Semantic and Channel-Aware Image Transfer in Satellite Networks

Semantic and Channel-Aware Image Transfer in Satellite Networks

M. Palena[1], K. Ma[2], F. Penna[2], R. Fantini[3], C. Amatetti[4], B. De Filippo[4], A. Muhammad[4], C. F. Chiasserini[1,2]
[1]*CNIT*, Parma, Italy; [2]*Politecnico di Torino*, Torino, Italy; [3]*TIM*, Torino, Italy; [4]*University of Bologna*, Bologna, Italy

2026 IEEE 51st Conference on Local Computer Networks (LCN) | 979-8-3195-3148-3/26/$31.00 ©2026 IEEE | DOI: 10.1109/LCN67947.2026.11660769

***Abstract*—Non-Terrestrial Networks (NTNs) are expected to support future large-scale remote sensing, but transmitting high-resolution satellite imagery is challenged by limited feeder links, time-varying channels, and constrained onboard computation. We propose AITACS (Adaptive Image Transmission with Asymmetric Computation over Satellites), an end-to-end framework for satellite-ground image transmission. AITACS combines semantic feature extraction and Deep Joint Source Channel Coding (DJSCC) to transmit task-relevant latent representations instead of pixels. To address satellite-ground computational asymmetry, encoder stages are compressed via structured pruning and quantization, while more expressive decoders are retained at the receiver. A semantic-feedback mechanism adapts representation size and transmission rate to channel conditions. Evaluated on ship detection using a realistic feeder-link model, AITACS improves accuracy by up to 3.7× and reduces transmission rate by over 50% compared to conventional DJSCC. Transmitter complexity and memory usage drop by 74% and 95%, respectively, with minimal performance loss.**

***Index Terms*—Joint source and channel coding, Deep Neural Networks, Semantic communications**

I. Introduction

Satellite imaging systems play a critical role in a wide range of modern applications, including environmental monitoring, disaster management, urban planning, and security surveillance [1]. Advances in sensing technologies have enabled the acquisition of increasingly high-resolution imagery with improved temporal frequency, significantly enhancing the quality and utility of satellite data. In parallel, recent developments in machine learning have enabled more efficient and accurate processing of such data, supporting advanced analytics and intelligent decision-making [2], [3]. However, the rapid increase in image resolution and acquisition rates has led to a substantial growth in data volume, posing significant challenges for satellite communication systems. In practical scenarios, especially for satellites operating in Low Earth Orbit (LEO), data transmission is constrained by limited bandwidth, strict power budgets, and time-varying channel conditions [1]. These limitations often hinder the reliable and timely delivery of raw imagery to ground stations, where computational resources are typically more abundant. As a result, the effectiveness of downstream applications that rely on satellite imagery can be significantly impacted [4].

This work was partially supported by SNS JU under the EU's HE research and innovation programme under the MultiX project (Grant Agreement No. 101192521) and by TIM S.p.A. This work has been carried out in the framework of the SNS-JU-funded project 6G-NTN²: Nexus.

To address these challenges, semantic communication (SemCom) is a promising paradigm that prioritizes the transmission of meaningful information rather than exact bit-level accuracy [5]. By focusing on preserving the data content semantic, this approach can improve communication efficiency in resource-constrained conditions. Notably, Deep Joint Source Channel Coding (DJSCC) has gained considerable attention due to its ability to jointly optimize compression and transmission through end-to-end learning [6]. Indeed, unlike conventional communication systems that rely on separate source and channel coding, DJSCC learns robust representations that are inherently resilient to channel impairments.

Despite its potential, the deployment of DJSCC-based semantic communication systems in satellite environments is significantly challenged by rapidly varying channel conditions. In practical satellite communication scenarios, especially in LEO systems, channel quality can fluctuate due to factors such as atmospheric effects, mobility, Doppler shifts, and intermittent link availability [1]. These dynamics make it difficult to maintain reliable and consistent transmission performance, as the communication system must continuously adapt to changing signal-to-noise ratios (SNR) and link conditions. Such variability directly impacts both the fidelity of reconstructed imagery and the effectiveness of downstream tasks, as models trained under static or slowly varying assumptions may fail to generalize under highly dynamic environments. As a result, designing robust communication strategies that can maintain semantic integrity under rapidly changing channel conditions remains an open and critical challenge. A further challenge in satellite communication systems is the strong asymmetry in computational capabilities between satellites and ground stations. Onboard hardware is constrained by limited processing power, memory, and energy budgets, as well as by the need to operate reliably under harsh radiation conditions. In contrast, ground stations can leverage significantly greater computational resources. However, most existing DJSCC architectures rely on relatively symmetric encoder–decoder pipelines, with comparable neural network complexity at both transmitter and receiver [6]. While suitable for terrestrial systems, such designs are difficult to deploy in satellite scenarios, where the onboard encoder must satisfy stringent hardware constraints. This mismatch makes the practical adoption of DJSCC in satellite communications particularly challenging.

To address these challenges, we propose Adaptive Image Transmission with Asymmetric Computation over Satellites

(AITACS), an end-to-end SemCom framework that integrates: *(a)* a realistic channel model for the NTN feeder-link, *(b)* a DJSCC architecture with asymmetrical encoder-decoder complexities, and *(c) task-driven feedback* to dynamically adapt to time-varying channel conditions. Unlike existing approaches, AITACS adopts a closed-loop design in which the receiver assesses the utility of the reconstructed data using task-oriented metrics and feeds this information back to the transmitter. This allows the system to dynamically adapt both the semantic representation size and the transmission rate, prioritizing task-relevant features under time-varying channel conditions.

