POLITECNICO DI TORINO
Repository ISTITUZIONALE

Goal-Oriented Communication With Adaptive Semantic Representation Selection for Satellite Image Classification

Original
Goal-Oriented Communication With Adaptive Semantic Representation Selection for Satellite Image Classification / Muhammad, A., De Filippo, B., Amatetti, C., Palena, M., Penna, F., Chiasserini, C.F., Vanelli-Coralli, A.. - (2026). (2026 IEEE Future Networks World Forum (FNWF) Lisbon (Por) 26–28 October 2026).

Availability:
This version is available at: 11583/3015749 since: 2026-09-21T09:00:12Z

Publisher:
IEEE

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 November 2026

Goal-Oriented Communication With Adaptive Semantic Representation Selection for Satellite Image Classification

A. Muhammad[1], B. De Filippo[1], C. Amatetti[1], M. Palena[2], F. Penna[3], C. F. Chiasserini[2,3], A. Vanelli-Coralli[1]
[1]*University of Bologna*, Bologna, Italy; [2]*CNIT*, Parma, Italy; [3]*Politecnico di Torino*, Torino, Italy;

***Abstract*—Earth observation satellites continuously capture Earth images, generating large volumes of visual data that must be transmitted over bandwidth-limited satellite-to-ground links for remote inference. Existing deep joint source-channel coding (DJSCC) methods optimize transmission for image reconstruction, requiring complex models and significant channel resources to preserve content that is largely irrelevant to semantic inference tasks such as image classification. This paper proposes an end-to-end goal-oriented semantic communication framework for image classification over a satellite-to-ground erasure channel, where the semantic encoder, erasure-adaptive feature selection gate and task decoder are jointly optimized using only the classification objective, transmitting task-relevant latent representations rather than aiming at reconstructing the original image. The encoder conditions semantic feature extraction on the estimated erasure probability via a multi-head self-attention module, while an adaptive gate dynamically determines the number of transmitted dimensions through learned importance scores, without any predefined transmission target. Experiments on CIFAR-10 demonstrate classification accuracies of 86.32% - 89.29% over the erasure probability $\epsilon \in [0, 0.8]$, while transmitting only 2.5% - 14.4% of the 1024 latent dimensions, consistently outperforming the baseline across the evaluated erasure range.**

***Index Terms*—Semantic Communication, Goal-Oriented Communication, Deep Joint Source-Channel Coding (DJSCC).**

I. Introduction

Traditional wireless communication systems were developed based on Shannon's information theory, the primary objective of which is the reliable transmission of bits over a noisy channel [1]. However, with the increasing deployment of non-terrestrial network (NTN) platforms, such as low-earth orbit (LEO) satellites and unmanned aerial vehicles (UAVs), for communication as well as Earth observation, remote sensing, and disaster monitoring, communication systems operating over these links are required to support not only reliable data delivery but also semantic information exchange that enables intelligent decision making at the receiver. In these applications, edge devices continuously capture visual data and transmit it over wireless links to a remote server for semantic inference, such as image classification [2]. Since wireless bandwidth and power resources are limited, the communication link becomes a critical bottleneck in conveying task-relevant semantic information, directly affecting inference accuracy. This has led to growing interest in goal-oriented communication (GOC) and semantic communications [2], [3], which aim to transmit only the information necessary for a specific task rather than reconstructing the original image [4], [5].

Recent advancements in deep learning have driven significant progress through deep joint source-channel coding (DJSCC), which jointly optimizes source compression and channel robustness in end-to-end trained neural networks [6]. Several researchers have extended this framework to handle varying channel conditions. The authors of [7] proposed the signal-to-noise Ratio (SNR)-adaptive DJSCC framework, where the decoder estimates the channel SNR and uses it to improve image reconstruction quality across different noise levels. Xu *et al.* [8] introduced SNR-conditioned attention blocks in the encoder that adaptively re-weight features based on the current channel state, improving peak SNR (PSNR) under varying SNR conditions. Yang and Kim [9] proposed a policy network with a gating mechanism that selects active feature groups per channel SNR using a single model, adapting the transmission rate without retraining. A more recent work has incorporated swin transformers [10] and non-linear transform coding [11] to further improve pixel-level image quality.

These frameworks have also been extended to the satellite setting. The authors in [12] proposed an end-to-end framework for efficient satellite-to-ground image transmission using semantic feature extraction and DJSCC, which reduces the onboard computational load through pruning and quantization while retaining more powerful decoding at the receiver. Their framework also includes a semantic feedback mechanism that adapts the transmitted representation and the transmission rate to the changing channel conditions. The authors in [13] presented an AI-assisted image processing pipeline that minimizes the amount of information transmitted over satellite links while preserving both semantic and perceptual fidelity. Their approach exploits the different relevance levels of the regions of interest of the image at the transmitter and applies generative AI at the receiver. Across all of these works, the design objective remains the recovery of the image at the receiver, whether at the pixel or at the perceptual level, and the performance is measured in terms of PSNR and structural similarity index measure (SSIM).

