Sheikh Salman Hassan [1], Loc X. Nguyen[2], Umer Majeed[2], Zhu Han[3], Choong Seon Hong[2], and Tharmalingam Ratnarajah[4]

[1]Affiliation not available
[2]Department of Computer Science and Engineering, Kyung Hee University
[3]Department of Electrical and Computer Engineering, University of Houston
[4]Department of Electrical and Computer Engineering, San Diego State University

February 06, 2026

Denoising-Enabled Semantic Communication for Robust Earth Observation in 6G Satellite Networks: A Swin Transformer Approach

Sheikh Salman Hassan*, Loc X. Nguyen†, Umer Majeed†, Zhu Han‡,
Choong Seon Hong†, and Tharmalingam Ratnarajah§
*Institute for Imaging, Data and Communications, The University of Edinburgh, Edinburgh, EH9 3BF, United Kingdom
†Department of Computer Science and Engineering, Kyung Hee University, Yongin, 17104, Republic of Korea
‡Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004-4005, USA
§Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182-1309, USA
Emails: shassan@ed.ac.uk, {xuanloc088, umermajeed, cshong}@khu.ac.kr, hanzhu22@gmail.com, t.ratnarajah@ieee.org

***Abstract*—Non-terrestrial networks (NTNs) in 6G enable persistent Earth observation (EO), yet satellite downlinks remain bandwidth-limited and vulnerable to path loss, Doppler, and small-scale fading. We propose a semantic communication (SC) pipeline for satellite EO that deploys a lightweight Swin Transformer encoder onboard and a Swin-based decoder with an optional plug-and-play denoiser at the gateway. Training uses channel-diverse pairs of the same scene to promote channel-invariant representations, while inference operates on a single image for reliability and low latency. Across representative operating points, the gateway denoiser consistently improves reconstruction quality while keeping the transmitted feature budget unchanged. On average, it recovers about 60% of the gap in peak signal-to-noise ratio (PSNR) and roughly 25% of the gap in multi-scale structural similarity index (MS-SSIM) relative to a higher-bit-rate baseline. These results show that relocating refinement to the gateway shifts the communication–computation balance toward lower downlink load at comparable quality.**

***Index Terms*—6G, non-terrestrial networks, semantic communication, Earth observation, denoising, Swin Transformer**

I. Introduction

Integrating non-terrestrial networks (NTNs) with terrestrial infrastructure is central to the 6G vision of resilient global connectivity [1]–[3]. In Earth observation (EO), continuous sensing from low Earth orbit (LEO) produces large volumes of optical and multispectral imagery that must reach the ground promptly for downstream analysis [4], [5]. Satellite downlinks, however, remain bandwidth-limited and sensitive to impairments such as severe free-space path loss, Doppler shifts, and small-scale fading, which cause time-varying link quality and deep fades [6], [7]. Traditional source–channel pipelines struggle under these conditions: heavy compression removes task-relevant detail, while strong channel coding consumes scarce spectrum and still suffers from cliff effects.

This work has been accepted for presentation at the 2026 IEEE Wireless Communications and Networking Conference (WCNC 2026). This research is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under projects EP/X04047X/2 and EP/Y037243/1 for the TITAN Telecoms Hub. *Dr. Sheikh Salman Hassan is the corresponding author.

Semantic communication (SC) offers an alternative by transmitting task-relevant representations rather than raw pixels, improving compactness and robustness for EO imagery [8], [9]. Yet, when semantic features traverse fading NTN links, residual artifacts often remain after decoding—textures weaken, long-range structures fragment, and reconstructions may fall short of downstream task requirements, even at reduced bitrates [10].

To address this limitation, we propose a Swin Transformer–enabled SC pipeline explicitly split between the satellite and the gateway. A lightweight Swin encoder on the satellite extracts compact semantic features, while the gateway hosts the Swin decoder and a plug-and-play post-decoder Swin denoiser. This split keeps the flight segment simple and shifts optional refinement to the ground. Shifted-window self-attention efficiently captures both local textures and long-range structure, well matching EO scenes with extended contours and repetitive patterns [11]. Training uses *channel-diverse pairs*—independent fades of the same scene—to promote feature consistency, while inference remains *single-image* to maintain reliability and low latency. While SwinJSCC has demonstrated the value of Swin backbones for end-to-end joint source–channel coding [12], our focus differs: we introduce a split semantic pipeline tailored to NTN constraints, with a gateway-side plug-and-play denoiser.