Our main contributions can be summarized as follows:

- *End-to-end semantic-aware DJSCC pipeline design:* We design an end-to-end architecture with a semantic encoder and a DJSCC in cascade, thus making the latter to process task-relevant features instead of pixel-level data. This allows for increased flexibility and adaptation capability. Further, we perform an end-to-end training of the pipeline using a loss function that accounts for both image reconstruction quality and image-domain perceptual quality.
- *Asymmetric encoder-decoder complexity:* We address the asymmetric computational capabilities of satellite and ground station systems by reducing the complexity of the onboard encoder via network quantization and pruning. This lowers the computational and memory requirements of transmitter-side processing while preserving the benefits of the SemCom pipeline under satellite hardware constraints.
- *Realistic feeder-link channel model*: We adopt a realistic NTN feeder-link channel model that captures the dominant line-of-sight propagation characteristics of satellite-ground station communications while accounting for atmospheric attenuation, shadowing, and scintillation.
- *Semantic feedback:* We introduce a semantic-level feedback mechanism based on the downstream task performance to guide transmission adaptation.

The rest of the paper is organized as follows. Sec. II discusses relevant related work and highlights the novelty of our study. Sec. III describes the reference scenario and the proposed semantic-aware DJSCC architecture. Sec. IV presents the experimental settings we used, shows the performance of our scheme, and compares it against state-of-the-art alternatives. Finally, Sec. V draws our conclusions.

II. Related Work

The seminal work in [6] demonstrated that convolutional neural networks (CNNs) can directly map images to channel inputs, significantly outperforming traditional separate source and channel coding schemes, especially in low Signal-to-Noise (SNR) and bandwidth-limited regimes. Subsequent works have extended this framework improving robustness, incorporating attention mechanisms and variable-rate transmission [7], [8].

More recently, the concept of SemCom has shifted the focus from accurate signal reconstruction to preservation of task-relevant information [9]. Several works have demonstrated that efficiency can be significantly increased integrating task-oriented objectives into the communication pipeline, allowing the system to prioritize semantically important features over pixel-level fidelity [10]–[12].

Further, adaptive transmission strategies have been designed to cope with time-varying wireless channels. Among these, traditional approaches rely on physical-layer metrics to adjust transmission parameters, while DJSCC enables adaptation by variable-rate encoding and bandwidth allocation [13]–[15]. However, these methods mainly depend on instantaneous channel knowledge and do not explicitly consider the impact of channel variations on downstream task performance.

Finally, feedback-based communications, widely studied to improve wireless transmissions reliability and efficiency, play an important role also in learning-based approaches for refining transmitted representations and enabling iterative transmission schemes [16]–[18]. Nevertheless, most existing approaches focus on signal-level or reconstruction-level feedback, rather than leveraging high-level semantic information.

Recent work in SemCom has studied the joint impact of channel quality and semantic compression on downstream task performance. In particular, [19] proposes a unified semantic loss model for DJSCC under Additive White Gaussian Noise (AWGN) and fading channels, demonstrating that metrics such as accuracy, pixel-wise fidelity, and structural similarity strongly depend on both SNR and compression ratio, and applying the model to communication-aware resource allocation for satellite Earth Observation systems.

Distinctiveness of our work. Existing DJSCC and SemCom approaches typically address robust transmission, rate adaptation, and task-oriented optimization separately, often under simplified channel models and without considering satellite computational constraints. In contrast, our work introduces an end-to-end framework for satellite image transmission that combines a realistic NTN feeder-link model, task-driven semantic adaptation, and an asymmetric encoder–decoder architecture based on encoder pruning and quantization.

III. Deep JSCC with Semantic-aware Adaptation

We consider a scenario in which a Low Earth Orbit (LEO) satellite equipped with an RGB camera transmits images to a ground station via a feeder link. The received data is processed for a downstream computer vision task (e.g., object detection), making task performance the primary objective rather than pixel-level reconstruction fidelity.

To cope with time-varying bandwidth and channel conditions, we propose AITACS (Adaptive Image Transmission with Asymmetric Computation over Satellites), an end-to-end framework that combines semantic communication, DJSCC, and feedback-driven adaptation (Fig. 1). The receiver evaluates reconstruction quality using task-oriented metrics and feeds this information back to the transmitter, which dynamically adjusts the semantic representation size and transmission rate. This closed-loop design enables efficient adaptation to channel variations while prioritizing task-relevant information.

The remainder of this section first describes the AITACS transmission pipeline, then presents the techniques used to accommodate the asymmetric computational capabilities of

the satellite–ground station system. Next, we introduce the semantic feedback and adaptation mechanism, followed by the training procedure for the overall framework.

A. End-to-End Pipeline: Basic Configuration

Transmitter side. Let $\mathbf{x} \in \mathbb{R}^{H\times W\times 3}$ be the input RGB image acquired by the satellite, where H and W correspond to the image height and width, respectively, and 3 is the color depth. The image is first processed by a semantic encoder $E_{\text{sem}}(\cdot)$, which projects the input into a compact latent representation. This stage serves both to reduce the spatial redundancy of the original image and to extract features that are most relevant for the downstream task, thereby facilitating more efficient and robust transmission through the subsequent DJSCC module. In our implementation, E_{sem} is realized as a lightweight CNN specifically designed for compact visual representation learning. More specifically, given the input image $\mathbf{x}$, the encoder processes the data through multiple convolutional layers combined with nonlinear activation functions and progressive downsampling stages. This progressively decreases the spatial resolution of the representation while increasing the level of semantic abstraction captured by the feature maps. The result is a compact latent semantic representation $\mathbf{s}$ given by

$$\mathbf{s} = E_{\text{sem}}(\mathbf{x};\boldsymbol{\theta}_{\text{sem}}), \quad \mathbf{s} \in \mathbb{R}^{H'\times W'\times C}, \tag{1}$$

where $\boldsymbol{\theta}_{\text{sem}}$ represents the set of trainable parameters of the semantic encoder. The dimensions H' and W' correspond to the reduced spatial resolution obtained after the downsampling operations, with $H' < H$ and $W' < W$, while C denotes the number of feature channels in the latent representation. Each channel captures different high-level semantic characteristics of the input image, enabling the extraction of increasingly abstract and task-relevant features across the network depth. The semantic encoder *maps the input image into a task-relevant feature space*, which typically also yields a compressed representation, i.e., $H'W'C < 3HW$. The semantic representation $\mathbf{s}$ is then transformed into a sequence of K complex-valued channel symbols through a DJSCC encoder $E_{\text{ch}}(\cdot)$, namely

