STM JOURNALS

Current Trends in
Signal Processing

ISSN: 2277-6176 (Online)
ISSN: 2321-4252 (Print)
Volume 16, Issue 2, 2026
May–August
DOI (Journal): 10.37591/CTSP

https://journals.stmjournals.com/ctsp

Review CTSP

Diffusion-Based Enhancement of Low-SNR Time-Frequency Signals

Pradnya Kharde[1,*], Sunil Kharde[2], Jyoti Lokhande[3], Sangita Wamane[4]

Abstract

Traditional enhancing techniques are useless in low signal-to-noise ratio (LSNR) situations because noise drastically interferes with communication signals. Based on an enhanced DiffBIR model, this paper suggests a dual-stage signal improvement approach that combines diffusion with deep learning. By combining the Inception module for multi-scale feature extraction with the Pixel Fusion Attention (PFA) module for significant region highlighting, the model improves signal recovery in the time-frequency domain. Experiments show that the enhanced DiffBIR model works better than conventional techniques in terms of noise reduction and time-frequency characteristic retention, with a wide range of potential applications in acoustic processing, radar, and communication. The DiffBIR method first converts communication signals into time-frequency representations using Short-Time Fourier Transform (STFT), enabling joint analysis of temporal and spectral characteristics. The enhancement framework consists of two stages. In the first stage, an improved restoration module based on SwinIR is employed, where standard convolution layers are replaced with multi-scale Inception modules to capture diverse frequency components effectively. Additionally, a Pixel Fusion Attention (PFA) mechanism is introduced to emphasize signal-dominant regions while suppressing noise at the pixel level, thereby improving structural preservation and feature clarity. In the second stage, a diffusion generation module refines the restored output by iteratively denoising latent representations, enabling recovery of high-frequency details and improving perceptual quality. A conditional guidance mechanism further enhances region-specific restoration, balancing global consistency and local detail reconstruction. Experimental results on synthetic datasets with multiple modulation schemes and noise types demonstrate that the proposed model significantly outperforms existing methods in terms of PSNR, SSIM, and LPIPS metrics. Moreover, it achieves superior performance in signal detection tasks, improving precision, recall, and F1-score across varying SNR levels. These findings demonstrate the suggested framework's resilience, effectiveness, and usefulness for actual LSNR communication systems.

***Author for Correspondence**
Pradnya Kharde
E-mail: pradnyabhingare12@gmail.com

[1,3]Assistant Professor, Department of Electronics & Telecommunication, Dr. Vitthalrao Vikhe Patil College of Engineering, Ahilyanagar, Maharashtra, India
[2]Assistant Professor, Department of Mechanical Engineering, Padmashree Dr. Vitthalrao Vikhe Patil Institute of Technology & Engineering (Polytechnic), Loni, Uttar Pradesh, India
[4]Assistant Professor, Department of Computer Science & Design, Dr. Vitthalrao Vikhe Patil College of Engineering, Ahilyanagar, Maharashtra, India

Received Date: April 10, 2026
Accepted Date: June 24, 2026
Published Date: June 25, 2026

Citation: Pradnya Kharde, Sunil Kharde, Jyoti Lokhande, Sangita Wamane. Diffusion-Based Enhancement of Low-SNR Time-Frequency Signals. Current Trends in Signal Processing. 2026; 16(2): 15–27p.

Keywords: DiffBIR model LSNR, signal recovery, time-frequency images communication signals, diffusion models, noise suppression

INTRODUCTION

In modern communication systems, signals often encounter low signal-to-noise ratio (LSNR) conditions, where noise severely masks the critical signal features. Such environments degrade the transmission quality and complicate downstream signal processing, including detection, classification, and demodulation. LSNR conditions lead to the loss of high-frequency components and distort time-frequency characteristics, making signal enhancement and restoration essential for reliable communication.

Traditional deep learning approaches, including Convolutional Neural Networks (CNNs) and Transformers, have shown promise in feature extraction for signal enhancement. However, they often struggle to recover the fine details of severely degraded signals. Recent advances in diffusion-based generative models have opened new possibilities for restoring complex signals, as demonstrated by methods such as DiffBIR, DiffIR, DiffUIR, and Diffface. However, these approaches exhibit critical limitations when applied to communication signal spectrograms.

1. Lack of specialized multi-scale frequency modeling to handle diverse signal bandwidths.
2. Standard convolution layers fail to extract features across multiple frequency bands simultaneously.
3. Uniform attention mechanisms cannot distinguish signal-dominant regions from noise-dominated regions, leading to suboptimal enhancement under LSNR conditions.
4. This study proposes an enhanced dual-stage DiffBIR framework for LSNR communication signal augmentation to address these gaps.

The key contributions are:

1. Inception modules were integrated into Residual Swin Transformer Blocks (RSTBs) for multi-scale feature extraction across diverse frequency components.
2. The Pixel Fusion Attention (PFA) mechanism is used for pixel-level attention and selectively emphasizes signal-dominant regions while suppressing noise.
3. Dual-stage architecture that balances global structure preservation in the restoration stage and high-frequency detail recovery in the diffusion generation stage.

Using the Short-Time Fourier Transform (STFT), the proposed framework first transforms the transmission signals into time-frequency pictures. The Restoration Module suppresses noise while preserving global structure, and the Diffusion Generation Module restores high-frequency details through iterative denoising. Extensive experiments demonstrate that this dual-stage approach outperforms existing methods in terms of PSNR, SSIM, LPIPS, and signal detection performance, confirming its potential for real-world LSNR communication scenarios.

MOTIVATION

This study is motivated by the need to accurately restore communication signals in extremely low SNR environments, where existing deep learning and diffusion-based methods fail to capture multi-scale frequency features and effectively distinguish signals from noise.