We evaluate the proposed approach under path loss, Doppler, and Rayleigh fading. The results show that shifting refinement to the gateway reduces the number of transmitted features while preserving reconstruction quality. On average, the denoiser recovers about 60% of the peak signal-to-noise ratio (PSNR) gap and roughly 25% of the multi-scale structural similarity index (MS-SSIM) gap relative to a higher-bit-rate baseline. Our main contributions are:

- A Swin-enabled semantic downlink that places a lightweight encoder on the satellite and a post-decoder Swin denoiser at the gateway, improving fidelity without modifying the transmitter.
- A channel-diverse, pairwise training strategy that enforces feature consistency while retaining single-image inference for reliable, low-latency operation.
- An NTN-grounded downlink formulation that accounts for path loss, Doppler, and Rayleigh fading, evaluated using

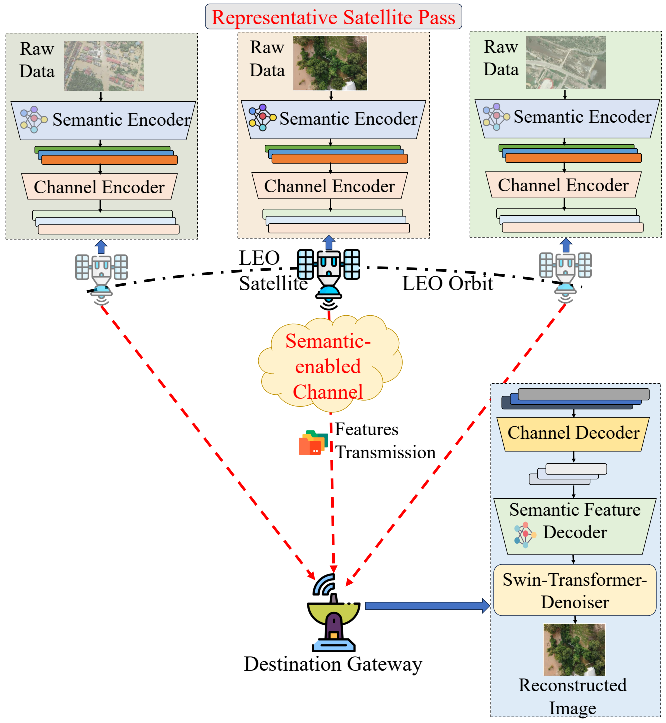

Fig. 1. SC–enabled 6G satellite EO downlink: onboard semantic encoding, NTN link, gateway decoding, and optional Swin-based denoising.

PSNR and MS-SSIM.

- An analysis of the communication–computation balance shows that relocating refinement to the gateway maintains quality while reducing the number of transmitted features and leaving onboard complexity unchanged.

II. System Model

A. Network and Signal Flow

We consider a 6G NTN downlink for EO with a constellation of LEO satellites and a terrestrial gateway, as shown in Fig. 1. Each satellite acquires an EO image $I \in \mathbb{R}^{H\times W\times C}$, processes it on board to extract a compact semantic representation F, then maps F to channel symbols, and transmits over the satellite link. At the gateway, the receiver performs channel and semantic decoding to obtain a reconstruction $\hat{I}$ and may apply a Swin Transformer denoiser to produce the output $\tilde{I}$.

B. Semantic Front-End and Channel Encoder

A semantic encoder E_α maps the input image to a d-dimensional latent, which can be defined as follows:

$$F \;=\; E_\alpha(I) \in \mathbb{R}^{d}, \tag{1}$$

where F denotes the semantic features being extracted from the raw image. A channel encoder C_β then produces a n-length complex codeword under an average power budget as:

$$X \;=\; C_\beta(F,\Gamma) \in \mathbb{C}^{n}, \qquad \frac{1}{n}\,\mathbb{E}\big[\|X\|_2^2\big] \le P, \tag{2}$$

where Γ denotes the available channel state information (CSI), such as receive signal-to-noise ratio (SNR) and Doppler. Additionally, the compression ratio (CR) $\in \{3/64, 4/64, 5/64\}$ controls n relative to the raw content.

C. NTN Downlink Channel

We adopt a narrowband, block-fading baseband model with path loss and Doppler as follows:

$$Y \;=\; HX + N, \qquad N\!\sim\!\mathcal{CN}(0,\sigma^2\mathbf{I}_n), \tag{3}$$

where H is the complex channel gain and can be represented as follows:

$$H(t) \;=\; \sqrt{G(d)}\,h\,e^{j2\pi f_d t}, \qquad G(d) = \frac{G_t G_r \lambda^2}{(4\pi d)^{\eta}}, \tag{4}$$

where $h \sim \mathcal{CN}(0,1)$ models flat Rayleigh small-scale fading, f_d is the Doppler shift due to satellite motion, and $G(d)$ is the large-scale gain determined by slant range d, antenna gains G_t, G_r, wavelength λ, and path-loss exponent η. Within a coherence block, H is constant; however, it varies independently across blocks. The receiver SNR can be defined as:

$$\mathrm{SNR} \triangleq \frac{\mathbb{E}[|H|^2]\,P}{\sigma^2}. \tag{5}$$

This abstraction follows common NTN practice for narrowband LEO links and aligns with 3GPP NTN studies and recent channel-model analyses [13]–[15].