$$\mathbf{z} = E_{\text{ch}}(\mathbf{s};\boldsymbol{\theta}_{\text{ch}}), \quad \mathbf{z} \in \mathbb{C}^{K}, \tag{2}$$

where $\boldsymbol{\theta}_{\text{ch}}$ denotes the trainable parameters of the DJSCC encoder and K is the number of transmitted symbols associated with each image. The role of $E_{\text{ch}}(\cdot)$ is to jointly perform source and channel coding by mapping the semantic features into a *compressed representation that is robust to channel impairments*. As such, we typically assume the DJSCC output size to be significantly smaller than that of the original image, i.e., $K \ll 3HW$.

The adopted DJSCC architecture builds upon the convolutional design proposed in [6], with the main modification that the encoder input is the semantic feature map $\mathbf{s}$ instead of the original image. The subsequent layers preserve the original structure and progressively generate a channel-adapted latent representation suitable for transmission over the wireless link. To satisfy the transmit power constraint, the encoded symbols are normalized to obtain $\tilde{\mathbf{z}} \in \mathbb{C}^K$ such that $\mathbb{E}\big[\|\tilde{\mathbf{z}}\|^2\big] = KP$, where P denotes the average transmit power.

Wireless channel. The normalized symbols $\tilde{\mathbf{z}}$ are mapped onto distinct time and/or frequency resources (e.g., OFDM symbols or subcarriers) and transmitted over a stochastic wireless channel affected by additive noise. The received signal is modeled as

$$\hat{\mathbf{z}} = \tilde{\mathbf{z}} + \mathbf{n}, \tag{3}$$

where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I}_K)$ is a vector of i.i.d. complex Gaussian noise samples with average power σ_n^2. For the NTN feeder-link scenario considered in this work, a realistic channel model involves a strong line-of-sight (LOS) component, with limited frequency selectivity due to multipath reflections. Examples include the NTN clustered delay line (CDL) models defined in 3GPP TR 38.811, such as NTN-CDL-C and NTN-CDL-D, where the first reflected path is approximately 10 dB weaker than the LOS path. For our practical application, we can effectively neglect the multipath reflections, and simply approximate the channel as frequency-flat. This motivates the AWGN model assumed.

Moreover, we do not explicitly model Doppler-induced temporal correlation. This assumption is justified by the feeder-link setting, where the ground station is fixed and the satellite trajectory is known, enabling effective Doppler compensation. In addition, consecutive observations in our scenario are separated by time intervals much larger than the channel coherence time, making different channel realizations effectively uncorrelated. To account for large-scale channel impairments such as atmospheric attenuation, shadowing, and scintillation, the noise variance σ_n^2 is assumed to vary across observations according to a log-normal distribution. Specifically, for each transmission, the instantaneous SNR in dB is sampled according to $SNR_{\text{dB}} \sim \mathcal{N}(\mu_{\text{SNR}}, \sigma^2_{\text{SNR}})$, where μ_{SNR} denotes the nominal operating SNR determined by the link budget, and σ^2_{SNR} models slow large-scale fluctuations of the channel conditions. The corresponding noise variance is then computed as $\sigma_n^2 = 10^{-SNR_{\text{dB}}/10}$. Furthermore, based on the ray-tracing-based evaluations reported in [20], the average K-factor can be over 50 dB at even 10° of elevation angle in Ka-, Q-, and V-band in suburban areas (a typical setting for NTN ground stations). This formulation preserves the simplicity of the AWGN model while incorporating realistic slow variations of the feeder-link channel quality.

Receiver side. At the receiver side, the noisy signal $\hat{\mathbf{z}}$ is first processed by the DJSCC decoder $D_{\text{ch}}(\cdot)$ to recover an estimate of the transmitted semantic representation:

$$\hat{\mathbf{s}} = D_{\text{ch}}(\hat{\mathbf{z}};\boldsymbol{\phi}_{\text{ch}}), \tag{4}$$

where $\boldsymbol{\phi}_{\text{ch}}$ denotes the trainable parameters of the decoder. The decoder adopts a convolutional architecture that mirrors the structure of the DJSCC encoder and is jointly optimized with it to compensate for both compression losses and channel distortions [6]. Through this process, the receiver can reconstruct a semantically consistent approximation of the transmitted feature representation even under adverse channel conditions.

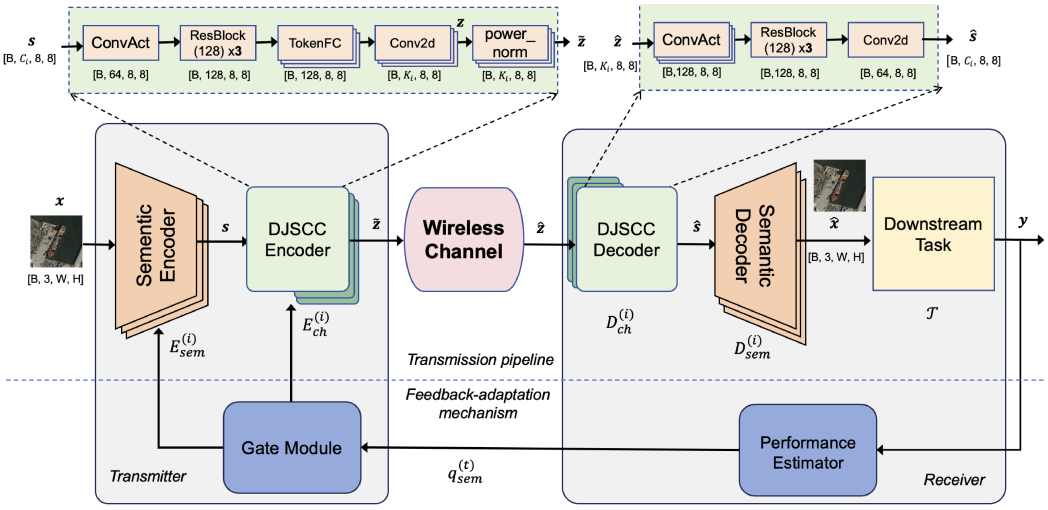


Fig. 1: Scheme of the AITACS framework. B is the batch size used during training (B=1 for inference). Top: internal architecture of a single branch of the DJSCC encoder/decoder, with C_i being the no. of feature channels in the semantic representation and K_i the no. of symbols produced by the encoder. The stacked boxes beneath the orange and green modules denote the different semantic/DJSCC encoder-decoder branches. For the DJSCC encoder/decoder some of the layers are shared across branches.

