

SemCom: Semantic-enhanced On-board Satellite Image Compression at Ultra-low Bitrates

Ziyuan Zhang¹, Jun Liu^{1*}, Hewu Li¹, Tianwei Zhang², Han Qiu^{1*}

¹Tsinghua University, China, ²Nanyang Technological University, Singapore
ziyuan-z23@mails.tsinghua.edu.cn, {juneliu, qiuhan}@tsinghua.edu.cn

Abstract

Although increasingly deployed for in-orbit earth observation (EO) imagery, existing neural image compression, following the design for natural images, struggles due to (1) a clear semantic structures gap and (2) computational inconsistencies between different in-orbit and ground hardware induce decoding errors. Observing EO imagery is dominated by large, semantically coherent regions (e.g., forest, urban), we propose SemCom, which uses a semantically partitioned codebook for efficient, region-aware bit allocation. SemCom introduces semantic-adaptive entropy coding with offline probability estimation, ensuring a fixed, platform-agnostic probability model for bit-exact decoding. Experiments on real-world satellite payloads show SemCom outperforms SOTA baselines with low in-orbit complexity.

Keywords

Satellite image compression, neural codec, ultra-low bitrates

ACM Reference Format:

Ziyuan Zhang¹, Jun Liu^{1*}, Hewu Li¹, Tianwei Zhang², Han Qiu^{1*}. 2026. SemCom: Semantic-enhanced On-board Satellite Image Compression at Ultra-low Bitrates. In *63rd ACM/IEEE Design Automation Conference (DAC '26)*, July 26–29, 2026, Long Beach, CA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3770743.3804035>

1 Introduction

Neural image compression (i.e., neural codecs) [31, 44] has achieved superior rate-distortion performance to traditional methods [43] for natural images. The mainstream neural codecs first transform an image into quantized latent representations, which are then losslessly entropy coded into a compact bitstream for transmission. The core of entropy coding relies on a learned entropy model, which functions as a unified global probability prior, to execute online probability estimation for the latent representation. For decompression, the identical entropy model calculates the same probabilities to decode the bitstream into quantized latent representations, which are then transformed into the decompressed image.

Recent research [45, 48] explores deploying neural codecs for compressing earth observation (EO) imagery in-orbit, driven by the need to leverage on-board computing resources (e.g., CPU [1],

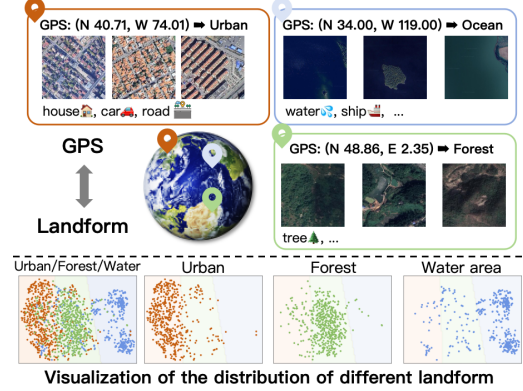


Figure 1: Satellite images have highly different visual patterns leading to distinct semantic structures.

FPGA [23], GPU [40], etc.) to satisfy downlink bandwidth constraints [16]. However, directly applying existing neural codecs designed for natural images to satellite images causes two issues. 1) **Semantic inefficiency.** The learned entropy model is a unified global probability prior [19, 28], which is inherently misaligned with satellite images, because they possess highly different statistical properties. As shown in Figure 1, satellite images are composed of distinct macroscopic semantic regions (such as urban, water area, and forest) with highly different visual patterns, which are clearly separable, corresponding to different geographical metadata (e.g., GPS coordinates). We visualize the distribution of the latent features in different regions in Figure 1 to show such distinct statistical properties. Thus, applying one global probability model to these diverse regions is suboptimal and inefficient for the bit allocations. 2) **Cross-platform unreliability.** The entropy model relies on online probability estimation [8, 24], where the probability distribution is calculated during the encoding and decoding processes. This process involves complex floating-point calculations, which may produce different results on different hardware (e.g., a satellite's onboard CPU/FPGA v.s. a ground station's GPU) due to differences in instruction sets and rounding precision [8]. Such computational inconsistencies may cause decoding errors at the ground station.

In this paper, to address the issues above, we propose SemCom, a vector-quantization (VQ) based satellite image compression method via a semantically partitioned codebook. The basic idea of SemCom is to learn discrete features for different macroscopic semantic regions and store them in a shared global codebook, encoding images as a sequence of codebook indices rather than learning a unified probabilistic prior for entropy coding. To enhance semantic efficiency, we propose semantic-guided codebook training, which utilizes a pre-trained segmentation model and contrastive learning to aggregate the discrete features belonging to the same semantic

* Corresponding authors.



This work is licensed under a Creative Commons Attribution 4.0 International License. *DAC '26, Long Beach, CA, USA*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2254-7/2026/07

<https://doi.org/10.1145/3770743.3804035>

Table 1: In orbit satellites with various hardware.