D. Gateway Decoding and Refinement

The gateway first performs channel decoding to recover semantic features and then reconstructs the image as:

$$\hat{F} \;=\; C_\phi^{-1}(Y,\Gamma), \qquad \hat{I} \;=\; D_\theta(\hat{F}). \tag{6}$$

An image-domain denoiser is optionally applied after semantic decoding as:

$$\tilde{I} \;=\; D_{\mathrm{ST}}(\hat{I}). \tag{7}$$

The semantic codec uses a lightweight Swin backbone with a split deployment, i.e., the semantic encoder E_α runs onboard the satellite, while the channel decoder C_ϕ^{-1} and semantic decoder D_θ run at the gateway (i.e., providing implicit noise suppression during reconstruction) [16]. The gateway-side Swin denoiser D_{ST} is an optional plug-and-play refinement stage applied after D_θ.

E. Fidelity Metrics

We quantify reconstruction quality using the PSNR and the MS-SSIM between the reference I and the output $\tilde{I}$ as follows [10]:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\frac{1}{HWC}\,\|I-\tilde{I}\|_2^2}\right), \tag{8}$$

where MAX_I is the maximum pixel value and H, W, C are the image dimensions and

$$\text{MS-SSIM}(I,\tilde{I}) \in [0,1], \tag{9}$$

with higher values indicating better perceptual quality across scales.

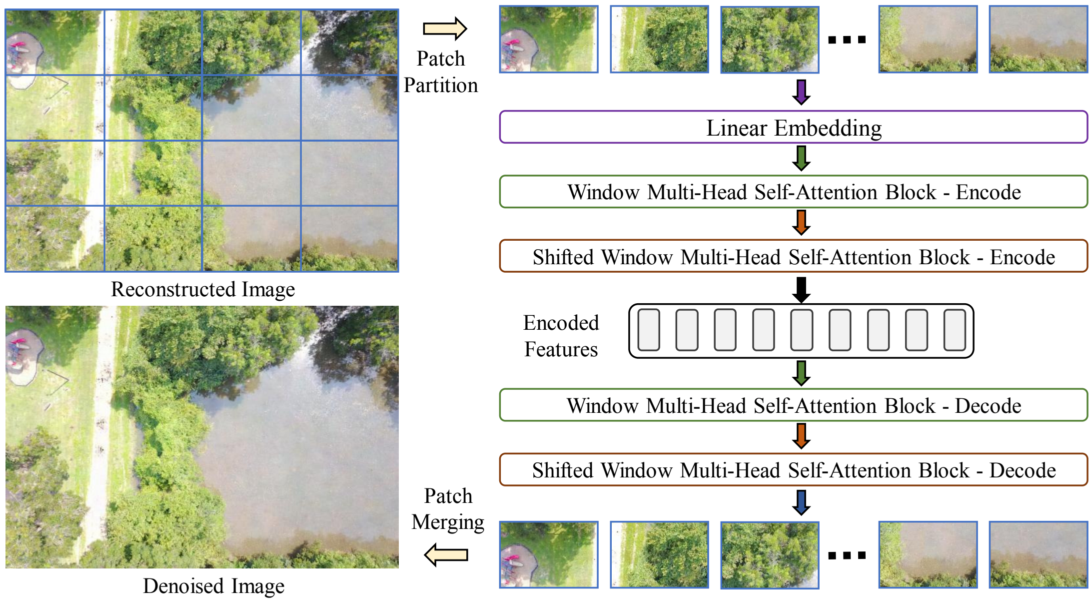

Fig. 2. Swin Transformer denoising: non-overlapping patch embedding, hierarchical windowed self-attention with cyclic shifting for cross-window interaction, and image reconstruction.

III. Proposed Framework: Swin-Enabled Semantic Downlink

A. Gateway Refinement Module

Building on the split deployment in Section II, we instantiate the gateway refinement as an *image-domain, post-decoder* Swin Transformer denoiser D_{ST} applied to $\hat{I}$ as shown in Fig. 2. D_{ST} uses windowed multi-head self-attention with cyclic shifting to exchange information across windows efficiently, which complements the decoder's implicit noise suppression without any transmitter changes. This bias is strong local modeling with controlled nonlocal aggregation, which is well-suited to EO scenes that mix fine textures and long contours. Prior Swin-based restorers (i.e., SwinIR, Restormer, Uformer) report high-fidelity reconstructions at scale, which supports our choice of a lightweight ST head for refinement [11], [17]–[19].