Objective

- Improve Signal Quality under LSNR Conditions.
- Replace standard convolutions with multi-scale Inception modules
- We Introduce Adaptive Attention Mechanism Pixel Fusion Attention (PFA), which emphasizes signal-dominant regions, suppresses noise-dominated regions, and achieves pixel-level refinement.
- Design a Dual-Stage Enhancement Framework Restoration Module and Diffusion Generation Module
- Improve Performance over Existing Methods
- Enhance Practical Signal Detection Performance increases Precision, Recall and F1-score

Scope

This study is limited to enhancing LSNR communication signals using a dual-stage diffusion-based framework applied to synthetic time-frequency representations, with an evaluation based on image quality metrics and signal detection performance under controlled noise conditions.

LITERATURE REVIEW

In low signal-to-noise ratio (LSNR) environments [1], communication signals are significantly corrupted by noise, which masks or distorts critical signal features and degrades transmission quality

and processing reliability. Signal augmentation and restoration [2, 3] are crucial for future signal processing tasks because high-frequency components and time-frequency features are lost under low signal-to-noise ratio (LSNR) conditions. To effectively enhance LSNR communication signals, this study suggests a dual-stage signal enhancement strategy based on the DiffBIR framework. This method uses creative model design and diffusion generating processes.

The process of the proposed method is as follows: to capture both temporal and spectral information, communication signals are first transformed into time-frequency pictures using the short-time Fourier transform (STFT). The improved DiffBIR framework was then applied to enhance these time-frequency images. Deep learning methods, including Convolutional Neural Networks (CNNs) and transformers [4, 5], have shown promise in feature extraction but still struggle to recover high-frequency details under extreme degradation. Recently, diffusion-based restoration methods have emerged as powerful solutions. DiffBIR [6] introduced blind image restoration by decoupling degradation removal and generative refinement; however, its first-stage restoration relies on standard convolutions that cannot capture the multi-scale frequency patterns inherent in communication spectrograms. DiffIR [7] explored adaptive diffusion-based denoising with improved noise scheduling; however, it lacks specialized attention mechanisms to differentiate signal-dominant from noise-dominated regions in time-frequency representations. Dif- fUIR [8] proposed selective hourglass mapping for universal image restoration; however, its uniform feature processing strategy is suboptimal for the highly heterogeneous energy distributions in LSNR spectrograms. Diffface [9] achieved effective blind face restoration via diffused error contraction; however, its domain-specific design does not generalize well to communication signal characteristics. These methods share three critical limitations when applied to communication signal spectrograms: (1) they lack specialized multi-scale frequency modeling for diverse bandwidth characteristics; (2) standard convolutions cannot extract features across multiple frequency bands simultaneously; and (3) uniform attention mechanisms fail to distinguish signal-dominant from noise-dominated regions under LSNR conditions, leading to suboptimal enhancement. To address these specific technical gaps in existing diffusion-based restoration methods, this study introduces an improved DiffBIR framework specifically designed for LSNR communication signal enhancement. The key technical contributions distinguishing our framework from the original DiffBIR are as follows: (1) Inception modules replace standard convolutions in Residual Swin Transformer Blocks (RSTBs) to enable multi-scale feature extraction across diverse frequency components; (2) a Pixel Fusion Attention (PFA) mechanism is introduced to provide pixel-level attention weighting, selectively emphasizing signal-dominant regions while suppressing noise in time-frequency images; (3) the dual-stage architecture is optimized to balance global structure preservation in the restoration stage and high-frequency detail recovery in the diffusion generation stage.

The first stage (Restoration Module) suppresses noise via multi-scale feature extraction and PFA-based pixel-level weighting, preserving the global signal structure. Iterative denoising generation is used in the second step (Diffusion Generation Module) to recover the high-frequency information. Experimental results demonstrate that this dual-stage framework achieves superior SSIM, PSNR, and LPIPS performance compared with traditional and deep learning methods. Ablation experiments validate the synergistic contributions of both modules.

METHODOLOGY

Overall Framework

To recover both high-frequency details and global structural characteristics for LSNR communication signals, this study proposes a dual-stage time-frequency image enhancement system. As illustrated in Figure 1, the framework consists of two stages: First, the Restoration Module extracts primary features, suppresses noise, and enhances global characteristics; Second, the Diffusion Generation Module utilizes a step-by-step denoising

Mechanism to restore high-frequency details and optimize local signal quality. The framework integrates a conditional input for iterative recovery, prioritizes key regions for enhancement, and outputs high-quality time-frequency images with improved structure and details.

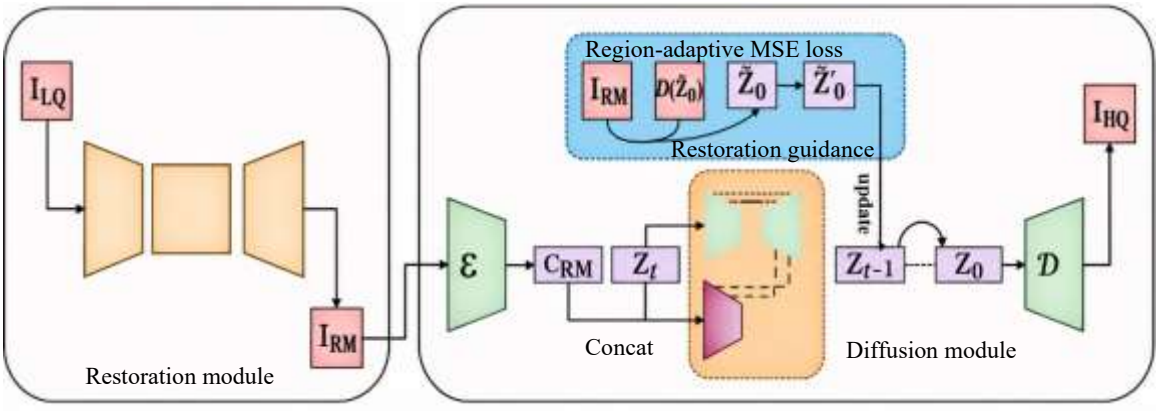

Figure 1. Overall network framework [6].

