of the flower stamen in the fourth image, and the texture details of the skin and sweater in the fifth image. That is to say, with a superior compression ratio, the SD_Semantic method can preserve more image details and exhibits superior performance in terms of human visual perception similarity. Compared with Deep-JSCC, the proposed SD_Semantic semantic coding method can preserve more image details while delivering clearer visual quality and better human perceptual quality.

3.2.2. Comparison of Algorithm Adaptability Under Low SNR Conditions

Under the fixed compression ratio in the above experimental setup, as channel quality degrades, the distortion in image transmission increases. To quantify this relationship, this thesis first simulated the bit error rate (BER) performance of 4QAM modulation under three distinct channel conditions (Gaussian channel, Rayleigh channel, and Rician channel) across varying SNR. The results, illustrated in Figure 6, include the Rician channel with a Rician factor of $K = 5$. In both Rayleigh fading and Rician fading channels, the BER becomes intolerable when the SNR drops below 20. Such channel conditions are quite common in the low-SNR satellite communication scenarios with severe interference considered in this

work. Furthermore, we conduct simulations to evaluate the performance of various coding and decoding schemes under the above channel conditions.

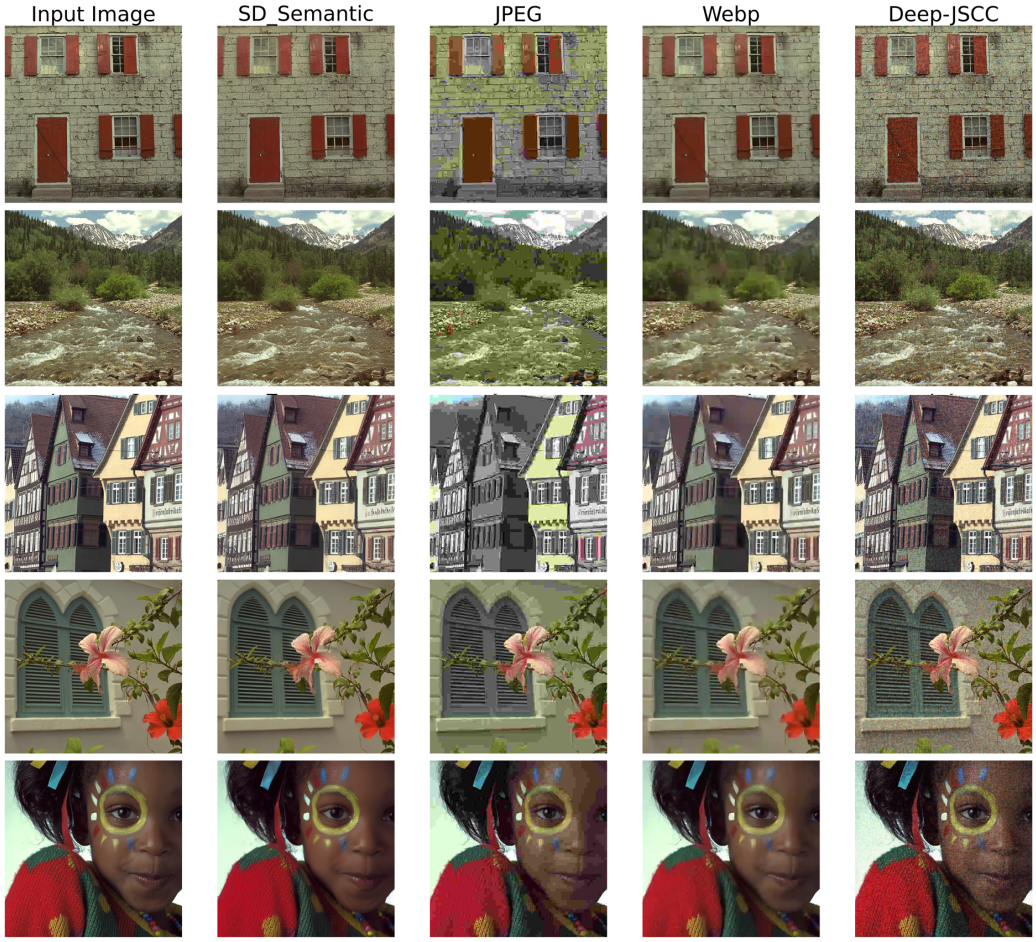


Figure 5. Visualized results of image restoration at SNR = 100 dB.

