Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3112

TASK-ORIENTED SEMANTIC COMMUNICATION FOR SMART CLASSROOM IMAGE ANALYTICS

Xue Lu[1], JiZhe Zhang[2*]
[1]*Foreign Languages College, Kashi University, Kashgar 844000, Xinjiang, China.*
[2]*College of Artificial Intelligence and Computer Science, Northwest Normal University, Lanzhou 730070, Gansu, China.*
[*]*Corresponding Author: JiZhe Zhang*

Abstract: Smart classrooms continuously generate large volumes of visual data that enable educational big data analytics, including learner-state assessment and classroom activity understanding. In practice, however, bandwidth-constrained and noise-impaired wireless links render conventional compress-then-transmit pipelines inefficient: pixel-fidelity-oriented compression typically requires high transmission rates to sustain downstream recognition performance. This paper studies a task-oriented semantic communication framework for image analytics over additive white Gaussian noise (AWGN) channels. We develop a unified multi-SNR, multi-rate training and evaluation protocol and benchmark a learned semantic link against a conventional DCT-based link under matched rate budgets. The proposed semantic link integrates (i) an encoder that extracts task-relevant latent representations, (ii) a rate-controllable channel-selection bottleneck that regulates the transmitted feature budget, and (iii) a decoder that reconstructs images for a fixed downstream classifier. Using CIFAR-10 as a reproducible testbed, we report task performance alongside perceptual quality metrics across a grid of SNR and rate settings. Experimental results indicate that the semantic link consistently sustains higher classification accuracy in low-to-medium SNR and low-rate regimes. In addition, PSNR/SSIM do not necessarily exceed the DCT baseline, revealing a task-perception mismatch that favors task-driven transmission. Overall, the proposed framework offers a practical methodology for designing communication pipelines that better support educational big data image analytics.
Keywords: Educational big data; Smart classroom; Semantic communication; Image classification

1 INTRODUCTION

Educational big data emphasizes continuous recording and computational analysis of the teaching–learning process to enable classroom diagnosis, learning analytics, and instructional optimization. With the deployment of smart classrooms, visual data collection has become routine and large-scale; images and videos are now key modalities for characterizing learning states and teaching activities. In addition, when edge devices upload captured content to edge or cloud servers, transmission is often constrained jointly by bandwidth budgets and channel noise. Conventional visual transmission pipelines typically prioritize pixel-level reconstruction quality: the source is first compressed, then transmitted over a noisy channel, and finally reconstructed at the receiver. Under low bit rates or low signal-to-noise ratios (SNRs), this paradigm is prone to structural distortions and noise amplification, which can impair the discriminative cues required by downstream recognition models. As a result, task performance may fluctuate, undermining the reliability and stability of educational visual analytics.

Semantic communication follows the principle of “transmit for the task.” Instead of targeting pixel fidelity, the transmitter extracts and conveys task-relevant semantic representations so that the receiver can maintain usable task-level performance under limited bandwidth and time-varying channels. Compared with reconstruction-centric pipelines, semantic communication focuses on preserving and robustly delivering information that is most relevant to the target task, and therefore has the potential to offer a better trade-off between task performance and resource consumption in educational visual analytics.

In this work, we focus on image classification as a fundamental visual analytics task. We establish a unified training and grid-based evaluation framework across multiple SNR and rate operating points. Under matched rate budgets and identical channel conditions, we compare the end-to-end performance of a learned semantic transmission pipeline against a conventional DCT-based pipeline. To ensure reproducibility and benchmark alignment, we conduct systematic evaluations on the CIFAR-10 dataset and report Top-1 accuracy together with PSNR/SSIM, thereby analyzing the relationship—and potential mismatch—between task performance and perceptual quality. The overall end-to-end evaluation pipeline is illustrated in Figure 1.

The main contributions of this paper are summarized as follows:

- We propose an end-to-end comparative framework for semantic communication in educational visual analytics, enabling fair evaluation of learned semantic pipelines and conventional pipelines under a unified channel model and a unified rate scale.
- We develop a joint training and evaluation protocol over a multi-SNR×multi-rate grid, covering discrete operating conditions and systematically reporting Top-1 accuracy and PSNR/SSIM.
- On CIFAR-10, we provide an in-depth comparison between the semantic pipeline and the DCT-based

- pipeline in the low-to-medium SNR and low-rate regimes, revealing that task accuracy and PSNR/SSIM may be inconsistent, and offering methodological insights for designing visual transmission links in educational big data systems.