The recovered semantic representation $\hat{\mathbf{s}}$ is then fed to the semantic decoder, which reconstructs the image in the pixel domain according to

$$\hat{\mathbf{x}} = D_{\text{sem}}(\hat{\mathbf{s}}; \boldsymbol{\phi}_{\text{sem}}), \tag{5}$$

where $\boldsymbol{\phi}_{\text{sem}}$ denotes the parameters of the semantic decoder. The decoder is implemented as a CNN with a structure complementary to that of the semantic encoder, progressively increasing the spatial resolution and refining the feature maps to produce the reconstructed image $\hat{\mathbf{x}}$.

B. Semantic-aware Feedback and Adaptation

To improve robustness under time-varying bandwidth and channel conditions, we augment the framework with a semantic-level feedback mechanism that dynamically adapts the transmission process according to the performance of the downstream task. Unlike conventional schemes based solely on physical-layer metrics, the proposed approach leverages application-level feedback to balance communication efficiency and task performance. Although consecutive channel realizations are modeled as statistically independent, consecutive satellite images are typically semantically correlated, making the semantic feedback from one transmission informative for subsequent ones.

Feedback generation at the receiver. Semantic feedback is generated at the receiver for each reconstructed image. After image reconstruction, the received sample $\hat{\mathbf{x}}$ is processed by a downstream task in order to assess the quality of the recovered semantic information. Without loss of generality, we consider object detection as the target task, given its relevance in many satellite imaging applications. Let $\mathcal{T}(\cdot)$ denote the object detection model employed at the receiver. The performance of the reconstructed image can then be quantified using standard task-oriented metrics such as Intersection over Union (IoU) or mean Average Precision (mAP). These metrics provide a quantitative measure of how well the image semantic content is preserved with respect to the task objective. Based on these indicators, a scalar semantic quality score $q_{\text{sem}}^{(t)}$ is computed for the image transmitted at time instant t through a performance estimator module $\mathcal{Q}(\cdot)$:

$$q_{\text{sem}}^{(t)} = \mathcal{Q}(\hat{\mathbf{x}}^{(t)}), \tag{6}$$

where $\mathcal{Q}(\cdot)$ represents the selected task-oriented performance metric.The resulting semantic feedback is returned to the transmitter and used to adapt the encoding strategy for subsequent transmissions. The receiver also estimates the instantaneous SNR from the reported Channel Quality Indicator (CQI) and feeds it back together with the semantic feedback.

Adaptation mechanism. After receiving the feedback, the transmitter stores the most recent L semantic quality measurements, namely $\{q_{\text{sem}}^{(t-z)}\}_{z=0}^{L-1}$, within a sliding observation window. This temporal aggregation mitigates short-term fluctuations and provides a more stable estimate of the downstream task performance under dynamic channel conditions. An aggregated semantic quality indicator is then computed as $\bar{q}_{\text{sem}}^{(t)} = \frac{1}{L}\sum_{z=0}^{L-1} q_{\text{sem}}^{(t-z)}$ (alternative aggregation strategies, such as filtering, could also be adopted).

To support adaptive operation, we define a set $\mathcal{M}$ of parallel transmission-reception branches, each associated with a different point in the trade-off between transmission rate and task performance. Each branch $i \in \mathcal{M}$ includes:

- a semantic encoder-decoder pair $\left(E_{\text{sem}}^{(i)}, D_{\text{sem}}^{(i)}\right)$ characterized by a latent representation with C_i feature channels;
- a corresponding DJSCC encoder-decoder pair $\left(E_{\text{ch}}^{(i)}, D_{\text{ch}}^{(i)}\right)$ generating K_i transmitted channel symbols.

Increasing either the number of the semantic feature channels C_i or the transmitted symbols K_i generally improves the downstream task performance, with the latter incurring in higher bandwidth consumption. We consider $|\mathcal{M}| = 3$ operating

TABLE I: Configuration of the $|\mathcal{M}|$ transmission-reception branches; the size is expressed in number of parameters

Branch id (i)	Description	C_i	K_i	Size (Tx)	Size (Rx)
1	AITACS-small	16	16	1.54M	1.56M
2	AITACS-medium	32	32	3.48M	3.53M
3	AITACS-large	64	64	6.27M	6.40M

modes, corresponding to *small*, *medium*, and *large* semantic representations. The specific values of C_i and K_i adopted for each branch are summarized in Table I.

Fig. 2 illustrates representative reconstructed images obtained for different transmission branches (columns) and channel SNR values (rows). As expected, increasing the semantic representation size C_i, the number of transmitted symbols K_i, or the channel SNR leads to higher reconstruction fidelity and improved downstream detection performance, as reflected by more accurate object localization. The detailed configuration parameters for the small, medium, and large operating modes are provided in Sec. IV.