Restoration Module

The first stage aims to remove degradation and enhance time-frequency features without introducing new content. SwinIR [10] (Image Restoration using Swin Transformer) was selected as the baseline framework after a comparative evaluation against other transformer-based networks, including Restormer and Swin2SR. The selection is based on three key advantages: (1) superior balance between computational efficiency and restoration performance, which is critical for practical deployment; (2) window-based self-attention mechanism that effectively captures both local details and global structures in time frequency spectrogram; and (3) preliminary experiments (Table 3) demonstrating better preservation of high-frequency components, with LPIPS of 0.1029 outperforming Restormer and Swin2SR, while maintaining competitive PSNR and SSIM metrics. After an experimental comparison, the SwinIR framework was improved to better meet the LSNR enhancement requirements. As shown in Figure 2, the SwinIR network is composed of three parts: an Image Reconstruction Module that combines the extracted features to recreate improved time-frequency images, a Shallow Feature Extraction Module that records multi-scale features, and a Deep Feature Extraction Module that stacks several Residual Swin Transformer Blocks (RSTBs). To handle the complexity of communication signal time-frequency images, the improved restoration module introduces two key designs. First, the standard convolution in the RSTBs was replaced with the Inception module[11] to form the LE_RSTB. As shown in Figure 3.

As shown in Figure 4, the inception module in the LE_RSTB captures features at different scales via parallel convolution kernels (3 × 3, 5 × 5, and 7 × 7), combined with max-pooling to extract global features. This design effectively distinguishes between signals and noise while preserving local details. It significantly improves background noise suppression and enhances feature contrast in the signal regions, providing high-quality conditional images for the subsequent Diffusion Generation Module.

Second, the PFA module was inserted before the upsampling layer in the Image Reconstruction Module to enhance key regions in the time–frequency spectrograms. As shown in Figure 5, the PFA module uses two 1 × 1 convolutions and a sigmoid activation function to generate a pixel-wise weight map, which is multiplied by the original input feature map. This pixel-level attention mechanism guides the model to focus on signal-dominant areas and suppress irrelevant noise, improving the restoration of fine details.

To train the improved restoration module, a classic communication signal degradation model is used to generate synthetic low-quality data (LQ). The training process adopts Mean Squared Error (MSE) loss to minimize the difference between the restored image (I_{RM}) and the high-quality (HQ) reference (I_{HQ}). The restoration process and loss function are defined as follows:

$$I_{RM} = SwinIR(\mathrm{I_{LQ}})\ L\ \mathrm{MSE} = \| I_{RM} — I_{HQ} \|_2 \quad (1)$$

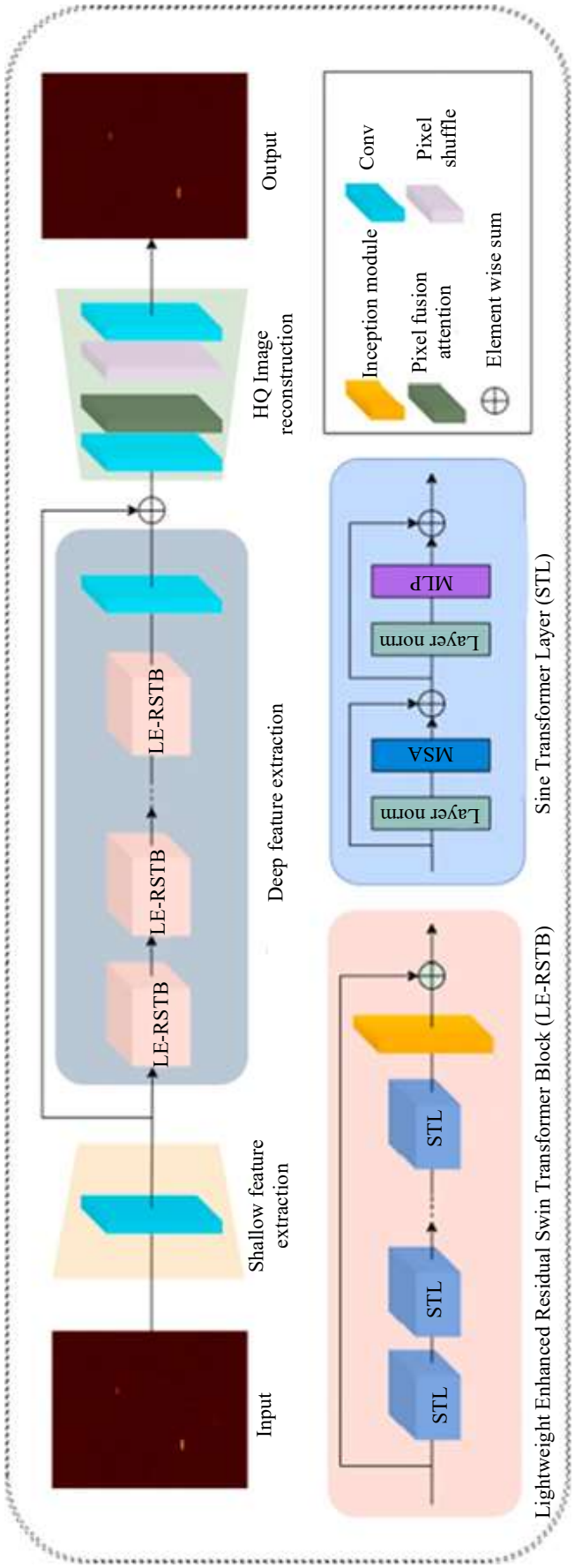

Figure 2. Improved SwinIR framework.[10].[11].

Where I_{RM} denotes the restored image (output of the Restoration Module), I_{LQ} the low-quality image, and I_{HQ} the high-quality reference image. The MSE loss function optimizes the model by minimizing the pixel-wise difference between I_{RM} and I_{HQ}.

Diffusion Generation Module

The second stage employs Stable Diffusion (a text-to-image latent diffusion model, Figure 1) to restore high-frequency details. The module operates in the latent space of a pre-trained autoencoder (encoder E , decoder D), optimized with a hybrid objective combining VAE, Patch- GAN, and LPIPS losses to reduce computational complexity.