The curves in Figure 7 depict the relationship between image decoding distortion and SNR under the Gaussian fading channel. As SNR decreases, image distortion (containing PSNR, SSIM, and LPIPS) gradually increases. The results demonstrate that our proposed semantic coding method based on pre-trained stable diffusion models (SD_Semantic) outperforms traditional methods under low-SNR conditions.

As shown in Figures 8 and 9, the advantages of the SD_Semantic method are more pronounced in Rayleigh fading channels and Rician fading channels, as the BER of these channels is higher than that of Gaussian fading channels under the same SNR.

When SNR is extremely low, the high BER may lead to the complete failure of signal recovery. Taking the Rayleigh channel as an example, Figure 10 visualizes the image transmission results with increasing SNR. The results intuitively demonstrate that the proposed SD_Semantic exhibits superior adaptability to low-SNR environments. Specifically, SD_Semantic progressively reconstructs images at low SNR levels, while image distortion gradually diminishes as SNR increases.

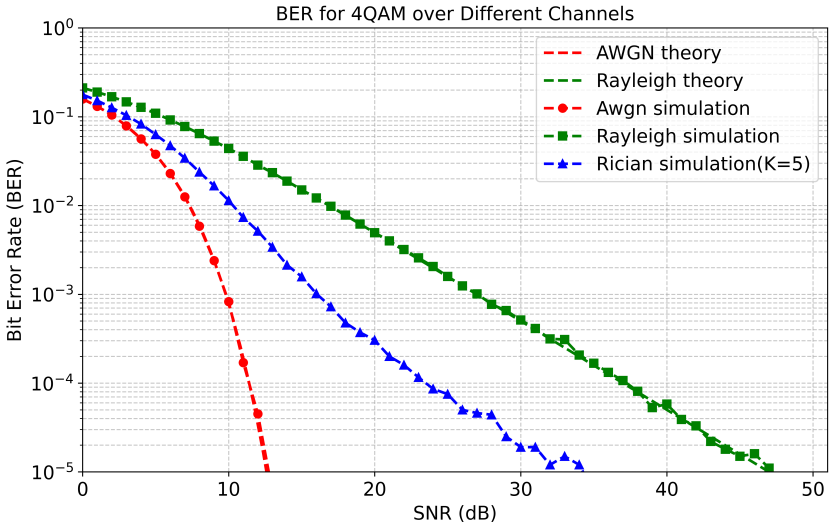


Figure 6. BER simulation results under different SNR conditions.

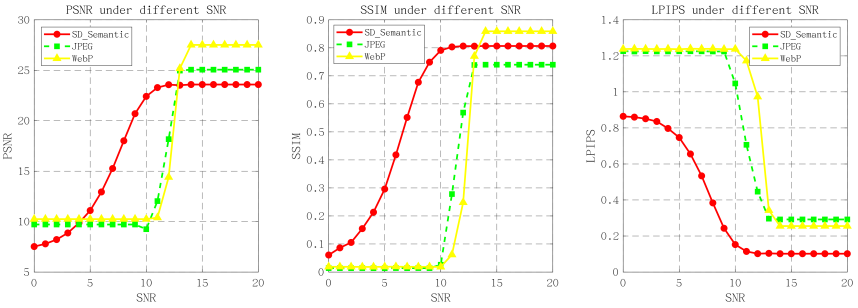


Figure 7. Image decoding distortion in Gaussian fading channel.

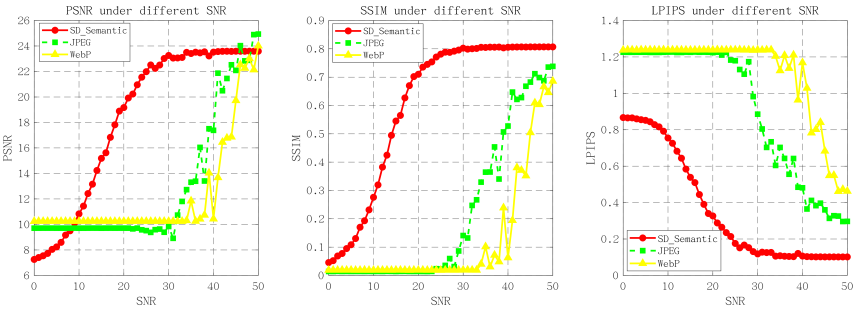


Figure 8. Image decoding distortion versus SNR in Rayleigh channel.