Satellites	Launch date	On board computers
Landsat 8 [1]	2013	RAD750 CPU
Sentinel-2 [2]	2015	Atmel ERC-32 CPU
Phi-Sat-1 [5]	2020	Intel® Movidius™ Myriad™ 2
Ingenuity [4]	2020	Qualcomm® Adreno™ 330
TechEdSat 13 [6]	2022	Intel Loihi Gen-1 Kapoho Bay
BroncoSat-1 [34]	2022	NVIDIA Jetson Nano
Chaohu-1 [40]	2022	NVIDIA Jetson Xavier NX
Taijing-3 [7]	2024	NVIDIA Jetson AGX Orin

category, thereby improving the codebook’s semantic representation. To tackle the cross-platform unreliability, we propose a semantic-adaptive entropy coding with offline probability estimation. By using pre-calculated probability models, SemCom ensures consistency across all hardware, thus preventing the decoding errors caused by computational inconsistencies. Moreover, such a strategy enables more precise probability distributions for different semantic regions, achieving higher compression efficiency.

We compare SemCom with 5 state-of-the-art (SOTA) baselines across 4 macroscopic semantic regions, considering 4 metrics. Extensive experiments show that SemCom significantly outperforms all SOTA baselines across different metrics at ultra-low bitrates, achieving a 3× higher compression ratio while maintaining similar decompressed image quality. Moreover, we deploy SemCom on real satellite payloads, considering 1 CPU and 2 GPU devices, to demonstrate its low in-orbit encoding complexity and ability to achieve 100% decoding accuracy with a high compression ratio in cross-platform scenarios.

2 Preliminaries

2.1 Satellite image compression methods

Transmitting high-fidelity satellite images back to Earth under downlink bandwidth constraints is a core challenge for Earth observation tasks. Early approaches, such as Mars rovers Viking [39], transmit raw data sequentially, pixel by pixel, leading to poor transmission efficiency. With the development of image compression algorithms, transform-based coding standards like JPEG [43] and JPEG2000 [41] are widely adopted to improve transmission efficiency. Driven by advancements in Artificial Intelligence (AI) and on-board computing, ESA’s satellite Phi-Sat-2 [21] deploys an in-orbit neural codec, significantly improving the compression ratio and reconstruction quality while prompting the exploration of neural image compression for satellite images.

Subsequent research has actively explored various ways to enhance neural codecs for satellite images. Considering both high-frequency and low-frequency information of satellite images, [45] uses discrete wavelet transform and designs a frequency domain encoding-decoding module to improve the reconstruction quality. To achieve extremely low bitrate compression, [46] leverages vector maps as guidance for pre-trained diffusion models. [18] proposes a mixed hyperprior network by exploring local and non-local redundancy to improve the probability estimation accuracy. Some works focus on onboard deployment. [21] deploy the compression model on VPU, extracting image features and reducing the image dimension by CAE model. [48] considers the computing resource constraints on satellites and designs a lightweight image encoder and a diffusion based compensation model for decoder.

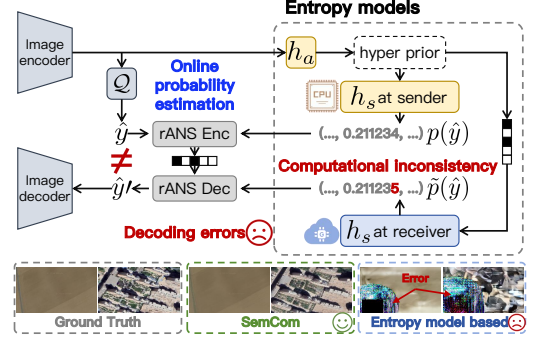


Figure 2: Pipeline for entropy model based algorithm.

However, the above methods are all entropy model based neural codecs, which makes them suffer from non-determinism issues when deployed on different types of hardware on satellites. Besides, none of them pay attention to the distinct macroscopic semantic regions of satellite images to further introduce conditions to improve the compression ratio and reconstruction quality.

2.2 On-board computing hardwares

On-board computers (OBCs) are the core of satellites’ in-orbit data processing. Unlike standard ground computing environments, satellite hardware exhibits high heterogeneity. With the development of nanosatellite technologies, the current in-orbit OBCs include both radiation-hardened CPUs [1, 2] and commercial-off-the-shelf (COTS) devices, such as GPUs [4, 6, 40] and VPUs [5], as shown in Table 1. Such hardware varies in architecture, instruction sets, and manufacturers, leading to a severe problem: **floating-point calculations lack consistency across different hardware**. More critically, satellites have long operational lifecycles, and no hardware can be changed after launch. Thus, to ensure that ground stations can correctly decode data returned from any satellite, the image compression algorithm must be cross-platform reliable.

2.3 Neural codecs

The mainstream neural codecs primarily consist of two main components: non-linear transformation [50] and entropy model [26], as shown in Figure 2. Non-linear transformation uses auto-encoder to transform the image and the latent representation into each other. The entropy model plays an important role in optimizing the compression ratio, as it estimates the probability distribution of the latents for entropy coding (such as arithmetic coding [38]), where more accurate estimation leads to lower bitrates. Hyperprior entropy model [9], autoregressive context modeling [35], discrete Gaussian mixture model [13], and spatial-channel contextual adaptive model [25] have all been proposed to improve the accuracy of probability estimation, thereby increasing the compression ratio.

Although entropy model based neural codecs have significantly improved the compression performance, they suffer from the non-determinism issue [8]. As illustrated in Figure 2, such methods rely on a neural network to dynamically extract the hyper prior from the latent feature, which is then used to estimate the latents’ probability distribution. The computational inconsistency in cross-platform scenarios inevitably produces different probability estimation on different hardware. As entropy coding requires the encoder and

decoder to share strictly identical probability models, such subtle differences may lead to catastrophic, cascading decoding errors.