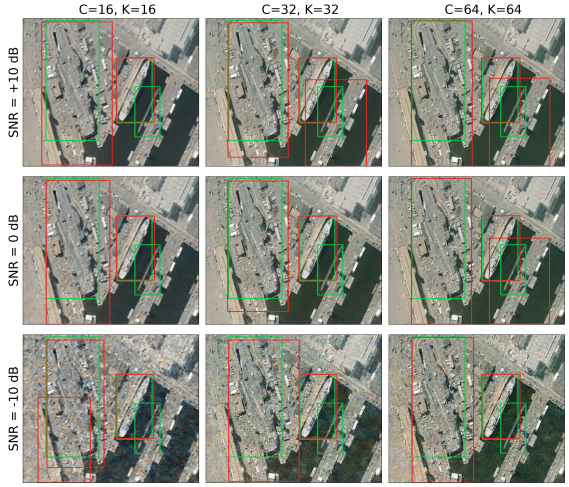


Fig. 2: Examples of reconstructed images and corresponding detection results under different transmission branches (columns) and SNR conditions (rows). Green bounding boxes denote the ground-truth objects, while red bounding boxes indicate the detection results.

At the transmitter, branch selection is performed by a lightweight *gate* module implementing a threshold-based adaptation strategy driven by the aggregated semantic feedback. Given the quality indicator $\bar{q}_{\text{sem}}^{(t)}$ computed at time t, the gate determines the transmission branch to be employed for the subsequent image transmission at time $t+1$. Specifically, two thresholds τ_1 and τ_2, with $\tau_1 < \tau_2$, are introduced, and the selected branch index i^* is defined as

$$i^* = \begin{cases} 1, & \bar{q}_{\text{sem}}^{(t)} < \tau_1, \\ 2, & \tau_1 \le \bar{q}_{\text{sem}}^{(t)} < \tau_2, \\ 3, & \bar{q}_{\text{sem}}^{(t)} \ge \tau_2. \end{cases} \tag{7}$$

Values of $\tau_1 = 0.55$ and $\tau_2 = 0.75$ are empirically selected. In practice, when the aggregated semantic quality degrades (e.g., due to harsher channel conditions), the system switches to a more robust configuration with higher C_i and K_i. Conversely, when performance is consistently satisfactory, a more compact representation is selected to improve transmission efficiency. This low-complexity mechanism enables the system to *dynamically adjust semantic compression and symbol rate based on a smoothed estimate of task performance.*

Finally, although AITACS is compatible with alternative semantic backbones, including Transformer-based ones, the evaluation in this work uses a convolutional encoder–decoder, which provides a favorable trade-off between reconstruction quality, computational burden, and training stability.

C. Managing Asymmetric Computational Capabilities

A key challenge in the considered satellite–to-ground station scenario is the asymmetry between transmitter- and receiver-side computational resources. To address this, we adopt an asymmetric design in which the onboard DNN encoders are compressed through structured pruning and post-training quantization, while the receiver-side decoders remain uncompressed. Specifically, compression is applied to both the semantic encoder $E_{\text{sem}}(\cdot)$ and the DJSCC encoder $E_{\text{ch}}(\cdot)$, as these modules are executed onboard the satellite.

Encoder pruning. We reduce the complexity of the transmitter-side models through structured channel pruning [21]. Rather than designing extremely small encoders from scratch, we first train the complete end-to-end pipeline and subsequently remove redundant encoder parameters. This strategy enables the network to exploit a larger representational capacity during training while achieving lower inference complexity with limited performance degradation.

To control the compression level, we introduce a pruning ratio $R \in [0, 1)$, representing the fraction of channels removed from each prunable layer. Let $\mathbf{W}^{(\ell)} \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}} \times k \times k}$ denote the convolutional kernel tensor at layer ℓ. For each output channel c, an importance score is computed as $\Gamma_c^{(\ell)} = \left\| \mathbf{W}_c^{(\ell)} \right\|_1$, where $\mathbf{W}_c^{(\ell)}$ denotes the weights associated with the c-th output channel. The channels are then ranked according to their importance scores, and the fraction R of channels with the lowest scores is removed together with the corresponding feature maps. Larger values of R therefore produce more aggressive compression and lower computational complexity, at the expense of a potentially higher performance degradation.

After pruning, the compressed encoder is fine-tuned jointly with the remaining network components in order to recover part of the performance loss induced by parameter removal. Structured pruning is particularly suitable for onboard deployment since it preserves dense tensor operations and avoids the overhead associated with sparse computation.

Model quantization. To further reduce memory footprint and computational complexity, we apply post-training quantization to the encoder modules. Specifically, floating-point weights and activations are mapped to low-precision fixed-point representations. Let w denote a floating-point parameter and let b be the number of quantization bits. Assuming a uniform quantizer over the dynamic range $[w_{\min}, w_{\max}]$, the quantized representation $\tilde{w}$ is obtained as

$$\tilde{w} = \text{round}\left(\frac{w - w_{\min}}{w_{\max} - w_{\min}} (2^b - 1) \right), \tag{8}$$

where 2^b represents the number of quantization levels. Increasing b provides finer quantization resolution and lower approximation error, while smaller values of b enable greater reductions in memory occupancy and arithmetic complexity. By combining structured pruning and quantization, AITACS achieves a lightweight transmitter-side implementation while preserving the benefits of the semantic-aware DJSCC pipeline.

D. Training the AITACS Pipeline

The different pipeline branches are trained independently in an offline fashion. For each branch configuration $i \in \mathcal{M}$, the semantic encoder-decoder pair $\left(E_{\text{sem}}^{(i)}, D_{\text{sem}}^{(i)}\right)$ is jointly trained together with the corresponding DJSCC encoder-decoder pair $\left(E_{\text{ch}}^{(i)}, D_{\text{ch}}^{(i)}\right)$. Training is conducted using a nominal operating SNR, μ_{SNR}, ranging from $-10\,\text{dB}$ to $10\,\text{dB}$. The overall training objective combines latent-domain consistency, pixel-level reconstruction accuracy, and perceptual image quality. Specifically, the loss function is defined as

$$\mathcal{L} = \lambda_{\text{lat}}\mathcal{L}_{\text{lat}} + \lambda_{\text{mse}}\mathcal{L}_{\text{mse}} + \lambda_{\text{lpips}}\mathcal{L}_{\text{lpips}}, \quad \text{where} \tag{9}$$