B. Learning Objectives and Training/Inference Protocol

To make the learned semantics robust to fading without increasing the test-time burden, training relies on *channel-diverse pairs* while inference remains strictly *single-image*. Channel-diverse pairs consist of two independently faded realizations of the same underlying scene, which encourages feature-level consistency and pushes the latent representation to become invariant to Rayleigh and Doppler fluctuations. Let $(I^{(a)}, I^{(b)})$ denote two views (i.e., augmentations) of the same scene traversing independent Rayleigh blocks $(H^{(a)}, H^{(b)})$. With $\hat{F}^{(k)} = C_{\phi}^{-1}(Y^{(k)}, \Gamma)$ and $\hat{I}^{(k)} = D_{\theta}(\hat{F}^{(k)})$, the codec is optimized using a reconstruction consistency objective inspired by neural joint source–channel coding (JSCC) for fading channels [20]:

$$\mathcal{L}_{\text{codec}} = \mathbb{E}_{H,N}\Bigg[\sum_{k\in\{a,b\}} \|I - \hat{I}^{(k)}\|_2^2 + \lambda_{\text{perc}}\,\mathcal{L}_{\text{perc}}\big(\hat{I}^{(k)}, I\big) + \mu\,\|\hat{F}^{(a)} - \hat{F}^{(b)}\|_2^2\Bigg]. \tag{10}$$

where $\mathcal{L}_{\text{perc}}$ is an optional perceptual term, and $\lambda_{\text{perc}} \geq 0$ and $\mu \geq 0$ weight the respective losses. When $\mu = 0$ (i.e., no consistency), the latent becomes more sensitive to fade variations, reducing robustness.

With $(E_\alpha, C_\beta, C_\phi^{-1}, D_\theta)$ frozen, the gateway denoiser learns the mapping $\hat{I} \mapsto \tilde{I}$ using:

$$\mathcal{L}_{\text{denoise}} = \mathbb{E}_{H,N}\Bigg[\sum_{k\in\{a,b\}} \|I - \tilde{I}^{(k)}\|_2^2 + \lambda_{\text{perc}}\,\mathcal{L}_{\text{perc}}\big(\tilde{I}^{(k)}, I\big)\Bigg]. \tag{11}$$

with $\tilde{I}^{(k)} = D_{\mathrm{ST}}(\hat{I}^{(k)})$. Training is staged, first optimizing the semantic codec and then the denoiser.

At test time, a single image I is processed through $E_\alpha \to C_\beta \to C_\phi^{-1} \to D_\theta$, and the denoiser D_{ST} can be enabled for higher fidelity or disabled for lower latency and energy.

C. Communication–Computation Trade-Off

Let q denote the quantization bits per semantic feature and s the effective spectral efficiency (i.e., bits/symbol, inclusive of channel code rate). For latent size d determined by the CR, the channel-use budget is as follows:

$$n = \left\lceil \frac{q\,d}{s} \right\rceil, \tag{12}$$

so lowering CR reduces d and thus n. On the flight segment, compute is dominated by the lightweight Swin semantic encoder E_α; on the gateway, compute is dominated by the semantic decoder D_θ and the optional denoiser D_{ST} (with the channel decoder C_ϕ^{-1} also contributing). The denoiser overhead scales approximately as:

$$\mathcal{C}_{\text{ST}} \propto H\,W\,(w_{\text{win}})^2\,L, \tag{13}$$

Algorithm 1 Swin-enabled SC downlink: training and inference

Require: Dataset $\mathcal{D}$; compression ratio (CR); hyperparameters μ, λ_{perc}; learning rates; channel parameters $(G(d), f_d)$; Rayleigh fading; batch size
Ensure: Trained codec $(E_\alpha, C_\beta, C_\phi^{-1}, D_\theta)$ and denoiser D_{ST}

1: **Stage 1: Train semantic codec** $(E_\alpha, C_\beta, C_\phi^{-1}, D_\theta)$
2: **for** epoch $= 1$ to E **do**
3:   **for** mini-batch $\{I\} \subset \mathcal{D}$ **do**
4:     Sample two views $(I^{(a)}, I^{(b)})$ (or augmentations) of each I
5:     **for** $k \in \{a, b\}$ **do**
6:       $F^{(k)} \leftarrow E_\alpha(I^{(k)})$; $X^{(k)} \leftarrow C_\beta(F^{(k)}, \Gamma)$
7:       Draw $H^{(k)}$ (Rayleigh) and $N^{(k)}$; $Y^{(k)} \leftarrow H^{(k)} X^{(k)} + N^{(k)}$
8:       $\hat{F}^{(k)} \leftarrow C_\phi^{-1}(Y^{(k)}, \Gamma)$; $\hat{I}^{(k)} \leftarrow D_\theta(\hat{F}^{(k)})$
9:     **end for**
10:     Compute $\mathcal{L}_{\text{codec}}$ as in (10) (reconstruction + perceptual + feature consistency)
11:     Update $\{\alpha, \beta, \phi, \theta\}$ by backprop on $\mathcal{L}_{\text{codec}}$
12:   **end for**
13: **end for**