Besides entropy model based neural codecs, some research [12, 22, 33] explores compressing images using vector-quantization. VQVAE [42] first introduces the concept of discrete latent space, transforming the continuous representation into discrete vectors using a trainable codebook. By transmitting the indices of discrete vectors and storing the same codebook in both sender and receiver ends, VQ-based compression methods bypass the online probability estimation, thereby solving the non-determinism issue. Existing methods change the codebook size K and the resolution of the latent features to adjust the bitrates. However, the fixed-length codes ($\log_2 K$ bits) resulting from VQ's uniform probability assumption are suboptimal for satellite images, as codeword usage frequency varies greatly across different macroscopic semantic regions.

3 Method

3.1 Overview

In this section, we describe in detail our VQ-based image compression method (section 3.2). To enhance the codeword semantic presentation, SemCom leverages a pre-trained segmentation model to inject the semantic map into the latent space and aggregate the codewords corresponding to the same category by contrastive learning (section 3.3). To ensure cross-platform reliability, meanwhile increasing the compression ratio, SemCom performs offline probability estimation with validation datasets to pre-calculate the codeword usage frequency for different semantic regions (section 3.4).

3.2 VQ-indices compression

Based on the highly structured semantic characteristics of satellite images, an intuitive approach is to train a specialized codebook for each semantic region. Such a codebook only contains features for its corresponding region, thereby achieving high-fidelity reconstruction with a smaller codebook size. However, due to the satellite images covering large geographic regions, diverse scene features often co-exist in satellite images. For instance, an urban region dominated by buildings will still contain “forest” features like trees. Training separate “urban” and “forest” codebooks would cause such features to be duplicated. This codeword overlap introduces unnecessary redundancy, reducing the compression performance.

To solve the above question, we choose to use a global codebook with semantic partition. As shown in Figure 3, SemCom consists of three components: image encoder $\mathcal{E}$, the codebook with a size of K , and image decoder $\mathcal{D}$. For an input image $x \in \mathbb{R}^{H \times W \times 3}$, the image encoder $\mathcal{E}$ compresses it into a continuous compact latent representation $z \in \mathbb{R}^{h \times w \times d}$, where d is the dimension of the vectors in the codebook. The continuous representation z is then quantized into the discrete space by looking up the nearest codeword e_k for each grid feature $z_{i,j}$, and can be presented as a sequence of indices $I_1, I_2, \dots, I_N, N = h \times w$, following:

$$I_{i \times w + j} = \arg \min_{k \in 1, 2, \dots, K} \|z_{i,j} - e_k\|. \quad (1)$$

The receiver converts the sequence of indices into discrete latent representations z_q by searching the codeword from the codebook and decodes z_q into a decompressed image $\hat{x}$ by $\mathcal{D}$.

3.3 Semantic-guided codebook training

Existing VQ-based image compression methods train their codebooks solely by minimizing reconstruction loss. Such semantic-agnostic training approaches yield codewords that are semantically chaotic in their spatial distribution. Thus, subsequent semantic-guided entropy coding (see detail in section 3.4) becomes inefficient as it cannot capture the high-frequency usage characteristics of codewords for each macroscopic category. To solve this problem, we propose semantic-guided codebook training, as illustrated in Figure 3. SemCom first leverages a pre-trained segmentation model to extract the segmentation map $\mathcal{M} \in \mathbb{R}^{h \times w}$ with m classes for latent representation z . For each codeword e_k , SemCom find the closest encoded latent feature $z_{i,j}$ as an anchor, following:

$$(i, j) = \arg \min_{i \in 1, \dots, h, j \in 1, \dots, w} \|z_{i,j} - e_k\|. \quad (2)$$

The target semantic class c_k for e_k is $\mathcal{M}(i, j)$. Note that, different from Equation 1, we try to find the nearest latent representation for each codeword in the codebook, thereby assigning the corresponding class as a semantic label, rather than finding the nearest codeword for the latent representation.

With the target class c_k , we take the codeword with the same class as positive samples and with different classes as negative samples. To further enhance the semantic structure of the discrete latent space and reduce codeword overlap between different macroscopic semantic regions, we propose semantic contrastive loss:

$$\mathcal{L}_{sc} = \log \left(\frac{\sum_{e_p \in P} e^{\cos(e_k, e_p)/\tau}}{\sum_{e_p \in P} e^{\cos(e_k, e_p)/\tau} + \sum_{e_n \in N} e^{\cos(e_k, e_n)/\tau}} \right), \quad (3)$$

where e_p is the positive sample and e_n is the negative sample, and τ is a scalar temperature parameter to control the sensitivity for hard negative samples.

3.4 Semantic-adaptive entropy coding

To further increase the compression ratio, previous methods utilize entropy coding to losslessly compress codebook indices. The naive assumption of uniform distribution results in the coding length of approximately $\log K$ bits per index. However, the distribution of codewords in satellite imagery is highly non-uniform and context-dependent. For example, in forest regions, the codeword presenting “tree” feature appears at much higher frequencies than the codeword presenting “wave” feature, and in water area, vice versa. Such observation implies that: 1) a uniform model is highly inefficient, and 2) even a single non-uniform global model is insufficient, as it fails to capture these strong, local statistical variations.