During diffusion, Gaussian noise with variance $\beta_t \epsilon(0, I)$ is added at each time step t, producing noisy latent features. The latent feature encoding is as follows:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_t \quad (2)$$

where $\epsilon \sim N(0, I)$ and $\alpha_t = 1 — \beta_t$ as t increases, z_t approaches a standard Gaussian distribution.

The model learns to predict noise ϵ_θ by selecting random time steps t.

The optimization objective is:

$$L_{\text{ldm}} = \mathbb{E}_{z,c,t,\epsilon}\left[\|\epsilon - \epsilon_\theta\left(z_t = \alpha_t z + \sqrt{1 - \alpha_t}\epsilon, c, t\right)2\right] \quad (3)$$

where ϵ_θ denotes the noise prediction network parameterized by θ, c represents the text condition (set to null in our implementation), and the expectation is taken over the training data distribution, random time steps t, and noise samples ϵ.

To enable conditional control, the VAE encodes I_{RM} into a latent representation C_{RM}, which is concatenated with z_t as an input to a UNet-based conditional network. Training was stabilized via zero-convolution initialization, with intermediate features modulated by multi-scale conditional features. The optimization objective for IRControlNet is as follows:

$$L_{\text{IRcontrolNet}} = \mathbb{E}_{z,c,t,\epsilon,E(I_{\text{reg}})}[\|\epsilon - \epsilon_\theta(z_t, c, t, C_{\text{RM}})\|_2^2] \quad (4)$$

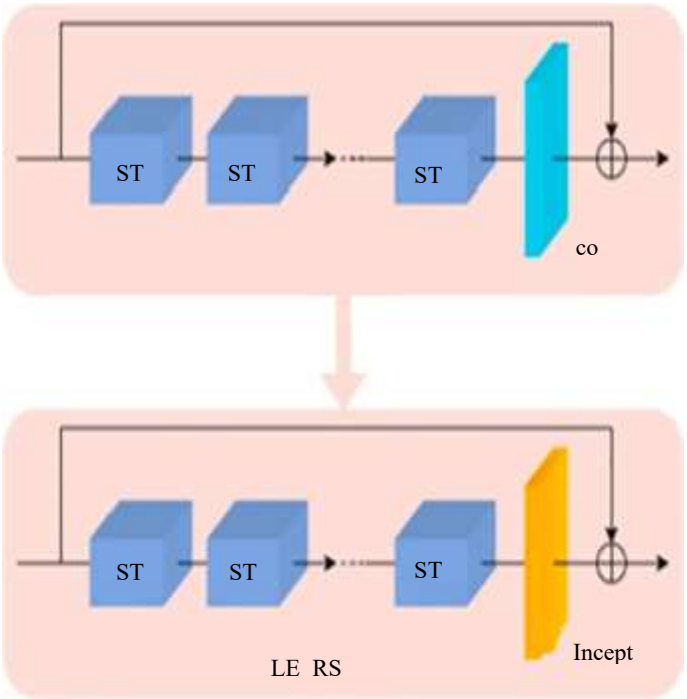

Figure 3. Inception module replacing conv module.[2].

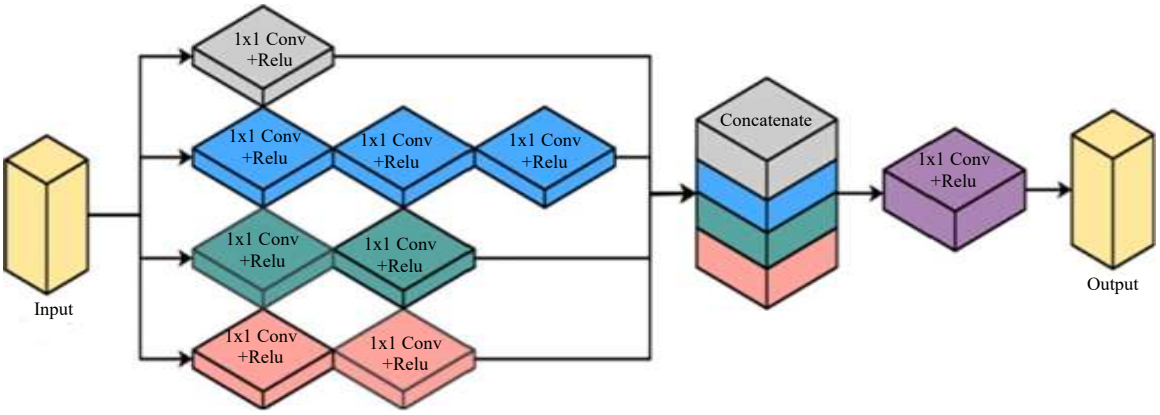

Figure 4. Inception module diagram.[2].

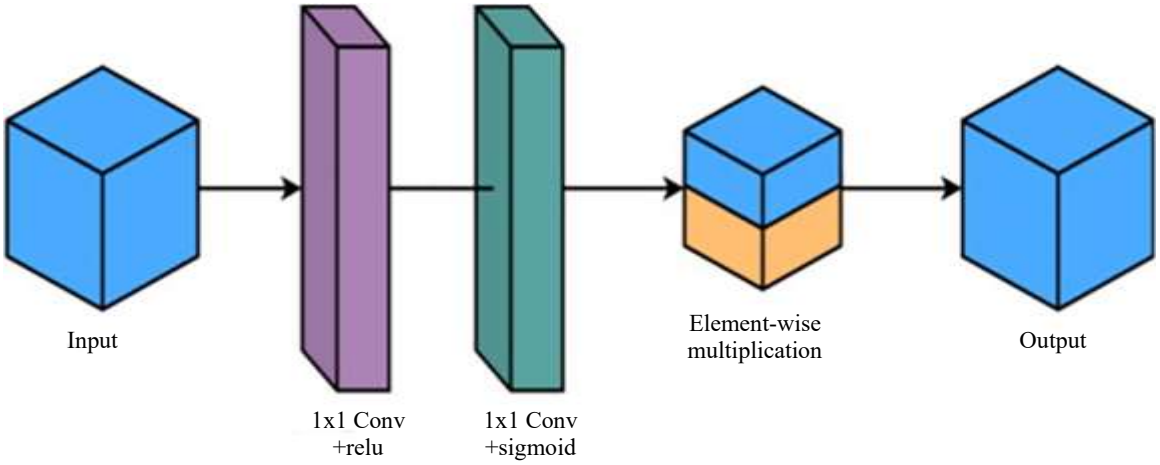

Figure 5. PFA module diagram [2].