14: **Stage 2: Train gateway denoiser** D_{ST} **(codec frozen)**
15: **for** epoch $= 1$ to $\tilde{E}$ **do**
16:   **for** mini-batch $\{I\} \subset \mathcal{D}$ **do**
17:     Generate $\hat{I}^{(a)}, \hat{I}^{(b)}$ with the *frozen* codec (lines 4–7)
18:     **for** $k \in \{a, b\}$ **do**
19:       $\tilde{I}^{(k)} \leftarrow D_{\mathrm{ST}}(\hat{I}^{(k)})$
20:     **end for**
21:     Compute $\mathcal{L}_{\text{denoise}}$ as in (11)
22:     Update parameters of D_{ST} only
23:   **end for**
24: **end for**

25: **Inference: single image with optional refinement**
26: Given I: $F \leftarrow E_\alpha(I)$; $X \leftarrow C_\beta(F, \Gamma)$; transmit over NTN
27: Recover $\hat{F} \leftarrow C_\phi^{-1}(Y, \Gamma)$; decode $\hat{I} \leftarrow D_\theta(\hat{F})$
28: **if** refinement enabled **then**
29:   $\tilde{I} \leftarrow D_{\mathrm{ST}}(\hat{I})$
30: **else**
31:   $\tilde{I} \leftarrow \hat{I}$
32: **end if**
33: **return** $\tilde{I}$

for image size $H \times W$, attention window size w_{win}, and L attention blocks (i.e., near-linear in image size for fixed w_{win}). As shown in Section IV, enabling D_{ST} permits operation at a lower CR with quality close to a higher-CR baseline, thereby reducing downlink symbols without disproportionate loss in PSNR or MS-SSIM. The overall process is summarized in Algorithm 1.

IV. Performance Evaluation

A. Simulation setup

We evaluate the split semantic downlink under the NTN channel of Section II, i.e., flat Rayleigh fading with elevation-dependent path loss and Doppler, plus additive thermal noise. The complex baseband per codeword follows $Y = HX + N$, where $H \sim \mathcal{CN}(0, 1)$ models small-scale fluctuations and $N \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$; the large-scale gain $G(d)$ and Doppler f_d follow (4). The receive SNR in (5) is swept over $\{1, 4, 7, 10, 13\}$ dB. The semantic codec is a lightweight Swin backbone; the Swin Transformer denoiser is applied *post decoder* without altering the transmitter. Our training process uses *DIV2K* with EO-oriented fine-tuning [21]; the codec is optimized with the joint objective in (10), and the denoiser with (11). Unless stated, results average 100 test images. We report the PSNR and the MS-SSIM per (8)–(9) at CR $= 3/64, 4/64, 5/64$. Our key simulation parameters and configuration are given in Table I.

TABLE I
Simulation parameters and experimental configuration

NTN channel / link	
Small-scale fading	Flat Rayleigh, $H \sim \mathcal{CN}(0, 1)$
Path loss	$G(d)$ per (4)
Doppler shift	± 50 kHz (LEO link)
Carrier frequency	f_c=20 GHz (Ka-band)
Orbit height	LEO, $h_s \approx 600$ km
Communication / Coding	
Compression ratio (CR)	$3/64,\ 4/64,\ 5/64$
Modulation	QPSK (baseline)
CSI Γ	Known at encoder/decoder (SNR sweep)
SNR sweep	$\{1, 4, 7, 10, 13\}$ dB (per (5))
Channel model	$Y = HX + N$, $N \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$
Training / Evaluation	
Optimizer	Adam (β_1=0.9, β_2=0.999)
Learning rate	5×10^{-4}
Batch size	B=20
Epochs	E=1000
Codec objective	(10) (recon.+perc.+consistency)
Denoiser objective	(11)
Semantic codec	Lightweight Swin backbone
Denoiser placement	Post semantic decoder (image domain)
Metrics	PSNR (8), MS-SSIM (9)
Dataset	DIV2K with EO fine-tuning
Image resolution	256×256
Averaging	Unless noted, over 100 test images

B. Results analysis

Fig. 3 shows PSNR versus SNR for CR values $3/64$, $4/64$, $4/64$+denoise, and $5/64$. Three consistent trends appear. First, less aggressive compression yields higher PSNR with no curve crossings. Second, the denoiser provides a near-uniform vertical lift to the $4/64$ curve, increasing from +0.30 dB at 1 dB to +0.45 dB at 4 dB and +0.55–0.61 dB for 7–13 dB. Third, $4/64$+denoise closely follows $5/64$ at medium and high SNR, with residual gaps of $0.29, 0.26, 0.25$ dB at 7, 10, and 13 dB. On average, the denoiser recovers about 60% of the PSNR gap and roughly 25% of the MS-SSIM gap relative to the higher-bit-rate baseline.

In Table II, PSNR and MS-SSIM increase monotonically with SNR and with larger bit allocation. At CR $= 4/64$, the denoiser improves PSNR at every operating point, e.g., 26.08→26.38 dB at 1 dB, and 27.85→28.46 dB at 13 dB. The corresponding MS-SSIM gains remain stable across SNR, with increments $\{+0.0040, +0.0040, +0.0037, +0.0036, +0.0035\}$. Over 1–13,dB, the PSNR slope of $4/64$+denoise is steeper than the baseline ($\approx$ 2.08 dB vs. $\approx$ 1.77 dB), indicating stronger recovery of high-frequency content as fades weaken.