To solve the problems, we propose a semantic-aware static probability model that leverages the unique characteristics of the satellite data. We first define R macroscopic semantic regions $\mathcal{R}_i, i = 1, 2, \dots, R$. For each macroscopic semantic region $\mathcal{R}_i$, we prepare a small validation dataset $\mathcal{D}_{val,i}$. Based on the well-trained image encoder, decoder, and codebook, we process all samples $x \in \mathcal{D}_{val,i}$ to find their corresponding codewords. We then perform frequency statistics to estimate a region-specific prior probability $P(k|\mathcal{R}_i)$ for each codeword e_k within $\mathcal{D}_{val,i}$. Such process yields R semantic-aware Cumulative Distribution Functions (CDFs), $\{\mathcal{CDF}\}_{i=1}^R$, which

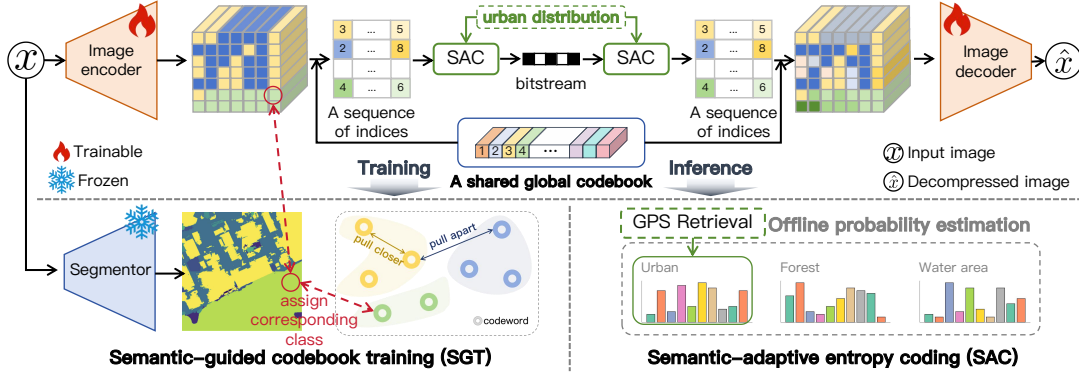


Figure 3: Overview of SemCom. SemCom leverages a shared global codebook and transmits the indices of codewords (section 3.2). For training, SemCom uses a pre-trained segmentation model to inject the semantic information into the codebook (section 3.3). For inference, the probabilities of different semantic regions are calculated offline and selected based on GPS data (section 3.4).

will be stored in both sender and receiver end. Note that all the computations above are performed in a one-time, offline process, which will not lead to cross-platform inconsistency.

At the inference stage, we use rANS [15] for efficient and fast entropy coding based on the pre-calculated semantic-aware CDFs. For an input image, SemCom first determines its macroscopic semantic region $\mathcal{R}_i$ by referencing its GPS metadata. Note that the GPS data is standard metadata and is transmitted to the ground by default in earth observation missions, so using the GPS data does not bring any additional overhead. The sender selects the corresponding CDF_i and uses rANS to encode the indices sequence. The receiver performs the identical GPS-based lookup to select the same CDF_i and decodes the bitstream. Such design ensures determinism by relying entirely on static lookup tables, eliminating any dynamic floating-point calculations. Moreover, SemCom is highly extensible, since the codebook contains general features of all satellite images, adding a new semantic region only requires computing a new CDF table from a small validation set for that category, without retraining the neural network.

3.5 Training details

We adopt an adversarial training strategy [20] to train the image encoder $\mathcal{E}$, image decoder $\mathcal{D}$, and the codebook together using the loss function as Equation 4.

$$\mathcal{L}_{fir} = \mathcal{L}_{rec} + \mathcal{L}_{per} + \sigma \mathcal{L}_{sc} + \mathcal{L}_{dis}, \quad (4)$$

where $\mathcal{L}_{rec}$ is the $\mathcal{L}_1$ loss between the decompressed image $\hat{x}$ and the origin image x , $\mathcal{L}_{per}$ is the perceptual loss to improve the visual fidelity at ultra-low bitrates, $\mathcal{L}_{sc}$ is the proposed semantic contrastive loss and $\mathcal{L}_{dis}$ is the discriminator loss [17]. To avoid the codebook collapse, we use the exponential moving average (EMA) as [49] following Equation 5.

$$e_k^{(t)} = e_k^{(t-1)}(1 - \alpha_k^{(t)}) + z_{(i,j)}^{(t)} \alpha_k^{(t)}, \quad (5)$$

where t is the iteration step, $z_{(i,j)}^{(t)}$ is the encoded latent feature from Equation 2, $\alpha_k^{(t)}$ is the decay value to control the update magnitude for codeword e_k , and can be calculated as $\alpha_k^{(t)} = \exp^{-1000 N_k^{(t)} K - \epsilon}$. $N_k^{(t)}$ is the average usage number of codeword e_k , and ϵ is a small constant to ensure the codeword update remains active.

4 Experiment

4.1 Experiment setup

Dataset. We use MEET dataset [30], which has 80 fine-grained categories as our training data and test data. For training dataset, we randomly select 500k images from the MEET dataset with all categories, and randomly crop the image to a resolution of 256×256 . For test dataset, we only consider the coarse-grained categories that can be divided by GPS data. We consider 4 macroscopic semantic categories, including urban, agricultural land, water area, and forest. We totally select 40 test images, with 10 images from each category, and crop them to a resolution of 256×256 .

Metrics. Bits per pixel (bpp) is used to measure the compression ratio. Image assessment metrics, such as PSNR and SSIM, are only devised for pixel-level distortion, which may lead to over-blurring at ultra-low bitrates. Therefore, following [22], we focus more on perceptual quality and consider 4 metrics for quantitative measures. For perceptual similarity comparison, we consider LPIPS [47] and DISTS [14]. For human preference and quality assessment, we consider PIEAPP [36] and MUSIQ [27].