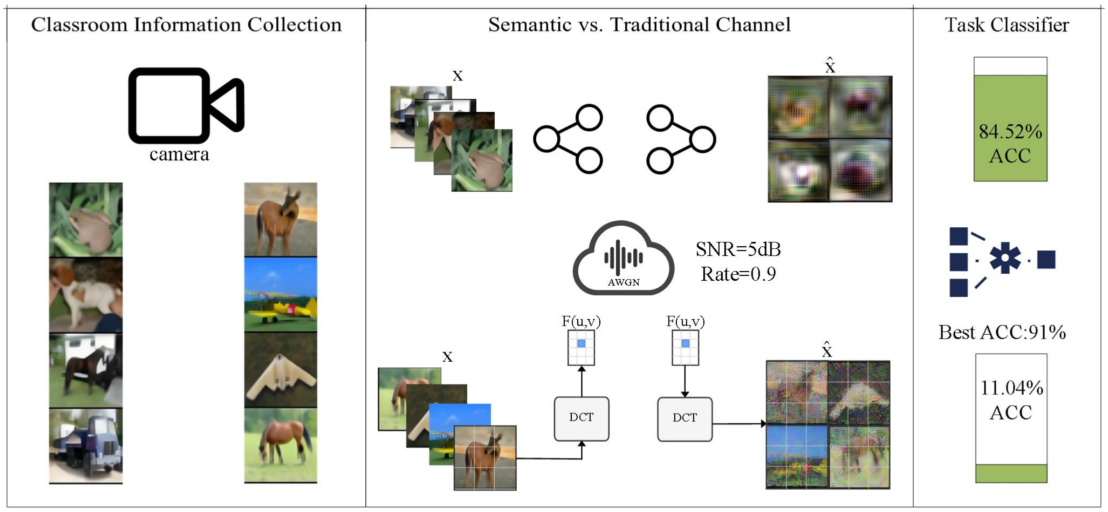

Figure 1 End-to-end Evaluation Pipeline for Task-oriented Image Transmission

2 RELATED WORK

2.1 Task-Oriented Transmission for Visual Analytics

Conventional multimedia communication pipelines are typically designed for *reconstructability*, optimizing pixel-level distortion metrics. However, in smart-classroom "capture–transmit–analyze" pipelines, the value of visual data is ultimately determined by downstream recognition and understanding tasks. Under joint constraints of low bit rates and noisy channels, improvements in pixel fidelity do not necessarily translate into better task performance; compression artifacts and channel noise may distort discriminative structures and degrade the statistics of task-relevant features, resulting in a "perceptual quality–task performance mismatch."

Task-oriented communication shifts the optimization objective from "pixel readability" to "task usability." A common approach is to directly transmit task-relevant representations via end-to-end learning and explicitly model the rate–relevance trade-off to improve edge inference under limited bandwidth[1]. In multi-device cooperative sensing, task-relevance constraints are further exploited to suppress redundant feature transmissions, reducing communication cost while controlling inference degradation[2]. Importantly, transmitting semantic features does not automatically guarantee privacy: under threats such as model inversion, feature-level leakage remains possible, which motivates joint optimization of privacy constraints and task utility for sensitive classroom visual data[3].

2.2 Semantic Communication Systems

Semantic communication emphasizes "conveying meaning" or "transmitting task-relevant information." Its core workflow is to extract semantic representations at the transmitter, deliver them robustly over the channel, and recover semantic elements at the receiver that are sufficient to support task-level decisions. End-to-end learning-based system frameworks provide a unified modeling paradigm for semantic encoding/decoding and training mechanisms[4]. Beyond individual models, survey studies systematically organize the field from the perspectives of semantic metrics, knowledge modeling, cross-layer co-design, and resource orchestration, and argue that semantic communication should be evaluated using multi-dimensional criteria—task utility, robustness, latency, and privacy—rather than relying solely on distortion measures[5–7]. These insights also motivate reproducible experimental protocols such as multi-SNR and multi-rate grid-based evaluations with end-to-end metric alignment[6].