A central outcome is that CR $= 4/64$ with denoising effectively approaches CR $= 5/64$ without denoising. The remaining PSNR gap to $5/64$ decreases to $\{0.44, 0.35, 0.29, 0.26, 0.25\}$ dB across SNR $\{1, 4, 7, 10, 13\}$ dB. In relative terms, the denoiser recovers $\{40.5\%, 56.3\%, 65.5\%, 69.4\%, 70.9\%\}$ of the baseline gap, averaging about 60.5%. For MS-SSIM, the average gap closure is about 25%, which shows that structural improvements track the distortion gains.

Fig. 4 overlays PSNR and MS-SSIM at CR $= 4/64$ with and without denoising. Both metrics improve pointwise across

TABLE II
PSNR AND MS-SSIM ACROSS SNR FOR DIFFERENT CRS WITH AND WITHOUT DENOISING UNDER RAYLEIGH FADING.

CR	Metric	SNR = 1 dB	SNR = 4 dB	SNR = 7 dB	SNR = 10 dB	SNR = 13 dB
3/64	PSNR	25.56	26.50	27.04	27.33	27.75
	MS-SSIM	0.8650	0.9012	0.9196	0.9287	0.9330
4/64	PSNR	26.08	26.95	27.45	27.72	27.85
	MS-SSIM	0.8784	0.9100	0.9256	0.9324	0.9354
4/64 w/ denoise	**PSNR**	**26.38**	**27.40**	**28.00**	**28.31**	**28.46**
	MS-SSIM	**0.8824**	**0.9140**	**0.9293**	**0.9360**	**0.9389**
5/64	PSNR	26.82	27.75	28.29	28.57	28.71
	MS-SSIM	0.8980	0.9259	0.9397	0.9459	0.9487

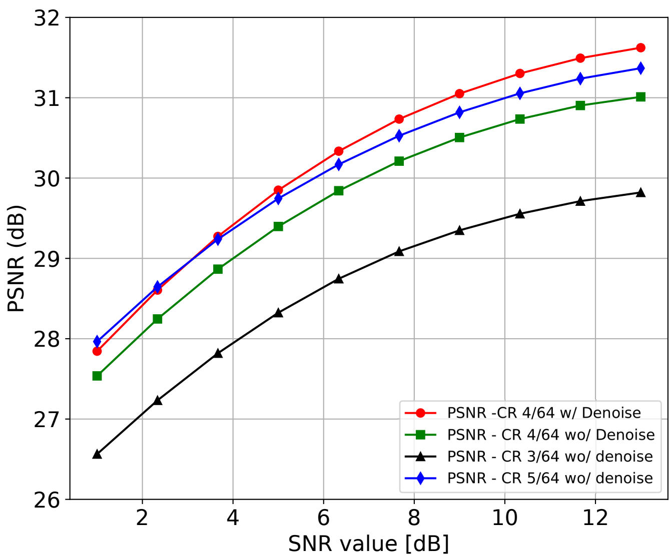

Fig. 3. PSNR vs. SNR under Rayleigh fading

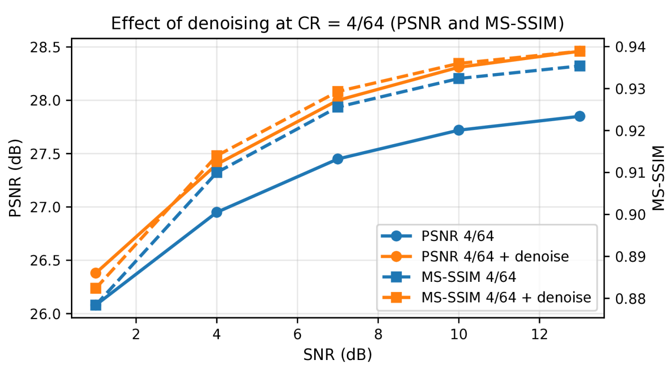

Fig. 4. PSNR and MS-SSIM vs. SNR under Rayleigh fading.

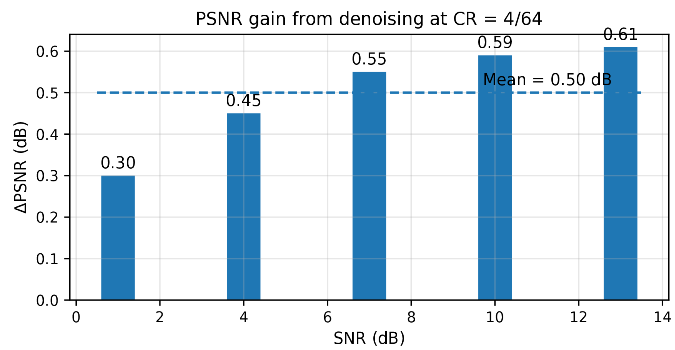

Fig. 5. Change in PSNR vs. SNR under Rayleigh fading.