$$\mathcal{L}_{\text{lat}} = \|\hat{\mathbf{z}} - \mathbf{z}\|^2, \qquad \mathcal{L}_{\text{mse}} = \|\hat{\mathbf{x}} - \mathbf{x}\|^2, \tag{10}$$

and $\mathcal{L}_{\text{lpips}}$ denotes the Learned Perceptual Image Patch Similarity (LPIPS) metric [22] computed between the original and reconstructed images. Unlike conventional pixel-wise losses, LPIPS evaluates perceptual similarity in a deep feature space extracted from a pretrained convolutional network; here LPIPS is computed using the AlexNet backbone [22]. More precisely, the metric compares feature activations across multiple network layers using a weighted distance measure, yielding a perceptual assessment that correlates more closely with human visual quality than purely pixel-domain criteria.

In all experiments, the loss weights are set to $\lambda_{\text{lat}} = 1.0$, $\lambda_{\text{mse}} = 0.2$, and $\lambda_{\text{lpips}} = 0.05$. The complete framework is trained end-to-end using the AdamW optimizer with a learning rate of 2×10^{-4} and a weight decay factor of 10^{-4}. Training is carried out for 20 epochs using a batch size of $B = 2$. The models are trained on the ShipRSImageNet ship detection dataset [23], composed of 2,198 training images, 550 validation images and 687 test images.

IV. Experimental Results

Experimental Setup. We evaluate the proposed framework on a satellite image transmission scenario with ship detection as the target downstream task. Semantic quality is assessed using a YOLOv12 [24] detector applied to the reconstructed images, measuring how effectively task-relevant information is preserved through the communication pipeline.

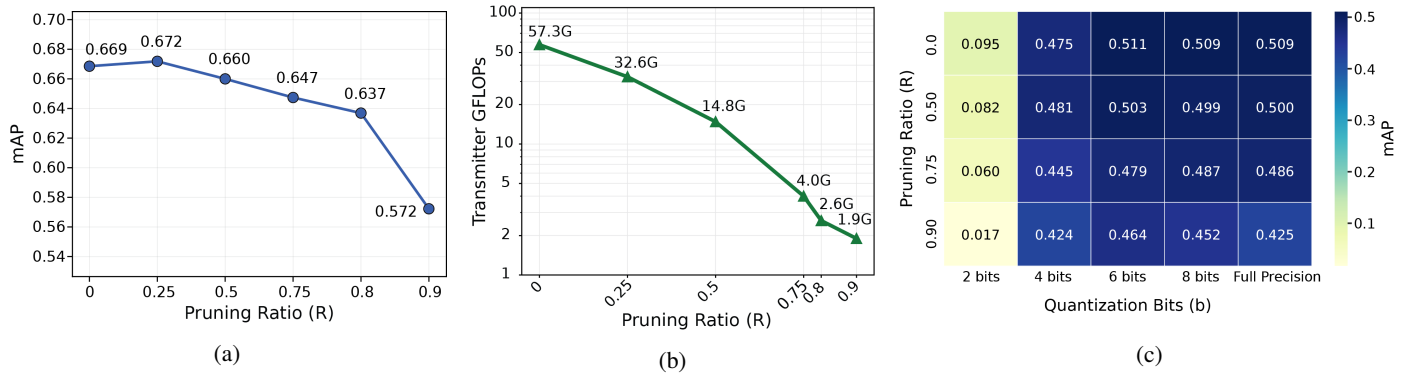
Detection performance is quantified using mean Average Precision (mAP) following the COCO protocol, i.e., AP averaged over Intersection-over-Union (IoU) thresholds from 0.50 to 0.95 in steps of 0.05 ($\text{AP}_{50:95}$). All experiments are run on an NVIDIA L40S GPU with 46 GB of memory.

Impact of pruning and quantization. To evaluate the asymmetric design introduced in Sec. III-C, we investigate the impact of encoder pruning and quantization on task performance. Specifically, for each branch, we assess the effect of the pruning ratio R and quantization precision b on downstream detection accuracy, model size, and computational cost, and then select an operating point based on a heuristic trade-off among these metrics. All results are obtained on the ShipRSImageNet validation set under a fixed channel condition of $\text{SNR} = 10$,dB. For brevity, we report results only for the *AITACS-medium* branch, as the other configurations exhibit similar behavior.

Figure 3a shows the downstream detection performance as a function of the pruning ratio R, while Fig. 3b reports the corresponding transmitter-side computational complexity in GFLOPs. Interestingly, light pruning ($R = 0.25$) preserves performance while reducing the computational cost by approximately 43%. This behavior can be attributed to a mild regularization effect arising from the removal of less informative channels during fine-tuning. A particularly desirable operating point is obtained with moderate pruning ($R = 0.50$) where the computational burden is reduced by approximately 74% with respect to the unpruned model, while the decrease in mAP remains below one percentage point. This result indicates that a large fraction of the encoder complexity can be removed without significantly affecting the preservation of task-relevant semantic information. However, as the pruning ratio increases beyond $R = 0.75$, the reduction in representational capacity becomes increasingly detrimental, leading to larger losses in detection performance. To further reduce onboard resource consumption, we combine structured pruning with post-training quantization. Figure 3c shows the resulting mAP for different combinations of pruning ratio R and quantization bit-width b. The results indicate that 6-bit and 8-bit quantization are essentially lossless across all considered pruning levels. In some cases, a slight performance gain is observed, which can be attributed to the regularization effect of quantization noise and the inherent variability of the evaluation metric. These findings highlight the robustness of the learned semantic representations to moderate quantization noise. In contrast, 4-bit quantization leads to a noticeable drop in detection accuracy, especially under aggressive pruning, while 2-bit quantization results in a severe performance degradation.