E(Ireg) denotes the latent encoding of the restored image from the first stage.

A restoration guidance module is introduced to enable region- adaptive detail adjustment during denoising, balancing global structure and local detail. The estimated clean latent z_0 is obtained from the noisy latent z_t:

$$\dot{z}_0 = \frac{z_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(z_t, c, t)}{\sqrt{\bar{\alpha}_t}} \tag{5}$$

An adaptive MSE loss function is used, giving low-frequency areas greater weights using a weight map ω (derived from the conditional image's gradient magnitude G(IRM). The loss is defined as

$$L(\dot{z}_0) = \sum \frac{1}{CHW} \|\omega \odot (D(\dot{z}_0) - I_{\text{RM}})\|_2^2 \tag{6}$$

where D is the pre-trained VAE decoder, and ω is a gradient-based weight map that assigns higher weights to smooth regions and lower weights to edge-rich regions of the IRM.

During optimization, the clean latent representation is updated at each time step t as follows:

$$z'_0 = \dot{z}_0 - S\nabla_{\dot{z}_0} L(\dot{z}_0) \tag{7}$$

Where S is the guidance scale that controls the influence of the conditional image I_{RM}. This mechanism enables a dynamic focus on different regions, improves image quality, and provides user-adjustable control.

EXPERIMENTS

Dataset, Implementation

To validate the proposed dual-stage framework under LSNR conditions, a synthetic communication time-frequency image dataset was generated in MATLAB. The dataset encompasses nine modulation schemes (BPSK, QPSK, 8PSK, 16QAM, 64QAM, FM, AM-DSB, AM-SSB,MSK) covering analog and digital systems, with a sampling rate of 2.4 MHz and a duration of 2 s. Signal parameters were randomly sampled: center frequencies from {3, 8, 11, 16, 22, 26} MHz, symbol rates from {1.2, 2.4,4.8,9.6, 19.2} kHz, symbol durations from {0.01, 0.03, 0.05, 0.08, 0.1,0.12, 0.2} s, and start times from {0.1, 0.3, 0.5, 0.8, 1.2, 1.5, 1.6} s. To comprehensively evaluate robustness under diverse interference, three noise types were introduced: (1) AWGN with SNR ranging from —25 to —15 dB; (2) Rayleigh fading with maximum Doppler shift $f_d = 50Hz$ simulating multipath propagation; (3) Impulsive noise following Middleton Class-A model ($A = 0.1$) representing electromagnetic interference. The dataset comprised 1200 images with a balanced modulation distribution, divided into training/validation/test sets at an 8:1:1 ratio (960/120/120). Randomization used seed 231 for reproducibility. Performance across noise scenarios is detailed in Table 1.

As shown in Table 1, DiffBIR(our) consistently outperformed the baseline DiffBIR across all three noise types. The improvement is most pronounced under AWGN (LPIPS reduced by 8.7%, PSNR improved by 7.6%), whereas Rayleigh fading and impulsive noise present greater challenges owing to non-stationary interference characteristics. Notably, even under impulsive noise, the most difficult scenario DiffBIR(our) achieves an SSIM of 0.6521 versus 0.6085 for the baseline DiffBIR, demonstrating the enhanced robustness of the proposed multi-scale and attention-based modules.

Implementation: The experiments were performed on an NVIDIA TITAN RTX (24GB) with PyTorch 2.0.1, Python 3.10, and CUDA 12.1. Because our work focuses on improving the first-stage restoration module, we compared it with five representative first-stage methods (SCUNet, BBCU, BSRNet, Restormer, Swin2SR, and original SwinIR). To validate the overall framework performance, we compared the complete dual-stage system with seven state-of-the-art methods, including single-stage approaches (Uformer and BSRGAN) and dual-stage diffusion models (DiffUIR, Diffface, DiffIR, ResShift, and DiffBIR). All baselines were retrained from scratch on our time-frequency dataset using official implementations with consistent settings. First-stage training employed 50 K iterations, a batch size of 4, AdamW optimizer (learning rate decaying from 2×10^{-4} to 1×10^{-6} via cosine schedule), MSE loss, and 256 × 256 input patches. For dual-stage methods, Stage 1 followed the first-stage protocol, while Stage 2 (ControlLDM) used 100 K iterations, batch size 4, AdamW optimizer (lr 1×10^{4}), 64 × 64 × 4 latent space, 1000 training time steps, and 50 DDIM inference steps. All metrics (LPIPS-VGG, PSNR, SSIM) were evaluated on the test set, and the experiment used seed 231 to ensure reproducibility.

Ablation Experiments

Ablation experiments were conducted to validate the performance contributions of key modules (Inception module, PFA module) and optimize their configurations. The experimental results show that using the Inception module alone, the PSNR increased from 37.8503 to 38.1847, the LPIPS decreased from 0.1029 to 0.1012, and the SSIM increased from 0.9342 to 0.9522. This confirms that the Inception module improves the model's ability to represent the signal features and suppress noise.

Table 1. Performance under different noise scenarios.