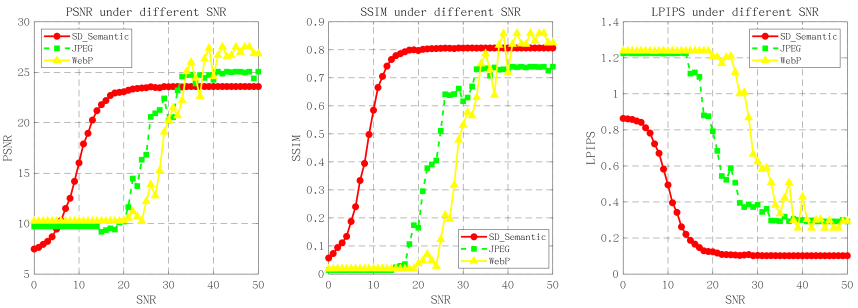


Figure 9. Image decoding distortion versus SNR in Rician channel ($K = 5$).

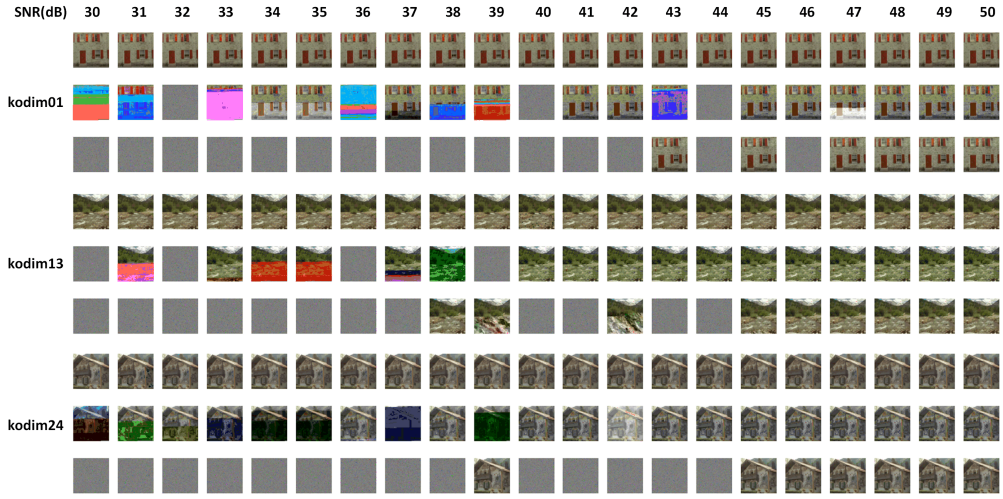


Figure 10. Visualized results of image transmission under SNR from 30 to 50 dB.

3.2.3. The Impact of Different Prompts on the SD_Semantic Method

Furthermore, different prompts exert distinct effects on image reconstruction. Taking the image "kodim15" as an example. This thesis tested the impact of varying textual descriptions on its restoration performance, as detailed in Table 3 The text description generated by Deepseek is as follows: "A close-up portrait of an adorable young girl. Her face has playful face paint, with a bright sun motif circling one eye, vibrant colors like yellow, red, and blue. Colorful ribbons adorn her hair. She's dressed in a cozy, multicolored knitted sweater with a retro–ike pattern. The lighting is gentle and even, creating a warm and lively atmosphere, photorealistic quality, high resolution, capturing the innocence and fun of childhood". As shown in Table 3, when the SNR is 10 dB, the distortion in transmitted semantic features is minimal. Consequently, image restoration relies less on prompt guidance, and while textual prompts can marginally enhance performance in distortion metrics, their impact remains statistically insignificant.

Table 3. Impact of different prompts on distortion in reconstructed images.

Prompts	PSNR	SSIM	LPIPS
No prompt	25.41834	0.86590	0.13966
High quality	25.42040	0.86594	0.13975
Masterpiece	25.42122	0.86593	0.13987
Best quality	25.41823	0.86591	0.13967
Highly detailed	25.41381	0.86586	0.13984
HDR	25.41112	0.86585	0.14000
Text description	25.41912	0.86597	0.13982

However, under low SNR conditions, the semantic spatial feature tensor is severely corrupted and may even misguide the image generation process, leading to significant distortion in the reconstructed image at the receiver compared to the original. This thesis attempted to reconstruct the image using only textual descriptions. In this study, for the Kodim15 image under Gaussian noise (SNR < 5 dB), the image generated solely by the Deepseek prompt as quoted above is shown in Figure 11b, with the original image in Figure 11a. The quantitative metrics between them are $PSNR = 10.7824$ dB, $SSIM = 0.2080$, and $LPIPS = 0.7060$. Although the reconstructed image fails to transmit the original visual content accurately, it maintains semantic perceptual similarity to the source image. The description text can be adaptively adjusted according to channel quality, ranging from several bits to hundreds of bits. A larger number of text bits enables more

detailed semantic description and achieves higher similarity between the generated image and the original image.