Despite these advances, most existing DJSCC methods optimize transmission for high-fidelity image reconstruction, an objective that differs fundamentally from semantic classification, where preserving task-relevant information matters more than reproducing visual detail. Since reconstruction requires

preserving the full image content, the encoders of these methods allocate channel resources to backgrounds and textures that are redundant for a classifier, which reduces inference accuracy under degraded channel conditions [14]. The work in [15] showed that training directly for the classification task produces more compact and discriminative representations, achieving higher inference accuracy under the same channel budget, and Kang *et al.* [16] confirmed this by transmitting only task-relevant semantic blocks, without reconstructing the image, which reduces overhead and improves scene classification accuracy in terrestrial and aerial deployments.

Shao *et al.* [17] proposed a variable-length information bottleneck framework that adjusts the number of transmitted features based on channel conditions for classification, but their approach relies on a fixed sparse prior structure and does not jointly train the encoder, the feature activation mechanism, and the classifier end-to-end as a unified adaptive system. Lyu *et al.* [18] proposed a gated JSCC framework with domain randomization to handle variable channel conditions while serving both image recovery and classification simultaneously, but the reconstruction loss dominates training, creating a direct conflict with the classification objective and limiting inference accuracy.

Furthermore, existing approaches handle channel variability using multiple independently trained models [19], [20], resulting in increased storage overhead and unstable classification under dynamic channels. From a goal-oriented communication perspective, image reconstruction is unnecessary for classification, as only task-relevant semantic information needs to be transmitted. Reconstruction-based methods, therefore, introduce additional communication and computational costs, limiting their applicability in resource-constrained wireless networks.

To address these challenges, we propose an end-to-end goal-oriented communication framework for image classification over an erasure channel, where the semantic encoder, feature selection gate, and task decoder are jointly optimized using only the classification objective. Unlike existing methods that rely on image reconstruction, fixed transmission targets, or multiple channel-specific models, the proposed framework employs a single trained model that dynamically adjusts the number of transmitted latent dimensions through an adaptive gating mechanism, without any predefined transmission target or manual regularization tuning.

Specifically, our main contributions of this paper are as follows:

- We propose an end-to-end framework for image classification over an erasure channel, which captures the short-term unavailability of the NTN link, where the semantic encoder, adaptive gate, and decoder are jointly optimized using a single classification objective.
- We introduce an erasure-probability-conditioned attention module that dynamically enhances task-relevant semantic features and suppresses redundant information according to channel conditions.
- We propose an erasure-aware feature selection gate that continuously adjusts the number of transmitted latent features based on the channel erasure probability, enabling efficient semantic communication under varying channel reliability.

The remainder of this paper is organized as follows. Section II describes the proposed GOC framework. Section III presents the experimental setup and results. Section IV concludes the paper.

II. Proposed Goal-Oriented Communication (GOC) Framework

We consider a GOC system for NTNs, in which a satellite platform captures an input image $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$ and transmits a compressed latent representation to a ground gateway station. Each captured image belongs to exactly one of a finite set of N_c predefined classes, and is associated with a ground-truth label $y \in \{1, \ldots, N_c\}$, where N_c denotes the number of target classes. We propose two model variants under this framework. The first, referred to as the proposed *regular* GOC, employs five convolutional layers at both the encoder and decoder with a latent dimension of $D = 1024$, targeting high-accuracy inference under varying link conditions. The second, referred to as the proposed *lightweight* GOC, reduces the encoder and decoder to three convolutional layers with $D = 256$, targeting resource-constrained onboard deployment scenarios. Both variants share the same gate architecture and are trained under an identical objective, enabling a direct comparison of representational capacity and architectural complexity under the same channel conditions.

The proposed framework, illustrated in Fig. 1, comprises three jointly trained modules: a channel-aware encoder $f_\phi^{(e)}$ and an erasure-adaptive feature selection gate $\mathcal{G}_\psi$, both deployed onboard the satellite, and a convolutional task decoder $f_\theta^{(d)}$ located at the ground gateway, with trainable parameters set ϕ, ψ, and θ, respectively. The gateway estimates the instantaneous erasure probability ϵ from the received signal quality and returns the estimate $\hat{\epsilon}$ to to the satellite as channel state information (CSI) via a dedicated uplink feedback channel, enabling the encoder and the gate to adapt the transmitted representation to the prevailing link conditions.