Implementation details. We jointly train the encoder $\mathcal{E}$, decoder $\mathcal{D}$ and the codebook using a batch size of 32. We first train for 50k iterations using the loss function in Equation 4 without discriminator loss and for another 100k iterations with discriminator loss. We use SegEarth-OV [29], an open-vocabulary method based on CLIP [37], as the pre-trained segmentation model and set $\sigma = 0.1$ to ensure the aggregation between codewords along the same macroscopic semantic categories. All training experiments are performed on 8× NVIDIA H800 GPUs using the Adam optimizer with a learning rate of $1.5e-4$. For inference stage, we use a validation set of 5k images for each category to pre-calculate the semantic-aware CDFs.

Baselines. We compare SemCom with 5 baselines, including traditional methods, VAE-based methods, and diffusion model based methods. For traditional methods, we consider two SOTA methods, BPG [3] and VTM [11]. For VAE-based methods, we choose HL_RS [45] and TCM [32], designed for satellite and natural images, respectively. The VAE-based methods are usually solely optimized for mean square error (MSE), which can lead to excessive image blurring, resulting in visual artifacts. We further enhance the above two methods with LPIPS-based perceptual loss following [10] to

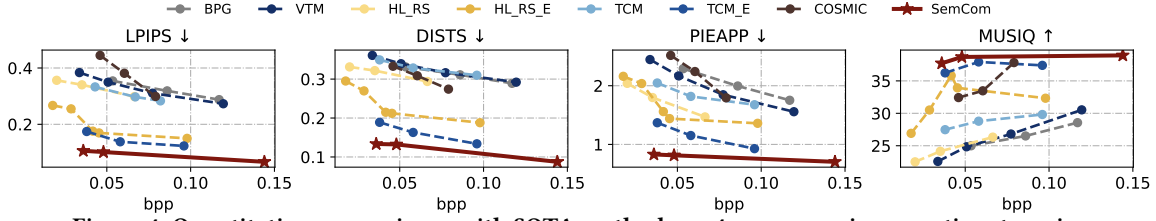


Figure 4: Quantitative comparisons with SOTA methods on 4 macroscopic semantic categories.

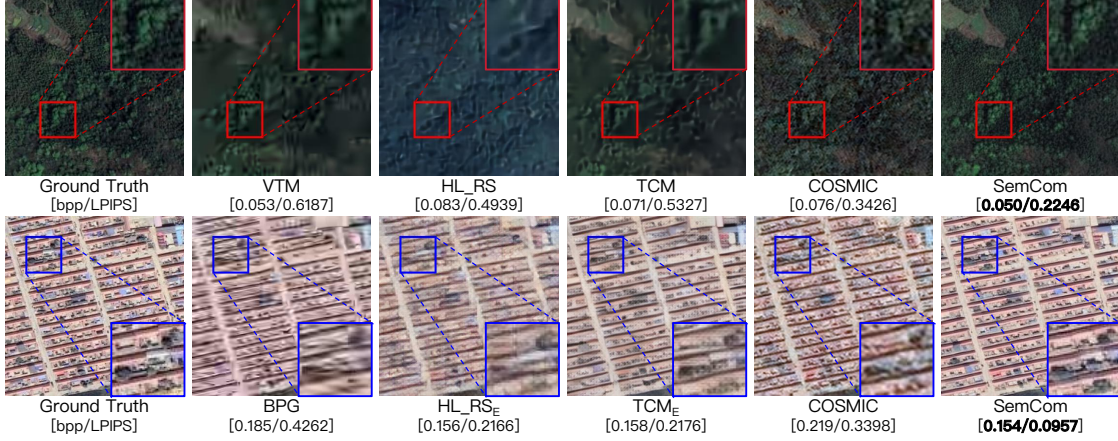


Figure 5: Visual results comparison of decompressed images in forest and urban regions.

improve the perceptual quality, denoted as HL_RS_E and TCM_E . For diffusion model based method, we consider COSMIC [48], a satellite image compression method with diffusion compensation.

4.2 Evaluation results

Quantitative comparisons. First, to compare the compression performance between SemCom and the baselines, we use the same devices for both the sender and receiver, ensuring the correct encoding and decoding. Figure 4 shows the average quantitative results on 4 macroscopic semantic categories. The results demonstrate that SemCom outperforms all baseline methods across all bitrates. At ultra-low bitrates, VAE-based methods optimized for MSE have poor perceptual quality of decompressed images. HL_RS_E and TCM_E incorporated with perceptual loss achieve significantly improved perceptual quality. However, they also rely on a unified probability prior to estimate probabilities for different macroscopic semantic regions, which leads to inefficient bit allocation and results in suboptimal compression performance. COSMIC uses a conditional diffusion model for compensation, however, at ultra-low bitrates, the latent features extracted by image encoder are not enough to guide the diffusion model to generate high-fidelity images, resulting in poor reconstruction quality. SemCom learns a codebook of rich features and transmits indices rather than quantized features, avoiding the problem of severe feature information loss at ultra-low bitrates. Furthermore, with a semantically partitioned codebook, SemCom computes distinct probability models for different macroscopic semantic regions, enabling more efficient bit allocation.