TABLE III
MS-SSIM ACROSS SNR (RAYLEIGH FADING)

SNR	1 dB	4 dB	7 dB	10 dB	13 dB
CR: 3/64	0.8650	0.9012	0.9196	0.9287	0.9330
CR: 4/64	0.8784	0.9100	0.9256	0.9324	0.9354
CR: 4/64 w/ denoise	**0.8824**	**0.9140**	**0.9293**	**0.9360**	**0.9389**
CR: 5/64	0.8980	0.9259	0.9397	0.9459	0.9487

SNR, and the separation widens at higher SNR, which reflects shallower fades and the denoiser's ability to suppress the remaining image-like corruption. The PSNR gain increases from $+0.30$ dB at 1 dB to $+0.45$ dB at 4 dB and $+0.55$–0.61 dB over 7–13 dB (mean ≈ 0.50 dB). MS-SSIM gains remain almost constant across the sweep (mean ≈ 0.0038), indicating consistent improvements in cross-scale structure rather than a pure distortion shift.

Fig. 5 shows the incremental gain ΔPSNR, which increases monotonically from 0.30 dB at 1 dB to 0.61 dB at 13 dB. The rise is nearly linear with SNR (Pearson $r \approx 0.94$), suggesting an SNR-scaled benefit: at low SNR, the channel corruption dominates and limits recoverable detail, whereas at moderate and high SNR, the residual distortion becomes more image-like and is suppressed more effectively by the denoiser. A trapezoidal AUC comparison shows that denoising recovers about 62% of the CR $= 4/64 \rightarrow 5/64$ PSNR gap and about 25% of the corresponding MS-SSIM gap, reinforcing that $4/64$+denoise can serve as an operational substitute for $5/64$ while transmitting fewer features. Similarly, Table III lists MS-SSIM for all SNR and CR settings, which highlights a stable gain of $+0.003$–0.004 at CR $= 4/64$, confirming that the improvements extend beyond PSNR.

Fig. 6 provides example reconstructions at SNR $=$ 0dB, a deep-fade regime where semantic decoding alone leaves structured residuals such as edge softening and texture collapse. Without refinement, fine contours and repetitive patterns (e.g., roofs, field boundaries) degrade noticeably. The Swin-based denoiser restores spatial coherence and high-frequency detail, raising MS-SSIM from 0.824 to 0.835 and PSNR from 25.02 to 25.47dB. These examples reinforce that a lightweight, post-decoder transformer yields measurable (PSNR) and perceptible (MS-SSIM and visual) improvements and brings CR $= 4/64$ close to CR $= 5/64$ while reducing feature transmission. The

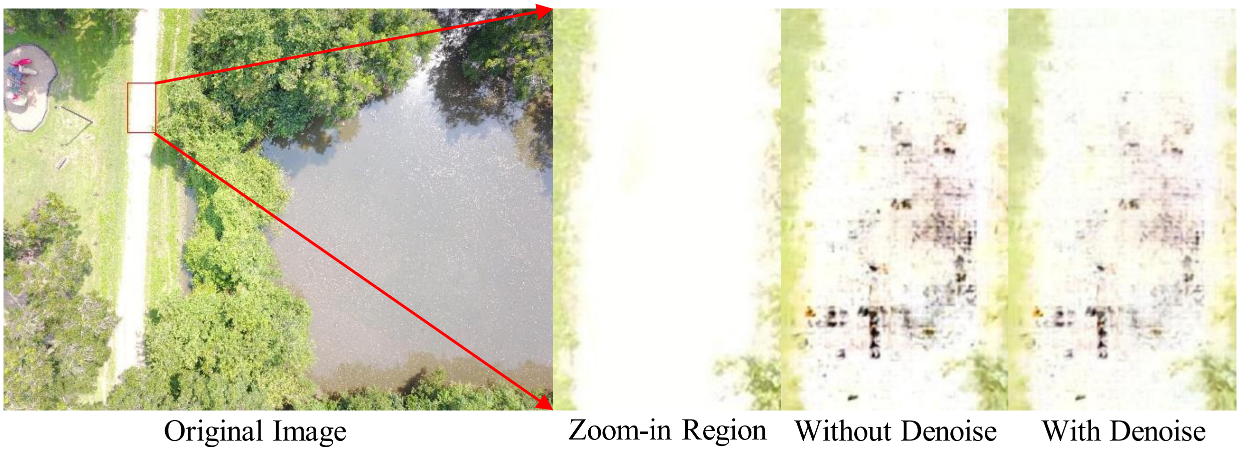

Fig. 6. Reconstructions at SNR = 0 dB (Rayleigh) with and without denoising.

sharper boundaries and more coherent textures also improve interpretability for EO tasks such as land-cover discrimination and object delineation.