A. End-to-End Model Flow

Transmitter. Given the input image $\mathbf{x}$ and the estimated erasure probability $\hat{\epsilon}$, the encoder produces a D-dimensional latent representation

$$\mathbf{z} = f_\phi^{(e)}(\mathbf{x}, \hat{\epsilon}) \in \mathbb{R}^D, \tag{1}$$

where the semantic encoder is a convolutional network augmented with a multi-head self-attention (MHA) module conditioned on $\hat{\epsilon}$. The receiver estimates $\hat{\epsilon}$ from the received signal and returns it via the dedicated feedback link described above. The MHA module is conditioned on $\hat{\epsilon}$ through a two-layer neural network that maps $\hat{\epsilon}$ to a channel embedding vector $\mathbf{v}_{\hat{\epsilon}} \in \mathbb{R}^d$, where d denotes the token embedding dimension.

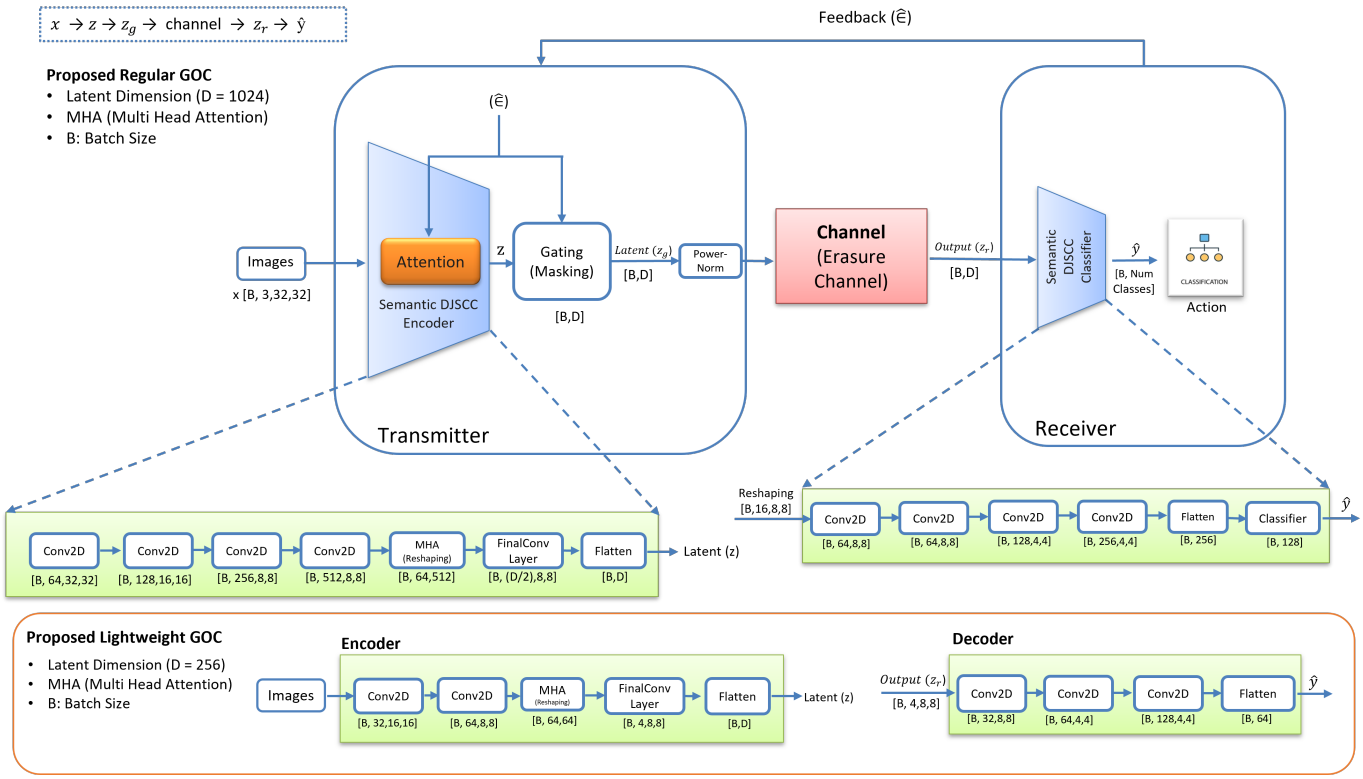

Fig. 1: Proposed GOC model pipeline.