Visual results. Figure 5 shows the example of decompressed images at different bitrates on distinct semantic regions. Traditional methods, such as BPG and VTM, bring severe block artifacts, especially at object edges, such as building outlines. As we illustrated

above, VAE-based methods optimized for MSE loss result in over-blurring, which obscures fine-grained details, the most useful information for satellite images. Although enhancing by perceptual loss mitigates the over-blurring issue, it introduces subtle grid-like artifacts, which also degrade visual quality. COSMIC leverages the diffusion model for detail compensation, achieving better visual effects. However, at ultra-low bitrates, the transmitted quant feature cannot provide enough conditions for the diffusion generation process. Consequently, the decompressed image fails to generate accurate details, such as the variations in forest color. Compared to baselines, SemCom maintains a higher similarity in structure and fine-grained details (e.g., clear building structures and variations in forest color), resulting in significant visual improvements.

4.3 Ablation study

We conduct the ablation study by removing **SGT (semantic-guided training)** and **SAC (semantic-adaptive entropy coding)**. The quantitative results are in Figure 6. Further, we analyze SemCom’s performance on different macroscopic semantic regions and investigate the impact of the validation dataset size.

W/o SGT. We remove the pre-trained segmentation model and the semantic contrastive loss when training the model. The results in Figure 6 show that SGT can help save 14% ~ 123% bitrates, demonstrating that SGT successfully organizes the discrete latent space, providing a critical foundation for efficient entropy coding.

W/o SAC. We remove SAC and use a uniform distribution for all semantic regions, as shown in Figure 6. The results show that SAC can achieve an average of about 10% bitrate saving. We further select two typical semantic regions, urban and water area, to compare SemCom’s performance in different regions, as shown in Table 2. For water area, SAC can save even 30% bitrates, because the water area

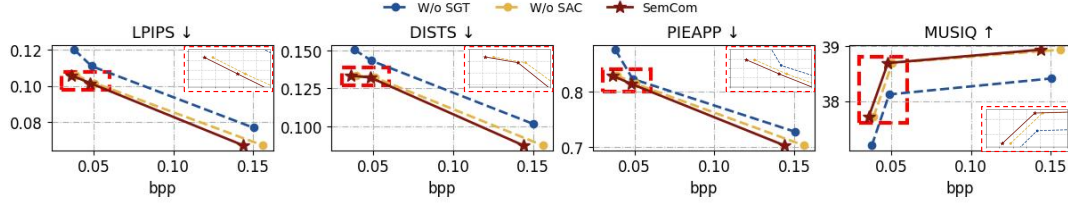


Figure 6: Ablation study with different variants of SemCom.

Table 2: Compression ratio (bpp) ablation of different semantic regions and validation dataset size evaluation.

	Urban			Water area		
W/o SAC	0.039	0.051	0.156	0.039	0.051	0.156
min	0.037	0.047	0.152	0.029	0.036	0.104
max	0.039	0.051	0.158	0.037	0.051	0.156
avg	0.038	0.048	0.153	0.032	0.046	0.121
$\mathcal{D}_{val} = 1k$	0.038	0.049	0.153	0.033	0.047	0.122
$\mathcal{D}_{val} = 3k$	0.038	0.049	0.153	0.032	0.046	0.121
$\mathcal{D}_{val} = 5k$	0.038	0.048	0.153	0.032	0.046	0.121
$\mathcal{D}_{val} = 10k$	0.038	0.048	0.153	0.032	0.046	0.121

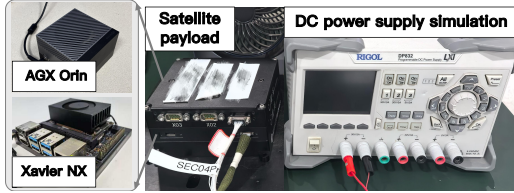


Figure 7: We evaluate SemCom on two real satellite payloads: Chaohu-1’s payload Nvidia Jetson Xavier NX and Taijing-3’s payload Nvidia Jetson AGX Orin.

only covers simple features, primarily using a very small subset of the codewords, resulting in a highly non-uniform probability distribution. In contrast, for urban, SAC only saves 2% ~ 4% bitrates. This is due to urban area contains not only buildings but also features like water and trees, using more codewords, leading to a more uniform probability distribution than water area.

Influence of validation dataset size. We further investigate the impact of validation dataset size on the estimation accuracy of the codeword usage distribution. As shown in Table 2, only a small validation dataset can achieve good estimation results, demonstrating that a small dataset can capture the true probability distribution, showing the practical and low validation complexity of SemCom.

4.4 Use case analysis in real satellite payload

We deploy SemCom on two real satellite payloads, Nvidia Jetson Xavier NX, the payload of Chaohu-1, and Nvidia Jetson AGX Orin, the payload of Taijing-3 (as shown in Figure 7), for evaluation.

Encoding complexity comparison. Table 3 illustrated the in-orbit encoding time comparison with traditional methods and neural codecs, with an input resolution of 256×256. To ensure a fair comparison with traditional methods, we evaluate BPG, VTM, and the ONNX version of SemCom on a CPU device. To compare with other neural codecs, we report two settings: 1) the performance of standard PyTorch models to enable the fair comparison with baselines, and 2) the performance of the TensorRT version, denoted

Table 3: In-orbit encoding time comparison on three testbeds. SemCom denotes the Pytorch model and SemCom * denotes the model is converted into ONNX version for CPU deployment and TensorRT version for GPU deployment.