Methods	Noise type	LPIPS ↓	PSNR ↑	SSIM ↑
DiffBIR (our)	AWGN	0.2626	26.4709	0.7080
DiffBIR (our)	Rayleigh	0.2897	24.8352	0.6745
DiffBIR (our)	Impulsive	0.3124	23.6418	0.6521
DiffBIR	AWGN	0.2876	24.5916	0.6852
DiffBIR	Rayleigh	0.3215	22.7643	0.6402
DiffBIR	Impulsive	0.3568	21.4957	0.6085

Ablation experiments were conducted to optimize the placement of the PFA module (Table 2). Placing the PFA module before upsampling (pre-Upsampling) achieved the best performance, with a PSNR of 40.1044 (vs. 22.5354 after upsampling and 40.0517 before output), and LPIPS and SSIM were also optimized. This placement enables the PFA module to dynamically emphasize key signal regions, providing high- quality feature inputs for the Diffusion Generation Module.

The combination of the Inception and PFA modules (pre-upsampling) significantly improved the performance (Table 2), with LPIPS, PSNR, and SSIM reaching 0.0968, 38.7333, and 0.9674, respectively.

Comparative Performance Analysis

To validate the performance of the first-stage Restoration Module, the improved SwinIR (SwinIR(our)) was compared with representative deep learning methods(SCUNet [12], BBCU [13], BSRNet [14], Restormer [15], Swin2SR [16], and original SwinIR).

As shown in Table 3, the original SwinIR already outperformed most methods (LPIPS = 0.1029, PSNR = 37.8502, SSIM = 0.9342), and SwinIR (our) achieved further improvements (LPIPS = 0.0968, PSNR = 38.7333, SSIM = 0.9672), outperforming all compared methods in perceptual quality, pixel-level accuracy, and structural similarity. This confirms the effectiveness of the Inception and PFA modules in enhancing the multi-scale feature extraction and pixel attention mechanisms.

To further validate the overall performance of the improved DiffBIR model (DiffBIR(our)) in LSNR image restoration tasks, it was compared with advanced diffusion-based methods (DiffUIR, Diffface, DiffIR, BSRGAN [17], Uformer [18], ResShift [19], and baseline DiffBIR).

As shown in Table 4, DiffBIR(our) achieves the best overall performance with LPIPS = 0.2626, PSNR = 26.4709, and SSIM = 0.7080.

Compared to the baseline DiffBIR, the improved model shows 8.7% improvement in LPIPS, 7.6% in PSNR, and 3.3% in SSIM, while reducing parameters by 0.18 M and FLOPs by 1.7%. The inference speed increased by 40.2% (from 0.82 to 1.15 fps), demonstrating that the Inception and PFA modules enhanced both the restoration quality and computational efficiency.

Table 2. Impact of different module designs and positions on the first stage.

Module configuration	Position setting	LPIPS ↓	PSNR ↑	SSIM ↑
Original Module	-	0.1029	37.8503	0.9342
Inception Module	-	0.1012	38.1847	0.9522
PFA (Pre-Upsampling)	Pre-Upsampling	0.1022	40.1044	0.9017
PFA (Post-Upsampling)	Post-Upsampling	0.3183	22.5354	0.3798
PFA (Pre-Output)	Pre-Output	0.1039	40.0517	0.8998
Combined (Inception + PFA)	Pre-Upsampling	0.0968	38.7333	0.9674

Table 3. Comparison of experimental results for different methods in the first stage.

Methods	LPIPS ↓	PSNR ↑	SSIM ↑
SCUNet	0.2270	24.8149	0.8706
BBCU	0.2202	27.1125	0.9096
BSRNet	0.1856	21.7891	0.9369
Restormer	0.1388	35.8762	0.8245
Swin2SR	0.1425	37.6531	0.8658
SwinIR	0.1029	37.8502	0.9342
SwinIR (our)	0.0968	38.7333	0.9672

Table 4. Overall performance comparison of different diffusion-based methods.

Methods	LPIPS ↓	PSNR ↑	SSIM ↑	Params (M) ↓	FLOPs (G) ↓	FPS ↑
DiffUIR	0.4381	24.5214	0.5761	156.32	387.45	2.37
DiffFace	0.4593	29.2453	0.5905	189.67	428.91	1.89
DiffIR	0.4550	19.5347	0.6946	201.48	521.77	1.24
BSRGAN	0.3918	24.1254	0.6187	16.70	548.36	3.56
Uformer	0.3542	25.8163	0.6487	50.88	233.04	4.82
ResShift	0.2979	29.9610	0.6253	312.54	892.16	0.68
DiffBIR	0.2876	24.5916	0.6852	279.78	664.69	0.82
DiffBIR (our)	0.2626	26.4709	0.7080	279.60	653.19	1.15

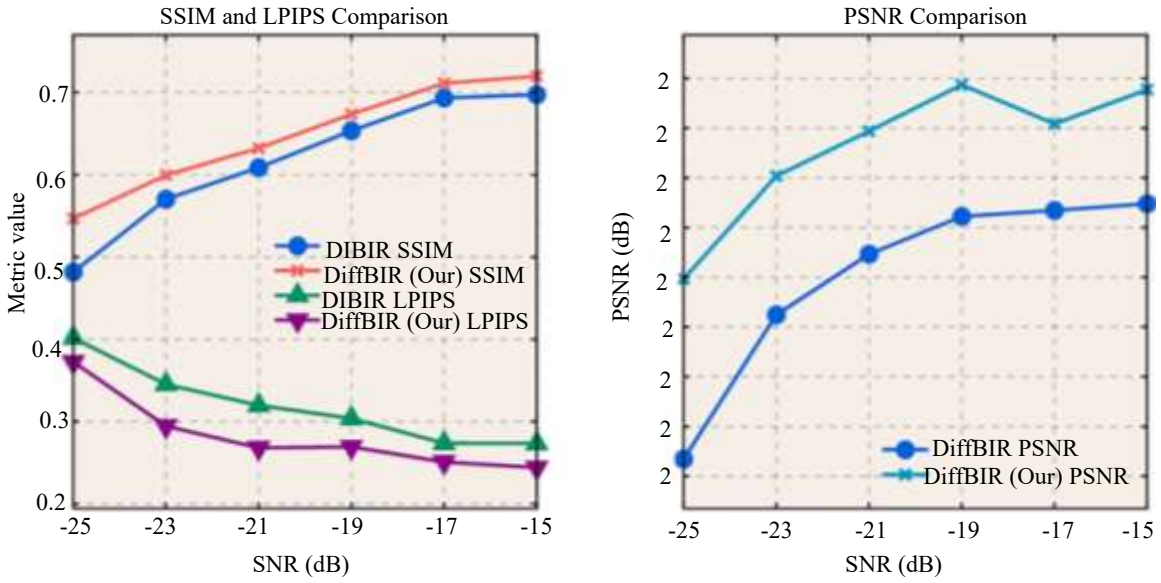

Figure 6. Evaluation metric curves of the improved model under different noise levels.