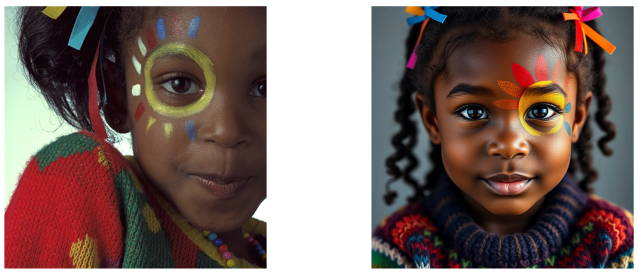


Figure 11. Comparison between the image generated solely from a textual description and the original image.

3.2.4. The Compatibility of the SD_Semantic Coding Method

The SD_Semantic coding method exhibits excellent compatibility, which can be hybridized with the traditional LDPC channel coding method to achieve superior anti-interference capability. Its superior source compression capability significantly conserves bandwidth resources while remaining compatible with LDPC codebooks exhibiting enhanced error correction capabilities. Figures 12 and 13 compare the performance of the combined semantic coding and LDPC(1/2) scheme against traditional JPEG+LDPC(1/2) and JPEG+LDPC(1/2). As demonstrated in Figure 13, even with higher compression ratios, the integration of semantic coding enables image transmission to adapt to lower SNR regions. Moreover, it demonstrates significant advantages in LPIPS under ultra-high compression ratios.

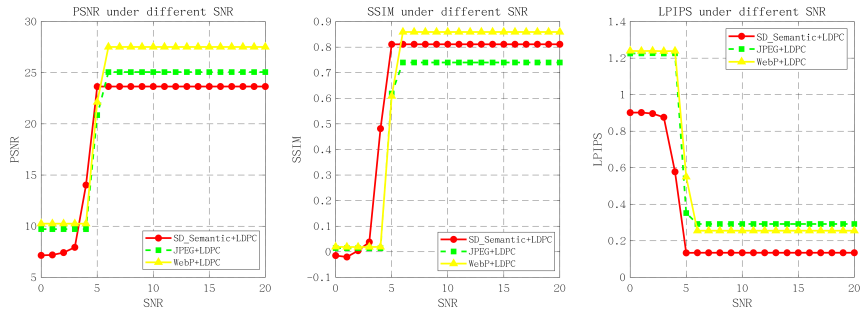


Figure 12. Image decoding distortion versus SNR in Gaussian channel.

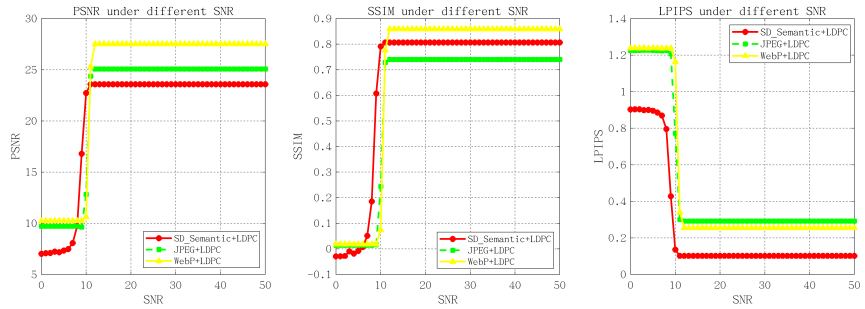


Figure 13. Image decoding distortion versus SNR in Rayleigh channel with LDPC coding.