Methods	CPU (s)		GPU (s)		bpp ↓	LPIPS ↓
	Intel(R) Xeon(R) Silver 4210R CPU	Nvidia Jetson Xavier NX	Nvidia Jetson AGX Orin			
BPG [3]	0.443	-	-		0.053	0.3541
VTM [11]	2.98 ~ 9.16	-	-		0.051	0.3497
HL_RS _E [45]	-	0.860	0.247		0.045	0.1694
TCM _E [32]	-	0.506	0.205		0.058	0.1379
COSMIC [48]	-	0.268	0.158		0.060	0.3812
SemCom	-	0.168	0.040		0.048	0.1015
SemCom *	0.369	0.055	0.017		0.048	0.1017

Table 4: Compression performance in cross-platform scenarios. Acc. (%) denotes the decoding accuracy.

Methods	Acc. (%) ↑	LPIPS ↓	DISTS ↓	PIEAPP ↓	MUSIQ ↑	bpp ↓
HL_RS [45]	0	0.7938	0.5932	7.4641	22.9706	0.042
TCM [32]	0	0.5225	0.4674	3.3570	34.0088	0.038
COSMIC [48]	0	0.7429	0.5523	5.3672	22.2853	0.046
SemCom	100	0.1059	0.1322	0.8140	37.7112	0.036

as SemCom *, which is the actual model deployed on the satellite payloads. The results demonstrate that SemCom is significantly faster than traditional methods, and also outperforms other neural codecs in the PyTorch comparison. Furthermore, SemCom * achieves greater acceleration with negligible performance degradation.

Cross-platform results. To demonstrate that SemCom can ensure accurate decoding across different platforms, we deploy the encoder on Nvidia Jetson AGX Orin and the decoder on NVIDIA H800 GPU, which has a different architecture from Orin. As shown in Table 4, due to the computational inconsistencies, all entropy model based neural codecs cannot decode correctly, resulting in low perceptual quality. SemCom can achieve 100% decoding accuracy with an offline probability estimation, while still maintaining the significant compression gains with semantic-adaptive entropy coding.

5 Conclusion

We present SemCom, a semantic-enhanced on-board satellite image compression method. We propose semantic-guided training to learn a semantically partitioned codebook and semantic-adaptive entropy coding with offline probability estimation to ensure a fixed, platform-agnostic probability model for bit-exact decoding. Extensive experiments on real-world satellite payloads indicate that SemCom can achieve superior and accurate compression at ultra-low bitrates with low in-orbit encoding complexity.

Acknowledgments

This work is supported by National Natural Science Foundation of China (62272266 and 62132009) and the Center of High Performance Computing, Tsinghua University.