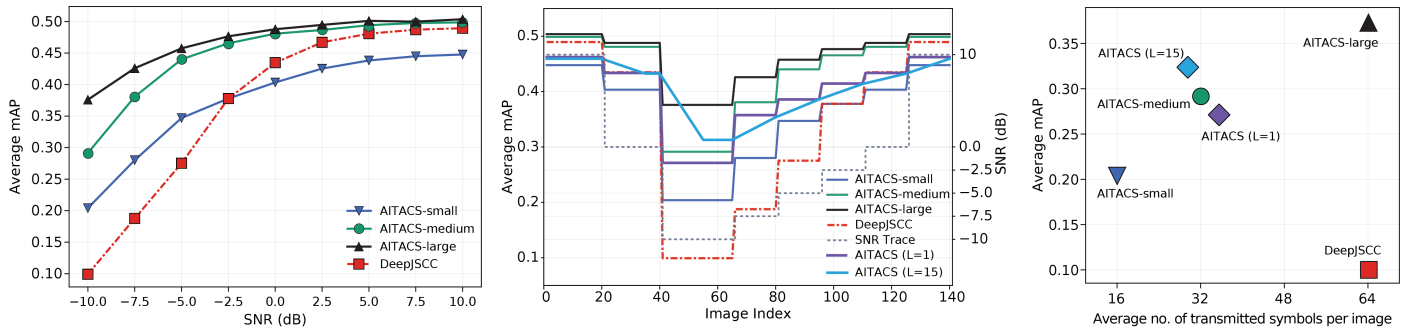
Overall, the results show substantial reductions in transmitter-side complexity yielded by our asymmetric design strategy. Moderate pruning combined with 6–8 bit quantization emerges as the most favorable operating region, offering significant savings in computational cost and memory footprint while preserving almost unchanged downstream task performance. Applying the selected compression configuration ($R = 0.5$, $b = 6$ bits) yields substantial transmitter-side savings across all AITACS branches, reducing computational complexity by approximately 74% in terms of GFLOPs and memory footprint by approximately 95% relative to the corresponding uncompressed encoders.

Baseline. We compare the proposed method with a pixel-based, purely DeepJSCC baseline. Specifically, we implement the DeepJSCC framework described in [6], using a fixed



(a) (b) (c)

Fig. 3: Impact of structured pruning and post-training quantization: (a) Downstream detection performance vs. the pruning ratio; (b) Transmitter-side computational complexity (GFLOPs); (c) Joint effect of pruning and quantization on detection performance.



(a) (b) (c)

Fig. 4: Performance evaluation of the AITACS framework. (a) Average mAP vs. SNR for the fixed AITACS branches and the DJSCC baseline. (b) mAP under time-varying channel conditions, comparing fixed branches (solid lines), adaptive variants (dashed lines) and the baseline (dash-dotted) over an average SNR trace (dotted line). (c) Average mAP vs. average number of transmitted symbols per image at SNR$= -10$ dB.

number of transmitted symbols equal to that of the largest AITACS branch ($K = 64$). The baseline is trained on the same dataset and under the same channel models and SNR range as AITACS. To ensure a fair comparison, the baseline encoder also undergoes pruning and quantization as described above.

Performance Under Static Channel Conditions.

We then evaluate the individual, non-adaptive AITACS operating modes under static channel conditions in order to quantify the trade-off among semantic representation size, transmission resources, and downstream task performance. Fig. 4a reports the average mAP as a function of the channel SNR. The analysis considers the three predefined branches of AITACS, namely *AITACS-small*, *AITACS-medium*, and *AITACS-large* and compares them against a conventional DJSCC baseline.

Several observations can be drawn from the results. First, increasing the branch size consistently improves detection performance across the entire SNR range, indicating that richer semantic representations and larger symbol budgets enable more robust information delivery. Second, the benefit of larger branches is particularly evident under challenging channel conditions, where the performance gap between the most compact and the largest configuration becomes substantial. As channel quality improves, however, the performance gap among the different branches gradually reduces, suggesting that the advantage of allocating additional semantic and transmission resources diminishes at high SNR. Furthermore, the proposed semantic-aware architecture generally achieves higher mAP than the pixel-oriented DeepJSCC baseline, with the only notable exception being the *AITACS-small* configuration in the SNR region above -2.5 dB. Overall, these findings highlight the effectiveness of task-oriented semantic transmission and reveal a clear performance–efficiency trade-off that can be exploited through adaptive branch selection.

Adaptive Transmission Under Time-Varying Channels. We next evaluate AITACS in a dynamic channel environment to assess the effectiveness of semantic feedback for online adaptation. To this end, we consider a predefined SNR trace in which channel quality gradually deteriorates from 10 dB to -10 dB before progressively recovering. Figure 4b compares the performance of the three fixed AITACS branches with that of two adaptive configurations using feedback window lengths of $L = 1$ and $L = 15$, as well as the baseline. The results show that the proposed feedback-driven adaptation mechanism effectively tracks channel variations and maintains higher task performance than static operating modes, particularly under adverse channel conditions. The adaptive configurations outperform the DJSCC baseline across most of the SNR trajectory and provide a substantial gain over *AITACS-small* during the low-SNR phase, while converging to similar performance as channel quality improves. The configuration with a longer feedback window ($L = 15$) also surpasses *AITACS-medium* under the most severe channel conditions, although this advantage grad-

ually diminishes with increasing SNR. Overall, these findings confirm that the adaptation mechanism enables AITACS to respond effectively to channel fluctuations.

Performance-rate Trade-off Analysis. Finally, to illustrate the performance–rate trade-offs offered by AITACS, we compare the adaptive and non-adaptive configurations with the baseline in terms of downstream task precision and the average number of transmitted symbols per image under a severely impaired channel (SNR $= -10$ dB), as shown in Fig. 4c.