V. Conclusion

We presented a semantic downlink for NTN-enabled EO that places a lightweight Swin-based encoder on the satellite and a gateway stack comprising the channel decoder, semantic decoder, and an optional post-decoder Swin denoiser. The system model incorporates path loss, Doppler, and flat Rayleigh fading, and the learning protocol uses channel-diverse pairs during training while relying on a single image at inference. Across the considered SNRs and compression ratios, the gateway denoiser consistently improves PSNR and MS-SSIM and narrows the gap to a higher-bit-rate baseline, with the strongest gains appearing in moderate-to-high SNR regimes. The results show that the denoiser recovers edges and textures that the baseline decoder cannot fully reconstruct, while remaining plug-and-play and requiring no changes to the transmitter. Future work includes extending the evaluation to Rician and shadowed NTN links, which introduce modest mobility and CSI uncertainty, exploring simple adaptive CR and on/off denoising under latency and energy limits, and studying multi-pass fusion and lightweight model compression.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, Oct. 2020.

[2] S. S. Hassan, L. U. Khan, Y. M. Park, M. Guizani, Z. Han, T. Ratnarajah, and C. S. Hong, "Quantum machine learning for 6G space–air–ground integrated networks: A comprehensive tutorial and survey," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 3710–3750, Sep. 2025.

[3] O. Y. Kolawole, S. Vuppala, M. Sellathurai, and T. Ratnarajah, "On the performance of cognitive satellite-terrestrial networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 668–683, Oct. 2017.

[4] A. Herrmann and H. Schaub, "Reinforcement learning for the agile earth-observing satellite scheduling problem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 5235–5247, Oct. 2023.

[5] Q. Zhao, L. Yu, Z. Du, D. Peng, P. Hao, Y. Zhang, and P. Gong, "An overview of the applications of earth observation satellite data: Impacts and future trends," *Remote Sens.*, vol. 14, no. 8, p. 1863, Apr. 2022.

[6] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. M. Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani, E. Lagunas, and B. Ottersten, "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 4, pp. 2633–2672, Aug. 2022.

[7] V.-P. Bui, T. Q. Dinh, I. Leyva-Mayorga, S. R. Pandey, E. Lagunas, and P. Popovski, "Semantic image encoding and communication for Earth observation with LEO satellites," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 1210–1224, Apr. 2025.

[8] Z. Liu and T. Ratnarajah, "Multi-user semantic communication for interactive speech dialogue with dynamic resource allocation," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 5, pp. 3901–3918, May 2025.

[9] L. X. Nguyen, S. S. Hassan, Y. M. Park, Y. K. Tun, Z. Han, and C. S. Hong, "SemSpaceFL: A collaborative hierarchical federated learning framework for semantic communication in 6G LEO satellites," *IEEE Trans. Commun.*, vol. 74, pp. 1269–1286, Nov. 2025.

[10] S. S. Hassan, L. X. Nguyen, Y. K. Tun, Z. Han, and C. S. Hong, "Semantic enabled 6G LEO satellite communication for Earth observation: A resource-constrained network optimization," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Cape Town, South Africa, Dec. 2024.

[11] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021.

[12] L. X. Nguyen, A. D. Raha, P. S. Aung, D. Niyato, Z. Han, and C. S. Hong, "A contemporary survey on semantic communications: Theory of mind, generative AI, and deep joint source-channel coding," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 2377–2417, Oct. 2025.

[13] 3GPP, "Solutions for new radio (NR) to support non-terrestrial networks," 3GPP, Tech. Rep. TR 38.821, 2020.

[14] R. Wang, M. A. Kishk, and M.-S. Alouini, "Modeling and analysis of non-terrestrial networks by spherical stochastic geometry: A survey," *IEEE Commun. Surv. Tutor.*, Apr. 2025.

[15] M. Sandri, M. Pagin, M. Giordani, and M. Zorzi, "Implementation of a channel model for non-terrestrial networks in ns-3," in *Proc. Workshop on ns-3 (WNS3 '23)*. Arlington, VA, USA: Association for Computing Machinery, 2023, pp. 28–34.

[16] L. X. Nguyen, S. S. Hassan, Y. K. Tun, K. Kim, Z. Han, and C. S. Hong, "Semantic communication enabled 6G-NTN framework: A novel denoising and gateway hop integration mechanism," *IEEE Trans. Wireless Commun.*, vol. 24, no. 12, pp. 10 149–10 165, Dec. 2025.

[17] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool *et al.*, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021.

[18] S. W. Zamir, A. Arora, S. K. Gupta *et al.*, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[19] Z. Wang, X. Cun, J. Bao, W. Zhou *et al.*, "Uformer: A general U-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[20] D. B. Kurka and D. Gündüz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 178–193, May 2020.

[21] R. Timofte, S. Gu *et al.*, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017.