The embedding is added to each of the S spatial tokens produced by the convolutional backbone prior to self-attention,

$$\tilde{\mathbf{h}}_i = \mathbf{h}_i + \mathbf{v}_{\hat{\epsilon}}, \quad i = 1, \ldots, S, \tag{2}$$

adapting the attention weights according to channel conditions. The feature selection gate then yields the gated vector

$$\hat{\mathbf{z}} = \mathbf{g}(\hat{\epsilon}) \odot \mathbf{z}, \tag{3}$$

where $\mathbf{g}(\hat{\epsilon}) \in \{0,1\}^D$ is a binary selection mask and $\odot$ denotes element-wise multiplication, as described in Section II-B. Prior to transmission, $\hat{\mathbf{z}}$ is normalized to unit power, yielding the transmitted vector $\mathbf{z}_{\text{tx}}$.

Channel: The communication channel is modeled as an erasure channel, used to capture the channel impairments resulting from the short-term unavailability of an NTN [21]. Each transmitted symbol is independently erased with probability $\epsilon \in [0, \epsilon_{\max}]$, so that the received signal is given by

$$z_{r,k} = \begin{cases} z_{\text{tx},k}, & \text{with probability } 1-\epsilon, \\ 0, & \text{with probability } \epsilon, \end{cases} \tag{4}$$

where $z_{\text{tx},k}$ represents the k-th transmitted symbol and $z_{r,k}$ denotes the corresponding received symbol. The maximum erasure probability is set to $\epsilon_{\max} = 0.8$. In the considered GOC system, ϵ serves as the only channel quality indicator and reflects the level of information loss affecting the classification task.

Receiver:

The received vector $\mathbf{z}_r \in \mathbb{R}^D$ is fed into the task decoder $f_\theta^{(d)}(\cdot)$, which produces the predicted class label

$$\hat{y} = f_\theta^{(d)}(\mathbf{z}_r) \tag{5}$$

where $\hat{y} \in \{1, \ldots, N_c\}$ is the inferred class label corresponding to the input image $\mathbf{x}$. The decoder implicitly learns robustness to variable erasure patterns through joint end-to-end training with the encoder and gate. Independently from the classification process, the receiver estimates the erasure probability ϵ from the received signal and feeds it back to the transmitter, enabling the encoder and gate to adapt the transmitted representation to the prevailing channel conditions, as indicated in Fig. 1.

B. Erasure-Adaptive Feature Selection Gate

The gate $\mathcal{G}_\psi$ controls the number of transmitted dimensions as a function of $\hat{\epsilon}$, without requiring a fixed target value. It is parameterized by a learnable importance score vector $\mathbf{s} \in \mathbb{R}^D$ and a small threshold network τ_ψ. For each latent dimension k, the binary gate activation is

$$g_k(\hat{\epsilon}) = \Big[\sigma\big((s_k - \tau_\psi(\hat{\epsilon}/\epsilon_{\max})) \cdot T\big) > 0.5\Big], \tag{6}$$

where σ is the sigmoid function and $T = 10$ is a temperature parameter. The threshold network τ_ψ is a four-layer MLP that maps the normalized erasure probability $\hat{\epsilon}/\epsilon_{\max} \in [0,1]$ to a scalar activation threshold, learning a distinct dimension selection profile for each channel condition. Gradients are

propagated through the hard threshold using the straight-through estimator [18]. The number of active, that is transmitted, dimensions at the estimated erasure level $\hat{\epsilon}$ is

$$K(\hat{\epsilon}) = \sum_{k=1}^{D} g_k(\hat{\epsilon}). \tag{7}$$

A key property of the proposed gate is that $K(\hat{\epsilon})$ increases with $\hat{\epsilon}$: under severe erasure, more dimensions are transmitted to compensate for channel losses, while under mild erasure, fewer dimensions suffice for accurate classification.

C. Gating-Regularized Training Objective

The encoder, gate, and decoder are jointly optimized by minimizing

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{g}}, \tag{8}$$

where $\mathcal{L}_{\text{cls}}$ is the label-smoothed cross-entropy classification loss, $\mathcal{L}_{\text{g}}$ is a gating regularization term defined below and $\lambda > 0$ is a non-negative weight controlling the trade-off between classification accuracy and transmission efficiency, updated adaptively during training.