4. Conclusions

To address the challenge of image transmission under low SNR and bandwidth-constrained conditions, we proposed an image semantic coding method, SD_Semantic,

which was based on pre-trained generative diffusion models and realized a zero-shot universal image compressor. The method integrated SNR feedback into the denoising network of the diffusion model, enhancing its adaptability to channel interference. Meanwhile, by performing quantization and entropy coding on feature tensors in the semantic space, the compression ratio was dynamically optimized. Simulation results demonstrated that while maintaining superior compression efficiency, The human perceptual quality is significantly improved compared with traditional coding and decoding methods. Furthermore, it robustly adapts to lower SNR scenarios in Gaussian noise channels, Rician fading channels, and Rayleigh fading channels, exhibiting broad application prospects in wireless image transmission scenarios with increasingly complex electromagnetic environments. The proposed algorithm is built upon the diffusion model and entails relatively high computational complexity. Given the limited computing resources of mobile communication devices, model lightweighting will be a key concern in our subsequent research.

Author Contributions: Conceptualization, S.L. and R.L.; methodology, S.L.; validation, S.L., J.Q. and Z.Y.; data curation, J.Q. and Z.Y.; writing—original draft preparation, S.L.; writing—review and editing, Y.Z.; supervision, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The validation dataset employed in this work is the widely used public dataset Kodak, which can be downloaded via the following URL: https://www.kaggle.com/datasets/sherylmehta/kodak-dataset (accessed on 20 March 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Silva, H.T.P.D.; Silva, H.S.; Figueiredo, F.A.P.; Anjos, A.A.D.; Souza, R.A.A. A survey on noise-based communication. *arXiv* **2025**, arXiv:2511.04011. [CrossRef]
2. Bourtsoulatze, E.; Kurka, D.B.; Gunduz, D. Deep joint source-channel coding for wireless image transmission. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 567–579. [CrossRef]
3. Tan, C.W. Optimal Power Control in Rayleigh-Fading Heterogeneous Wireless Networks. *IEEE/ACM Trans. Netw.* **2016**, *24*, 940–953. [CrossRef]
4. Zhang, P.; Xu, W.; Gao, H.; Niu, K.; Xu, X.; Qin, X.; Yuan, C.; Qin, Z.; Zhao, H.; Wei, J.; et al. Toward wisdom-evolutionary and primitive-concise 6G: A new paradigm of semantic communication networks. *Engineering* **2022**, *8*, 60–73. [CrossRef]
5. Shi, G.; Xiao, Y.; Li, Y.; Xie, X. From semantic communication to semantic-aware networking: Model, architecture, and open problems. *IEEE Commun. Mag.* **2021**, *59*, 44–50. [CrossRef]
6. Kalfa, M.; Gok, M.; Atalik, A.; Tegin, B.; Duman, T.M.; Arikan, O. Towards goal-oriented semantic signal processing: Applications and future challenges. *Digit. Signal Process.* **2021**, *119*, 103134. [CrossRef]
7. Wang, Y.; Han, H.; Feng, Y.; Zheng, J.; Zhang, B. Semantic communication empowered 6g networks: Techniques, applications, and challenges. *IEEE Access* **2025**, *13*, 28293–28314. [CrossRef]
8. Niu, K.; Zhang, P. A mathematical theory of semantic communication. *arXiv* **2024**, arXiv:2401.13387.
9. Jankowski, M.; Gunduz, D.; Mikolajczyk, K. Wireless image retrieval at the edge. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 89–100. [CrossRef]
10. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–20 June 2022; pp. 10684–10695.
11. Karn, A.; Kumar, S.; Kushwaha, S.K.; Katarya, R. Image synthesis using gans and diffusion models. In *Proceedings of the 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, 21–22 April 2023*; IEEE: New York, NY, USA, 2023; Volume 1, pp. 1–6.
12. Kingma, D.P.; Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. *arXiv* **2018**, arXiv:1807.03039. [CrossRef]

13. Li, C.H.Z.; Wang, X.; Hu, H.; Wyeth, C.; Bu, D.; Yu, Q.; Gao, W.; Liu, X.; Li, M. Lossless data compression by large models. *Nat. Mach. Intell.* **2025**, *7*, 794–799. [CrossRef]
14. Deletang, G.; Ruoss, A.; Duquenne, P.-A.; Catt, E.; Genewein, T.; Mattern, C.; Grau-Moya, J.; Wenliang, L.K.; Aitchison, M.; Orseau, L.; et al. Language modeling is compression. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 March 2024.
15. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
16. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *Camb. Explor. Arts Sci.* **2024**, *2*. [CrossRef]
17. Luo, C. Understanding diffusion models: A unified perspective. *arXiv* **2022**, arXiv:2208.11970. [CrossRef]
18. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2366–2369.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.