References

- [1] 2013. Landsat8. <https://www.eoportal.org/satellite-missions/landsat-8-ldcm#spacecraft>.
- [2] 2015. Sentinel-2. https://sentinels.copernicus.eu/documents/247904/349490/S2_SP-1322_2.pdf.
- [3] 2018. BPG image format. <https://bellard.org/bpg/>.
- [4] 2020. Ingenuity. <https://www.qualcomm.com/products/mobile/snapdragon/smartphones/snapdragon-8-series-mobile-platforms/snapdragon-processors-801>.
- [5] 2020. Phisat. https://www.esa.int/Applications/Observing_the_Earth/Phisat.
- [6] 2023. TechEdSat. https://ntrs.nasa.gov/api/citations/20230011460/downloads/Brainstack_SSC23_Final.pdf.
- [7] 2024. Taijing-3. https://space.skyrocket.de/doc_sdat/taijing-3.htm.
- [8] Johannes Ballé, Nick Johnston, and David Minnen. 2018. Integer networks for data compression with latent-variable models. In *International Conference on Learning Representations*.
- [9] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. 2018. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436* (2018).
- [10] Yochai Blau and Tomer Michaeli. 2019. Rethinking lossy compression: The rate-distortion-perception tradeoff. In *International Conference on Machine Learning*. PMLR, 675–685.
- [11] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J Sullivan, and Jens-Rainer Ohm. 2021. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 10 (2021), 3736–3764.
- [12] Marlene Careil, Matthew J Muckley, Jakob Verbeek, and Stéphane Lathuilière. 2023. Towards image compression with perfect realism at ultra-low bitrates. In *The Twelfth International Conference on Learning Representations*.
- [13] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. 2020. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7939–7948.
- [14] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. 2020. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence* 44, 5 (2020), 2567–2581.
- [15] Jarek Duda. 2013. Asymmetric numeral systems: entropy coding combining speed of huffman coding with compression rate of arithmetic coding. *arXiv preprint arXiv:1311.2540* (2013).
- [16] Rudy Emrick, Pedro Cruz, Nuno B Carvalho, Steven Gao, Rudiger Quay, and Patrick Waltereit. 2014. The sky's the limit: Key technology and market trends in satellite communications. *IEEE Microwave Magazine* 15, 2 (2014), 65–78.
- [17] Patrick Esser, Robin Rombach, and Bjorn Ommer. 2021. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12873–12883.
- [18] Chuan Fu and Bo Du. 2023. Remote sensing image compression based on the multiple prior information. *Remote Sensing* 15, 8 (2023), 2211.
- [19] Haisheng Fu, Jie Liang, Zhenman Fang, Jingning Han, Feng Liang, and Guohe Zhang. 2024. Weconvene: Learned image compression with wavelet-domain convolution and entropy model. In *European Conference on Computer Vision*. Springer, 37–53.
- [20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.
- [21] Giorgia Guerrisi, Fabio Del Frate, and Giovanni Schiavon. 2023. Artificial intelligence based on-board image compression for the ϕ -sat-2 mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), 8063–8075.
- [22] Jinpei Guo, Yifei Ji, Zheng Chen, Kai Liu, Min Liu, Wang Rao, Wenbo Li, Yong Guo, and Yulun Zhang. 2025. OSCAR: One-Step Diffusion Codec Across Multiple Bit-rates. *arXiv preprint arXiv:2505.16091* (2025).
- [23] Ahmed Hanafi, Mohammed Karim, Ibtissam Latachi, Tajjeeddine Rachidi, Sanae Dahbi, and Smail Zouggar. 2017. FPGA-based secondary on-board computer system for low-earth-orbit nano-satellite. In *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 1–6.
- [24] Dailan He, Ziming Yang, Yuan Chen, Qi Zhang, Hongwei Qin, and Yan Wang. 2022. Post-training quantization for cross-platform learned image compression. *arXiv preprint arXiv:2202.07513* (2022).
- [25] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. 2022. Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5718–5727.
- [26] Dailan He, Yaoyan Zheng, Baocheng Sun, Yan Wang, and Hongwei Qin. 2021. Checkerboard context model for efficient learned image compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14771–14780.
- [27] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. 2021. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 5148–5157.
- [28] Chao Li, Shanzhi Yin, Chuanmin Jia, Fanyang Meng, Yonghong Tian, and Yongsheng Liang. 2024. Multirate progressive entropy model for learned image compression. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 8 (2024), 7725–7741.
- [29] Kaiyu Li, Ruixun Liu, Xiangyong Cao, Xueru Bai, Feng Zhou, Deyu Meng, and Zhi Wang. 2025. Segearth-ov: Towards training-free open-vocabulary segmentation for remote sensing images. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 10545–10556.
- [30] Yansheng Li, Yuning Wu, Gong Cheng, Chao Tao, Bo Dang, Yu Wang, Jiahao Zhang, Chuge Zhang, Yiting Liu, Xu Tang, et al. 2025. Meet: A million-scale dataset for fine-grained geospatial scene classification with zoom-free remote sensing imagery. *IEEE/CAA Journal of Automatica Sinica* 12, 5 (2025), 1004–1023.
- [31] Yuqi Li, Haotian Zhang, Li Li, and Dong Liu. 2025. Learned image compression with hierarchical progressive context modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 18834–18843.
- [32] Jinming Liu, Heming Sun, and Jiro Katto. 2023. Learned Image Compression with Mixed Transformer-CNN Architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1–10.
- [33] Qi Mao, Tinghan Yang, Yinyao Zhang, Zijian Wang, Meng Wang, Shiqi Wang, Libiao Jin, and Siwei Ma. 2024. Extreme image compression using fine-tuned vqgans. In *2024 Data Compression Conference (DCC)*. IEEE, 203–212.
- [34] Michael L. Pham, Christopher P. Artates. 2022. The Quickly Universally Integrated CubeSat. <https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=4662&context=smallsat>.
- [35] David Minnen, Johannes Ballé, and George D Toderici. 2018. Joint autoregressive and hierarchical priors for learned image compression. *Advances in neural information processing systems* 31 (2018).
- [36] Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. 2018. Pieapp: Perceptual image-error assessment through pairwise preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1808–1817.
- [37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.
- [38] Jorma Rissanen and Glen G Langdon. 1979. Arithmetic coding. *IBM Journal of research and development* 23, 2 (1979), 149–162.
- [39] Gerald A Soffen and Clayton W Snyder. 1976. The first Viking mission to Mars. *Science* 193, 4255 (1976), 759–766.
- [40] Spacety. 2022. Chaohu-1 SAR satellite. <https://www.satelliteevolution.com/post/two-new-satellites-launched-including-the-2nd-sar-satellite-with-improved-capabilities>.
- [41] David S Taubman, Michael W Marcellin, and Majid Rabbani. 2002. JPEG2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging* 11, 2 (2002), 286–287.
- [42] Aaron Van Den Oord, Oriol Vinyals, et al. 2017. Neural discrete representation learning. *Advances in neural information processing systems* 30 (2017).
- [43] Gregory K Wallace. 1991. The JPEG still picture compression standard. *Commun. ACM* 34, 4 (1991), 30–44.
- [44] Yichong Xia, Yimin Zhou, Jinpeng Wang, Baoyi An, Haoqian Wang, Yaowei Wang, and Bin Chen. [n. d.]. DiffPC: Diffusion-based High Perceptual Fidelity Image Compression with Semantic Refinement. In *The Thirteenth International Conference on Learning Representations*.
- [45] Shao Xiang and Qiaokang Liang. 2024. Remote sensing image compression based on high-frequency and low-frequency components. *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), 1–15.
- [46] Yixuan Ye, Ce Wang, Wanjie Sun, and Zhenzhong Chen. 2025. Map-Assisted remote-sensing image compression at extremely low bitrates. *ISPRS Journal of Photogrammetry and Remote Sensing* 223 (2025), 159–172.
- [47] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.
- [48] Ziyuan Zhang, Han Qiu, Maosen Zhang, Jun Liu, Bin Chen, Tianwei Zhang, and Hewu Li. 2024. Cosmic: Compress satellite images efficiently via diffusion compensation. *Proceedings of the NeurIPS, Vancouver, BC, Canada* (2024), 10–15.
- [49] Chuanxia Zheng and Andrea Veldai. 2023. Online clustered codebook. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 22798–22807.
- [50] Renjie Zou, Chunfeng Song, and Zhaoxiang Zhang. 2022. The devil is in the details: Window-based attention for image compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 17492–17501.