Channel-adaptive gating loss: The gating term penalizes the expected number of dimensions that arrive intact at the decoder. For the i-th sample with true erasure probability ϵ_i, the expected number of surviving dimensions is $K(\hat{\epsilon}_i)(1-\epsilon_i)$, and the gating loss is defined as the batch average

$$\mathcal{L}_{\text{g}} = \frac{1}{B \cdot D} \sum_{i=1}^{B} K(\hat{\epsilon}_i)\,(1-\epsilon_i)^2, \tag{9}$$

where B is the batch size, $\epsilon_i \sim \mathcal{U}(0, \epsilon_{\max})$ is the true erasure probability of the i-th sample drawn independently at each training iteration, $\hat{\epsilon}_i$ is its estimated counterpart used by the gate, and $K(\hat{\epsilon}_i) = \sum_{k=1}^{D} g_k(\hat{\epsilon}_i)$ is the number of active dimensions for that sample. The squared survival factor $(1-\epsilon_i)^2$ imposes a larger gradient penalty when ϵ_i is small (reliable channel), encouraging the gate to reduce $K(\hat{\epsilon}_i)$ for efficient transmission, and a smaller penalty when ϵ_i is large (lossy channel), allowing the gate to increase $K(\hat{\epsilon}_i)$ to compensate for erasure losses. This asymmetric pressure is the source of the reverse gating property described in Section II-B.

Primal-dual adaptive λ: To avoid manual tuning of λ, we formulate the training objective as a constrained optimization problem in which $\mathcal{L}_{\text{cls}}$ is minimized subject to a constraint on the gating regularization loss $\mathcal{L}_{\text{g}}$, which controls the number of active transmitted dimensions. The weight λ serves as the Lagrange multiplier or dual variable, associated with this constraint and is updated at the end of each training epoch via gradient ascent on the log-barrier augmented dual objective, [22], $\mathcal{D}(\lambda) = \lambda\bar{\mathcal{L}}_g - \alpha \log \lambda$ yielding

$$\log \lambda \leftarrow \text{clip}\big(\log \lambda + \eta_\lambda \left(\lambda \bar{\mathcal{L}}_{\text{g}} - \alpha\right), \log \lambda_{\min}, \log \lambda_{\max}\big), \tag{10}$$

where $\bar{\mathcal{L}}_{\text{g}}$ is the epoch-averaged gating loss, η_λ is the dual step size, $\alpha > 0$ is the sole hyperparameter governing the target transmission efficiency, and $\text{clip}(\cdot, \log \lambda_{\min}, \log \lambda_{\max})$ constrains $\log \lambda$ to the interval $[\log \lambda_{\min}, \log \lambda_{\max}]$, ensuring λ remains within a stable operating range throughout training. At the fixed point of (10), the equilibrium satisfies

$$\lambda^* = \frac{\alpha}{\bar{\mathcal{L}}_{\text{g}}^*}, \tag{11}$$

so that the gate converges to a channel-dependent optimal number of active dimensions $K^*(\hat{\epsilon})$ that balances classification accuracy and transmission efficiency without a prescribed target. As ϵ decreases and the channel becomes more reliable, $K^*(\hat{\epsilon})$ decreases and the gate progressively retains only the most task-discriminative latent dimensions, which in turn improves classification accuracy. Reducing α drives the system toward a smaller $K^*(\hat{\epsilon})$ and a more compact representation, while increasing α relaxes the constraint and allows a larger active set.

D. Training Procedure

Training proceeds in three sequential phases. In the first phase, the gating regularization is inactive and the model is trained on $\mathcal{L}_{\text{cls}}$ alone, establishing a stable latent representation before transmission pressure is introduced. In the second phase, the effective regularization weight increases linearly from zero to the current value of λ, preventing abrupt disruption of the learned representations at the onset of gating regularization. In the third phase, the full objective in (8) is minimized and λ is updated each epoch via (10). Network parameters are updated jointly using the AdamW optimizer with a one-cycle cosine annealing schedule. Specific hyperparameter values and hardware details are provided in Section III-A.

III. Performance Evaluation

A. Experimental Setting

We evaluate the proposed system on CIFAR-10, a standard image classification benchmark comprising 50,000 training and 10,000 test images of size $32\times32\times3$ across 10 classes. Training images are augmented with random cropping with a padding of 4 pixels and random horizontal flipping, followed by per-channel normalization using the dataset statistics. All experiments are implemented in PyTorch and conducted on an NVIDIA T4 GPU via Google Colab. Models are trained with a batch size of 128 over 130 epochs using the AdamW optimizer [23] with a peak learning rate of 6×10^{-4}, a warm-up fraction of 0.30, weight decay of 10^{-4}, gradient norm clipping at 1.0, and label smoothing of 0.1. The three training phases described in Section II-D span over 30, 15, and 85 epochs, respectively. The regularization weight λ is initialized to $\lambda_0 = 0.05$ and updated via dual ascent with step size $\eta_\lambda = 5.0$, log-barrier strength $\alpha = 10^{-4}$, and bounds $\lambda_{\min} = 10^{-4}$ and $\lambda_{\max} = 2.0$. Per-sample erasure probabilities are drawn from $\epsilon \sim \mathcal{U}(0,\, \epsilon_{\max})$ at each training iteration.