The baseline exhibits the least favorable performance–rate trade-off, requiring 64 symbols per image while achieving an average precision of only 0.1. The fixed-branch variants of AITACS significantly improve this trade-off, delivering higher task performance while using equal or fewer transmission resources. As expected, larger model branches achieve better performance at the cost of a higher transmission rate. In particular, *AITACS-large* achieves the best performance (0.37), corresponding to a $3.7\times$ improvement over the baseline while using the same number of transmitted symbols. Smaller variants offer more efficient operating points: *AITACS-medium* and *AITACS-small* reduce the symbol rate by 50% and 75%, respectively, while still improving performance by $2.9\times$ and $2\times$ over the baseline, highlighting the benefits of semantic-aware transmission even at significantly lower rates.

Among adaptive variants, *AITACS (L=15)* consistently outperforms *AITACS (L=1)* in both task performance and communication cost. In terms of transmission efficiency, *AITACS (L=15)* reduces the average rate by 56.3% relative to the baseline, compared to 43.7% for *AITACS (L=1)*. A similar trend is observed for task performance, where *AITACS (L=15)* improves precision by $3.3\times$, compared to $2.6\times$for *AITACS (L=1)*. Overall, *AITACS (L=15)* provides the best trade-off.

Finally, as the feedback window size decreases, the average number of transmitted symbols increases and task performance slightly degrades. This indicates that shorter windows yield less stable channel estimates, leading the policy to adopt more conservative (higher-rate) transmission decisions, whereas larger windows enable more reliable adaptation.

V. Conclusions

This paper proposed AITACS, an adaptive semantic communication framework for image transmission over NTN feeder links. Combining semantic feature extraction, DJSCC, task-oriented feedback, and asymmetric encoder–decoder architectures, AITACS adapts both semantic representation and transmission rate to varying channel conditions while meeting satellite computational constraints. Results show that AITACS improves downstream detection accuracy by up to $3.7\times$ while reducing the average number of transmitted symbols by more than 50% compared to a conventional DJSCC baseline. Furthermore, transmitter-side pruning and quantization reduce computational complexity and memory footprint by over 74% and 95%, respectively, with minimal performance degradation.

References

[1] I. Leyva-Mayorga and a. et "Satellite edge computing for real-time and very-high resolution earth observation," *IEEE ToC*, 2023.

[2] G. Fontanesi and a. et "Artificial intelligence for satellite communication and non-terrestrial networks: A survey," 2023.

[3] M. Affek and J. Szymański, "A survey on the datasets and algorithms for satellite data applications," *IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[4] H.-F. Chou, S. Solanki, V. N. Ha, L. Chen, S. L. Ma, H. Al-Hraishawi, G. Eappen, and S. Chatzinotas, "Edge ai empowered physical layer security for 6g ntn: Potential threats and future opportunities," *ArXiv*, vol. abs/2401.01005, 2023.

[5] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Comm. Surveys & Tutorials*, 2023.

[6] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE TCCN*, 2019.

[7] M. Yang and H.-S. Kim, "Deep Joint Source-Channel Coding for Wireless Image Transmission with Adaptive Rate Control," in *IEEE ICASSP*, 2022.

[8] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gündüz, "Deep joint source-channel coding for semantic communications," *IEEE Comm. Mag.*, 2023.

[9] Y. Liu, X. Wang, Z. Ning, M. Zhou, L. Guo, and B. Jedari, "A survey on semantic communications: Technologies, solutions, applications and challenges," *Digital Communications and Networks*, 2024.

[10] J. Wu, C. Wu, Y. Lin, T. Yoshinaga, L. Zhong, X. Chen, and Y. Ji, "Semantic segmentation-based semantic communication system for image transmission," *Digital Communications and Networks*, 2024.

[11] N. Li and Y. Deng, "Goal-oriented semantic communication for wireless image transmission via stable diffusion," in *IEEE ICC*, 2025.

[12] M. Palena, J. A. Ayala-Romero, A. Garcia-Saavedra, and C. F. Chiasserini, "SPIFF: Selective preservation of image fidelity for bandwidth-constrained heterogeneous networks," in *IEEE INFOCOM*, 2026.

[13] D. B. Kurka and D. Gündüz, "Bandwidth-agile deep joint source-channel coding for wireless image transmission," *IEEE ToC*, 2020.

[14] H. Wu, Y. Shao, C. Bian, K. Mikolajczyk, and D. Gündüz, "Deep Joint Source-Channel Coding for Adaptive Image Transmission Over MIMO Channels," *IEEE TWC*, 2024.

[15] M. Xu, C.-T. Lam, Y. Liang, T. Qiu, B. K. Ng, and S.-K. Im, "MVJSCC: Adaptive lightweight DeepJSCC for semantic image transmission," *IEEE Wireless Comm. Lett.*, 2025.

[16] D. B. Kurka and D. Gündüz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE J. on Selected Areas in Information Theory*, 2020.

[17] H. Wu, Y. Shao, E. Ozfatura, K. Mikolajczyk, and D. Gündüz, "Transformer-aided wireless image transmission with channel feedback," *IEEE TWC*, 2024.

[18] H. Li, G. Zhang, K. Zhou, Y. Cai, and G. Yu, "Coarse-to-fine: A dual-phase channel-adaptive method for wireless image transmission," *IEEE TWC*, 2026.

[19] T. T. Nguyen, T.-D. Le, V. N. Ha, D.-D. Tran, H. Nguyen-Kha, D.-H. Tran, C. L. Marcos-Rojas, J. C. Merlano-Duncan, and S. Chatzinotas, "Toward a unified semantic loss model for deep JSCC-based transmission of EO imagery," 2026.

[20] 6G-NTN project, "Report on unified and data-driven air-interface for 6G-NTN," Deliverable D4.5, Dec. 2025.

[21] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural network with pruning, trained quantization and huffman coding," *arXiv: Computer Vision and Pattern Recognition*, 2015.

[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE CVPR*, 2018.

[23] Z. Zhang, L. Zhang, Y. Wang, P. Feng, and R. He, "ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021.

[24] Y. Tian, Q. Ye, and D. Doermann, "YOLO12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, 2025.