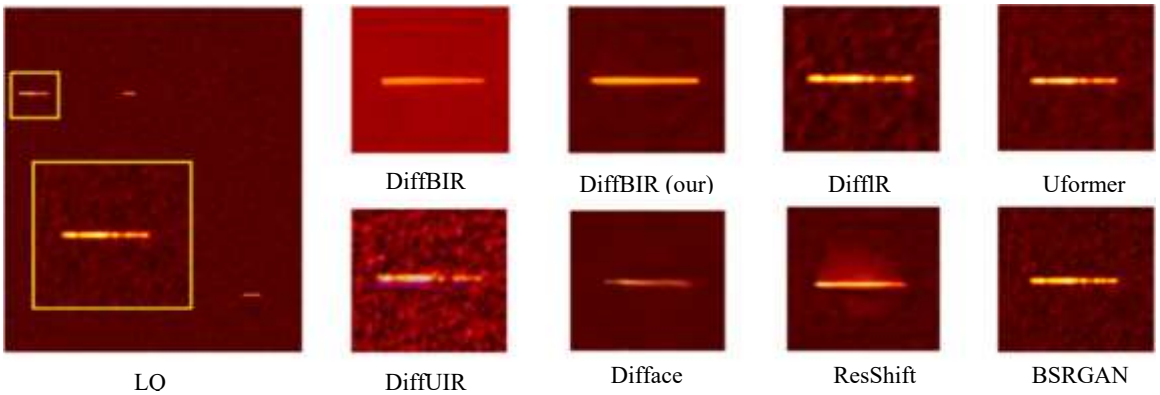

Figure 7. Visual comparison of different enhancement methods. (Left: Low- quality spectrogram with zoomed region (yellow box). Right: Same region enhanced by eight methods) [6].

To more intuitively demonstrate the effects of the improved model, it was tested across SNR levels from -25 dB to -15 dB, combining quantitative metric line charts and subjective visual comparisons. As shown in Figure 6, in the quantitative metrics, DiffBIR(our) outperforms other methods in SSIM, PSNR, and LPIPS across all noise levels. Specifically, under high noise conditions (-25 dB to -21 dB), DiffBIR(our) shows minimal performance degradation, indicating stronger robustness and detail recovery capabilities. Meanwhile, the LPIPS curve confirms that DiffBIR(our) generates time-frequency images with perceptual quality closer to real signals, outperforming other methods.

As shown in Figure 7, DiffBIR(our) demonstrates outstanding detail restoration capability in subjective visual comparisons, exhibiting sharper edges, higher overall consistency, and significantly reduced artifacts and blurring. This method effectively avoids common issues, such as line discontinuities, structural breaks, and detail loss, observed in other approaches. In contrast, competing methods exhibit various deficiencies, such as detail loss and blurring (DiffUIR, Diffface), edge blurring in local regions (DiffIR, DiffBIR), severe artifacts and discontinuities (ResShift), insufficient restoration with noise residues (Uformer), and unnatural texture artifacts (BSRGAN). These observations confirm the superiority of DiffBIR(our) for LSNR-blind image restoration tasks.

Signal Detection Performance Evaluation

To validate the engineering applicability of the proposed method beyond visual quality metrics, we conducted signal detection experiments to assess task-oriented effectiveness. The detection pipeline consists of three stages: (1) enhanced time-frequency images undergo grayscale conversion, binarization, and connected component analysis to extract the structural features of the target signal regions; (2) extracted regions are matched against the ground truth using the Euclidean distance as the similarity metric with a threshold δ=0.15 in the normalized feature space; (3) performance is quantified using precision (P), recall (R), and F1-Score, which measure the harmonic mean of P and R, providing a balanced accuracy assessment across SNR levels.

As shown in Table 5, DiffBIR(our) consistently achieves superior detection performance across SNR levels from -25 dB to -15 dB. The precision increased rapidly with SNR improvement, reaching 0.4426 at -15 dB, significantly outperforming the baseline DiffBIR (0.3948), DiffIR (0.3206), ResShift (0.3796), and the unenhanced baseline (0.2840). Recall also demonstrates strong performance, reaching 0.4176 at -15 dB, indicating robust coverage of true signal regions under complex noise conditions. The F1-Score confirms the overall superiority, achieving 0.4298 at -15 dB compared to 0.3864 (DiffBIR), 0.3181 (DiffIR), and 0.3761 (ResShift). Notably, at extremely low SNR levels (-25 dB to -21 dB), where the unenhanced baseline fails completely (F1<0.03), DiffBIR (our) maintains functional detection with F1-Scores ranging from 0.1042 to 0.2789. These results validate that the proposed framework substantially enhances downstream signal processing tasks, demonstrating strong practical applicability in real-world communication systems.

Table 5. Signal detection metrics (precision, recall, and F1-score).