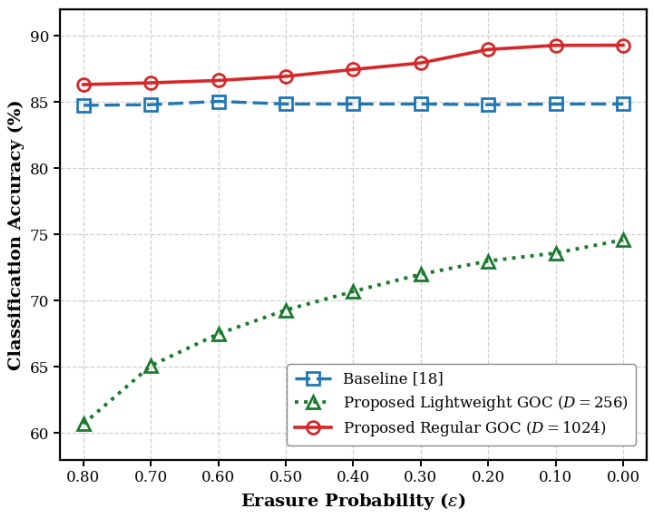

Fig. 2: Classification accuracy of the proposed *regular* GOC, proposed *lightweight* GOC, and baseline [18] versus erasure probability ϵ.

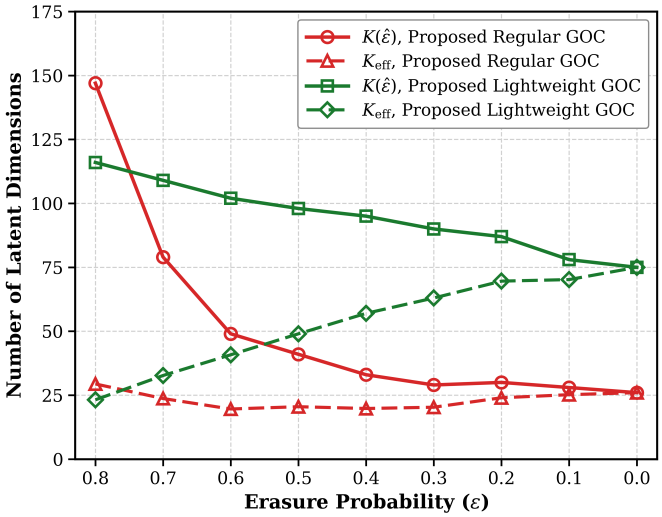

Fig. 3: Number of active transmitted latent dimensions $K(\hat{\epsilon})$ and effective number of dimensions received at the decoder, $K_{\text{eff}} = K(\hat{\epsilon})(1-\epsilon)$, versus the erasure probability ϵ, for the proposed regular GOC and the proposed lightweight GOC.

B. Experimental Results

For comparison, we reproduce the gated deep JSCC framework of [18] as a baseline. The original framework was designed for simultaneous image recovery and classification over an AWGN channel, using a gated net trained with domain randomization. We adapt it to the erasure channel setting and retain only the classification objective, keeping the original architecture unchanged.

Fig. 2 compares the classification accuracy of the proposed *regular* GOC, the proposed *lightweight* GOC, and the baseline [18] over $\epsilon \in [0, 0.8]$. The baseline maintains nearly constant accuracy across the entire erasure range, since domain randomization provides robustness to channel variability, but does not exploit improving channel conditions to enhance classification accuracy. In contrast, the proposed *regular* GOC improves from 86.32% at $\epsilon = 0.8$ to 89.29% at $\epsilon = 0$, consistently outperforming the baseline with an increasing margin as the channel becomes more reliable. The *lightweight* GOC follows a similar trend but achieves lower accuracy across all conditions, as the reduced latent dimension $D = 256$ and shallower architecture limit the amount of discriminative information available for classification. These results demonstrate that the proposed GOC more effectively exploits the erasure channel by preserving richer task-relevant features while progressively reducing the transmitted latent dimensions, thereby achieving superior classification performance over both the baseline and the lightweight model.