2.3 Learned JSCC and Visual Semantic Links

Deep joint source–channel coding (Deep JSCC) integrates compression and channel adaptation through end-to-end learning. A key advantage is its *graceful degradation* behavior under varying SNR, making it well suited for wireless image transmission in uncertain channel conditions. Recent efforts largely follow two directions. The first develops controllable rate capacity mechanisms to enable a single model to operate across multiple working points and facilitate practical deployment [8]. The second extends Deep JSCC to more realistic and challenging channels (e.g., MIMO) and adopts stronger representation learning designs to improve robustness and generalization[9]. In addition, incorporating auxiliary semantics or guidance information to form complementary semantic spaces has been shown to enhance semantic effectiveness and noise resilience, offering reusable building blocks for task-relevant semantic enhancement[10]. Overall, while these studies establish important foundations for end-to-end semantic links, systematically achieving scalability in signaling overhead, low latency, and task utility in dynamic multi-hop or cooperative settings remains an open challenge.

To summarize these research streams and clarify how they inform our study, Table 1 provides a taxonomy across three directions, highlighting their objectives, representative methodological patterns, and key for the design and evaluation of task-oriented image transmission.

Table 1 A Taxonomy of Related Work across three Research Directions

Research Direction	Objective	Method
Task-oriented transmission for visual analytics [1]–[3]	Transmit task-relevant representations; maximize task utility rather than pixel fidelity; explicitly balance privacy and utility	End-to-end learned feature encoding/decoding with channel-robust mapping; relevance constraints for multi-device cooperation; privacy regularization and adversarial learning
Semantic communication systems and evaluation frameworks [4]–[7]	Convey meaning and task-relevant information; rely on knowledge and semantic metrics; require cross-layer co-design	End-to-end semantic transceiver frameworks; system architectures with protocol and resource coordination; multi-SNR and multi-task evaluation methodologies
Learned JSCC and visual semantic links [8]–[10]	End-to-end integration of compression and channel adaptation; graceful degradation across SNR; pursue rate adaptivity and semantic enhancement	Adaptive coding across rate points; MIMO-aware adaptive JSCC; auxiliary semantics for robustness and utility

3 METHODS

3.1 Problem Setup and Channel Model

Let the input image be $x\in\mathbb{R}^{H\times W\times 3}$ with class label $y\in\{1,\ldots,K\}$. The receiver performs classification on the reconstructed image using a pre-trained and fixed classifier $C(\cdot)$throughout all comparative evaluations. The predicted label is $\hat{y}$, and we report Top-1 accuracy as task-level metrics.

We consider an additive white Gaussian noise (AWGN) channel for transmitting the latent representation:

$$\tilde{s}=s+\varepsilon,\varepsilon\sim N(0,\sigma^2 I). \quad (1)$$

For each discrete signal-to-noise ratio (SNR) operating point, $\sigma^2$is set accordingly to reflect the desired noise level. The end-to-end performance is evaluated over a predefined SNR set to capture robustness under channel variations.

We apply an average power constraint to the transmitted latent symbols. Specifically, for each mini-batch we normalize $z_r$to satisfy $\mathrm{E}[\|s\|_2^2]/n=1$, where $n$is the number of transmitted scalar symbols. Under this normalization, the noise variance is set as $\sigma^2=10^{-\mathrm{SNR}/10}$for a given SNR in dB.

3.2 Semantic Communication Link with a Rate-Controllable Bottleneck

Our semantic link consists of an encoder f_θ, a rate-controllable bottleneck $G_r(\cdot)$, an AWGN channel, and a decoder g_ϕ. The encoder extracts a latent semantic representation:

$$z=f_\theta(x). \quad (2)$$

3.2.1Rate control via channel-wise bottleneck

To enable controllable rate operation, we introduce a bottleneck operator $G_r(\cdot)$parameterized by a retention ratio $r\in(0,1]$. Given r, the bottleneck forms a compressed representation

$$z_r=G_r(z;r), \quad (3)$$

where $G_r(\cdot)$can be implemented as **(i)** top-k channel selection or **(ii)** randomized gating , with $k=\lceil r\cdot d\rceil$ for a latent dimension d. This mechanism provides a unified "rate knob" that maps directly to the fraction of transmitted latent channels.

3.2.2Channel transmission and reconstruction

The compressed latent $z_r$is transmitted over the AWGN channel, producing $\tilde{z}$. The decoder reconstructs the image:

$$\hat{x}=g_\phi(\tilde{z}). \quad (4)$$

Finally, the receiver obtains the task prediction using the fixed classifier:

$$\hat{y}=C(\hat{x}). \quad (5)$$

Figure 2 illustrates the overall semantic link architecture.