Methods	**Metric**	**-25 dB**	**-23 dB**	**-21 dB**	**-19 dB**	**-17 dB**	**-15 dB**
Base	P	0.0035	0.0097	0.0123	0.0170	0.2612	0.2840
Base	R	0.0930	0.0953	0.1004	0.1117	0.2742	0.3255
Base	F1	0.0067	0.0176	0.0218	0.0294	0.2676	0.3033
DiffBIR (our)	P	0.1106	0.1251	0.2954	0.3528	0.4135	0.4426
DiffBIR (our)	R	0.0983	0.1214	0.2639	0.3082	0.3987	0.4176
DiffBIR (our)	F1	0.1042	0.1233	0.2789	0.3289	0.4060	0.4298
DiffBIR	P	0.0849	0.0895	0.2512	0.2988	0.3592	0.3948
DiffBIR	R	0.0838	0.1003	0.2155	0.2724	0.3750	0.3783
DiffBIR	F1	0.0843	0.0947	0.2321	0.2851	0.3670	0.3864
DiffIR	P	0.0087	0.0128	0.0153	0.0475	0.3057	0.3206
DiffIR	R	0.1008	0.0938	0.1520	0.1607	0.2610	0.3157
DiffIR	F1	0.0160	0.0223	0.0268	0.0723	0.2819	0.3181
ResShift	P	0.0121	0.0122	0.0280	0.0588	0.3467	0.3796
ResShift	R	0.0781	0.0981	0.1702	0.1728	0.3565	0.3727
ResShift	F1	0.0209	0.0217	0.0478	0.0871	0.3516	0.3761
Methods	Metric	-25 dB	-23 dB	-21 dB	-19 dB	-17 dB	-15 dB
Base	P	0.0035	0.0097	0.0123	0.0170	0.2612	0.2840

Future Scope

Future work can focus on applying the proposed dual-stage DiffBIR framework to real-world LSNR communication signals to improve generalization and robustness. Efforts can also be made to optimize the network for lightweight deployment, thereby enabling real-time processing on resource-constrained devices. The framework can be extended to multimodal signal enhancement and task-specific applications, such as demodulation or automatic signal classification. Additionally, exploring efficient or hybrid diffusion strategies may further improve high-frequency detail recovery while reducing the computational overhead. These directions will broaden the applicability of the model to practical communication systems.

CONCLUSION

In this study, an improved DiffBIR framework for LSNR communication signal enhancement using a dual-stage approach is proposed. The Inception module and Pixel Fusion Attention (PFA) enable effective noise suppression and multi-scale feature extraction, whereas the diffusion-based generation stage restores high-frequency details. The experiments demonstrated superior performance in LPIPS, PSNR, SSIM, and signal detection compared to existing methods, highlighting the robustness and practical applicability of the proposed framework in weak-signal environments.

REFERENCES

1. Li C, Guo H, Tong S, Zeng X, Cao Z, Zhang M, et al. NELoRa: towards ultra-low SNR LoRa communication with neural-enhanced demodulation. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems; 2021 Nov 15. p. 56-68. ACM.
2. Wang J, Huang S, Huo Z, Zhao S, Qiao Y. Bilateral enhancement network with signal-to-noise ratio fusion for lightweight generalizable low-light image enhancement. Sci Rep. 2024;14(1):29832.
3. Kumar R, Patil M. Improved the image enhancement using filtering and wavelet transformation methodologies [preprint]. SSRN. 2022 Jul 22. SSRN:4182372.
4. Jiang K, Wang Q, An Z, Wang Z, Zhang C, Lin CW. Mutual Retinex: combining transformer and CNN for image enhancement. IEEE Trans Emerg Top Comput Intell. 2024;8(3):2240-52.
5. Xiao H, Wang X, Wang J, Cai JY, Deng JH, Yan JK, et al. Single image super-resolution with denoising diffusion GANs. Sci Rep. 2024;14(1):4272.
6. Lin X, He J, Chen Z, Lyu Z, Dai B, Yu F, et al. DiffBIR: toward blind image restoration with generative diffusion prior. In: European Conference on Computer Vision; 2024 Sep 29. Cham: Springer Nature Switzerland; 2024. p. 430-48.
7. Xia B, Zhang Y, Wang S, Wang Y, Wu X, Tian Y, et al. DiffIR: efficient diffusion model for image restoration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2023. p. 13095-13105.
8. Zheng D, Wu XM, Yang S, Zhang J, Hu JF, Zheng WS. Selective hourglass mapping for universal image restoration based on diffusion model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024. p. 25445-55.
9. Yue Z, Loy CC. DifFace: blind face restoration with diffused error contraction. IEEE Trans Pattern Anal Mach Intell. 2024;46(12):9991-10004.
10. Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R. SwinIR: image restoration using Swin Transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 1833-44.
11. Si C, Yu W, Zhou P, Zhou Y, Wang X, Yan S. Inception Transformer. In: Advances in Neural Information Processing Systems. 2022;35:23495-509.
12. Zhang K, Li Y, Liang J, Cao J, Zhang Y, Tang H, et al. Practical blind image denoising via Swin-Conv-UNet and data synthesis. Mach Intell Res. 2023;20(6):822-36.
13. Xia B, Zhang Y, Wang Y, Tian Y, Yang W, Timofte R, et al. Basic binary convolution unit for binarized image restoration network [preprint]. arXiv. 2022 Oct 2. arXiv:2210.00405.
14. Li Z, Liu Y, Chen X, Cai H, Gu J, Qiao Y, et al. Blueprint separable residual network for efficient image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 833-43.

15. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH. Restormer: efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 5728-39.
16. Conde MV, Choi UJ, Burchi M, Timofte R. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration. In: European Conference on Computer Vision; 2022 Oct 23. Cham: Springer Nature Switzerland; 2022. p. 669-87.
17. Cao J, Liang J, Zhang K, Li Y, Zhang Y, Wang W, et al. Reference-based image super-resolution with deformable attention transformer. In: European Conference on Computer Vision; 2022 Oct 23. Cham: Springer Nature Switzerland; 2022. p. 325-42.
18. Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H. Uformer: a general U-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 17683-93.
19. Yue Z, Wang J, Loy CC. ResShift: efficient diffusion model for image super-resolution by residual shifting. In: Advances in Neural Information Processing Systems. 2023;36:13294-307.