Fig. 3 shows the number of active transmitted dimensions $K(\hat{\epsilon})$ and the effective number of dimensions received at the decoder, $K_{\text{eff}} = K(\hat{\epsilon})(1-\epsilon)$, for both proposed variants, while Table I shows $K(\hat{\epsilon})$ together with the classification accuracy of the proposed regular GOC. For the regular GOC, $K(\hat{\epsilon})$ reduces from 147 to 26 dimensions as $\hat{\epsilon}$ decreases from 0.8 to 0, corresponding to 14.4% and 2.5% of D, while the accuracy improves from 86.32% to 89.29%, consistently with the reverse gating behaviour described in Section II-B. Over the same range, K_{eff} remains between 20 and 29 dimensions, indicating that the gate compensates the erasure losses so that the decoder receives an approximately constant task-relevant payload. The lightweight GOC follows the same trend, with $K(\hat{\epsilon})$ decreasing from 116 to 75 dimensions, but over a narrower range, and its effective number of dimensions decreases from 75 to 23 as ϵ increases. The reduced latent dimension therefore limits the extent to which erasure losses can be compensated, which accounts for the lower accuracy reported in Fig. 2. These results confirm that the proposed framework learns a compact task-oriented representation, and that the achievable compression depends on the representational capacity of the encoder.

TABLE I: Gate activity and classification accuracy of the proposed *regular* GOC scheme on CIFAR-10.

Erasure ε	K **Active**	**Active** K/D **(%)**	**Accuracy (%)**
0.80	147	14.4	86.32
0.70	79	7.7	86.45
0.60	49	4.8	85.63
0.50	41	4.0	86.94
0.40	33	3.2	87.46
0.30	29	2.8	87.95
0.20	30	2.9	88.97
0.10	28	2.7	89.28
0.00	26	2.5	89.29

Fig. 4 and Table II present the per-class confusion matrices and the precision, recall, and F1-score of the proposed *reg-*

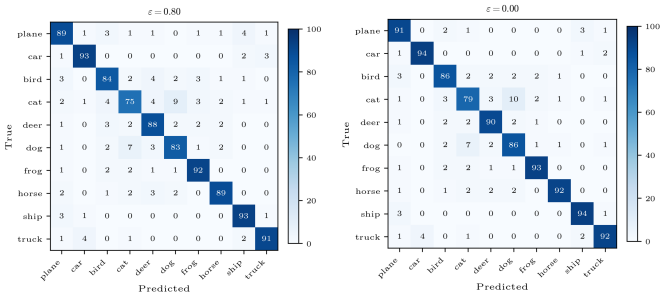

Fig. 4: Confusion matrices of the proposed *regular* GOC on CIFAR-10 dataset.

TABLE II: Per-class Precision (P), Recall (R), and F1-Score (F1) (%) of the proposed *regular* GOC on CIFAR-10 over an erasure channel.

	(ε=0.80)			(ε=0.50)			(ε=0.40)			(ε=0.00)		
Class	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Airplane	85.6	89.0	87.3	86.3	88.0	87.1	85.6	89.0	87.3	88.3	91.0	89.7
Automobile	93.0	93.9	93.5	92.2	94.0	93.1	93.1	94.9	94.0	95.9	94.0	94.9
Bird	86.6	84.0	85.3	85.0	85.0	85.0	85.9	85.0	85.4	87.8	86.0	86.9
Cat	80.6	73.5	76.9	83.3	75.0	78.9	82.6	76.0	79.2	82.3	79.0	80.6
Deer	84.6	88.0	86.3	86.0	86.0	86.0	86.3	88.0	87.1	90.0	90.0	90.0
Dog	83.8	83.8	83.8	82.4	84.0	83.2	84.0	84.0	84.0	83.5	86.0	84.7
Frog	90.2	92.9	91.5	91.9	91.0	91.5	92.9	91.0	91.9	93.0	93.0	93.0
Horse	91.8	89.0	90.4	92.8	90.0	91.4	91.8	90.0	90.9	96.8	92.0	94.4
Ship	93.0	94.9	93.9	91.2	93.0	92.1	93.0	93.0	93.0	93.1	94.0	93.5
Truck	91.0	91.9	91.5	86.5	91.8	89.1	88.5	92.9	90.6	86.8	92.0	89.3
Macro avg	**88.0**	**88.1**	**88.0**	**87.8**	**87.8**	**87.7**	**88.4**	**88.4**	**88.3**	**89.8**	**89.7**	**89.7**

ular GOC. The confusion matrices retain a strong diagonal structure at all values of ϵ, and as ϵ decreases from 0.80 to 0 the macro-averaged F1-score increases from 88.0% to 89.7%, a gain of 1.7 percentage points, accompanied by fewer off-diagonal errors. The largest confusion occurs between the semantically similar cat and dog classes, where 9 to 10% of cat samples are misclassified as dogs at every operating point, so that the cat class attains the lowest F1-score, between 76.9% and 80.6%. In contrast, automobile, frog, ship, and horse exceed 93% under favourable channel conditions. The macro-averaged F1-score varies by less than two percentage points over the full erasure range, indicating that channel quality primarily affects fine-grained discrimination between visually similar classes, while the learned representations remain robust for classes with distinctive visual characteristics.

IV. Conclusion

This paper proposed an end-to-end goal-oriented communication framework for image classification over a satellite-to-ground channel, in which a channel-aware encoder, erasure-adaptive feature selection gate, and task decoder are jointly optimized using only the classification objective. The proposed framework dynamically adjusts the number of transmitted latent dimensions through a primal-dual adaptive gating mechanism, without predefined transmission targets or manual regularization tuning. Experiments on CIFAR-10 demonstrate consistent classification accuracy of 86.32%–89.29% across $\epsilon \in [0, 0.8]$, while transmitting only 2.5%–14.4% of the latent dimensions.

Future work will investigate the impact of the channel feedback delay, extensions to multi-satellite cooperative inference, and more complex earth observation tasks such as object detection and semantic segmentation under resource-constrained onboard processing conditions.

V. Acknowledgement

This work was partially supported by UNITY-6G project, which received funding from the SNS JU under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101192650 and the SNS JU 6G-NTN2: Nexus project. The views expressed are those of the authors and do not necessarily represent the project. The Commission is not liable for any use that may be made of any of the information contained therein.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[2] T. M. Getu, G. Kaddoum, and M. Bennis, "Semantic communication: A survey on research landscape, challenges, and future directions," *Proceedings of the IEEE*, vol. 112, no. 11, pp. 1649–1685, Nov. 2024.

[3] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K.-K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, Jan. 2023.

[4] E. Calvanese Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, May 2021.

[5] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next-generation semantic communication networks," *IEEE Communications Surveys and Tutorials*, vol. 27, no. 1, pp. 37–76, 2025.

[6] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[7] M. Ding, J. Li, M. Ma, and X. Fan, "SNR-adaptive deep joint source-channel coding for wireless image transmission," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, pp. 1555–1559.

[8] J.-l. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2315–2328, Apr. 2022.

[9] M. Yang and H.-S. Kim, "Deep joint source-channel coding for wireless image transmission with adaptive rate control," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 5193–5197.

[10] K. Yang, S. Wang, J. Dai, X. Qin, K. Niu, and P. Zhang, "SwinJSCC: Taming Swin transformer for deep joint source-channel coding," *IEEE Transactions on Cognitive Communications and Networking*, Jul. 2024, doi: 10.1109/TCCN.2024.3424842.

[11] J. Dai, S. Wang, K. Tan, Z. Si, X. Qin, K. Niu, and P. Zhang, "Nonlinear transform source-channel coding for semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 8, pp. 2300–2316, Aug. 2022.

[12] M. Palena, F. Penna, R. Fantini, C. Amatetti, B. De Filippo, A. Muhammad, and C. F. Chiasserini, "Semantic and channel-aware image transfer in satellite networks," in *IEEE LCN 2026*, 2026.

[13] 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 - IEEE Conference on Computer Communications*, 2026, pp. 1–10.

[14] P. Zhang, Y. Liu, Y. Song, and J. Zhang, "Advances and challenges in semantic communications: A systematic review," *National Science Open*, vol. 3, no. 4, 2024, doi: 10.1360/nso/20230029.

[15] J. Shao, Y. Mao, and J. Zhang, "Learning task-oriented communication for edge inference: An information bottleneck approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 197–211, Jan. 2022.

[16] X. Kang, B. Song, J. Guo, Z. Qin, and F. R. Yu, "Task-oriented image transmission for scene classification in unmanned aerial systems," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5181–5192, Aug. 2022.

[17] J. Shao, Y. Mao, and J. Zhang, "Task-oriented communication for multi-device cooperative edge inference," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 73–87, Jan. 2023.

[18] Z. Lyu, G. Zhu, J. Xu, B. Ai, and S. Cui, "Semantic communications for image recovery and classification via deep joint source and channel coding," *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 8388–8404, Aug. 2024.

[19] F. Binucci, P. Banelli, P. Di Lorenzo, and S. Barbarossa, "Analog versus digital pulse amplitude modulation for goal-oriented wireless communications," in *Proc. European Signal Processing Conference (EUSIPCO)*, Sep. 2023, pp. 1415–1419.

[20] C. Wang, R. Olsson, S. Forssström, and Q. He, "Deep semantic inference over the air: An efficient task-oriented communication system," arXiv:2508.12748v2, Jan. 2026.

[21] E. Recayte and C. Amatetti, "Multi-satellite noma-irregular repetition slotted aloha for iot networks," in *ICC 2025-IEEE International Conference on Communications*. IEEE, 2025, pp. 2024–5209.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[23] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019.