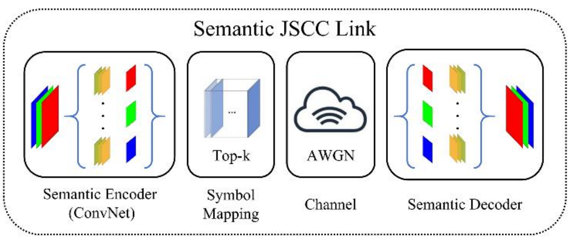

Figure 2 Semantic Link Architecture

3.3 End-to-End Training Objective

During training, we optimize $\{f_\theta, g_\phi\}$ end-to-end while keeping $C(\cdot)$ fixed. To balance reconstructability and task usability, we adopt a weighted combination of reconstruction loss and task loss:

$$\mathcal{L}=\lambda_1 \| x-\hat{x} \|_2^2+\lambda_2 \, \mathrm{CE}(C(\hat{x}),y), \tag{6}$$

where $\mathrm{CE}(\cdot)$ denotes cross-entropy loss and λ_1,λ_2 control the trade-off between perceptual fidelity and classification performance. This objective encourages the learned link to preserve information that is both visually meaningful and discriminative for the downstream task under channel and rate constraints.

3.4 Unified Multi-SNR and Multi-Rate Training and Grid-Based Evaluation

To ensure robustness across operating conditions, we adopt a unified multi-condition training strategy. For each mini-batch during training, we randomly sample an SNR value and a rate ratio r, and apply the corresponding noise level and bottleneck setting. This forces the semantic link to learn representations that generalize across heterogeneous channel qualities and communication budgets.

At evaluation time, we perform a grid traversal over all SNR–rate combinations. For each pair (SNR, r), we compute:

- Task metrics: Top-1 accuracy from $C(\hat{x})$;
- Perceptual metrics: PSNR/SSIM between x and $\hat{x}$, to quantify reconstruction quality and analyze potential mismatch with task accuracy.

The discrete operating sets used in our experiments are summarized in Table 2.

Table 2 Configuration of SNR and Rate Operating Points

Item	Setting
Rate set r	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
SNR set (dB)	-5,0,5,10,15,20,30

4 EXPERIMENTS AND DATA ANALYSIS

4.1 Dataset and Experimental Setup

We conduct experiments on CIFAR-10, a widely used benchmark for image classification. CIFAR-10 contains 60,000 color images of size 32×32 spanning 10 categories. Its standardized setting and moderate scale make it suitable for evaluating both the robustness and reproducibility of visual transmission pipelines under controlled conditions. CIFAR-10 is adopted as a reproducible benchmark to isolate the effect of link design; extending to classroom-specific datasets and tasks is left as future work.

To ensure that performance differences are attributable to transmission impairments rather than variations in the recognition backend, the receiver-side classifier $C(\cdot)$ is first trained on the original CIFAR-10 training set under a clean setting, i.e., without channel noise or link-induced distortions, and is then kept fixed throughout all comparative experiments. The semantic link adopts convolutional residual architectures for both the encoder and decoder. Rate control is implemented via a Top-k channel selection bottleneck aligned with the retention ratio r. Training uses the AdamW optimizer with a linear warm-up followed by cosine annealing, and the total number of training epochs is set to 80.

To cover diverse operating conditions, we evaluate the link over a discrete grid of SNR and rate points. Specifically, the retention-ratio set is $r\in\{0.1,0.2,\ldots,0.9\}$, and the SNR set is $\{-5,0,5,10,15,20,30\}\mathrm{dB}$. All results reported below are computed on this unified grid.

4.2 Evaluation Metrics

We assess performance from two complementary perspectives: task utility and perceptual quality. Task utility is primarily measured by Top-1 accuracy, reflecting the receiver's classification capability after transmission. Perceptual quality is quantified using PSNR and SSIM between the reconstructed image and the original input, capturing pixel-level fidelity and structural similarity, respectively. Importantly, all metrics are computed under identical settings to ensure a fair and consistent comparison across links.

4.3 Comparative Results and Discussion

Across the multi-SNR and multi-rate grid, the semantic link exhibits stronger task robustness in the low-to-medium SNR and low-rate regimes. When r is small or channel quality is poor, the degradation of Top-1 accuracy is noticeably more gradual, indicating that task performance remains relatively stable under stringent communication budgets and adverse channel conditions. In contrast, the conventional baseline is more susceptible to the compounded effect of aggressive information reduction and channel noise, which can lead to the loss of key structural cues and an amplification of perturbations at the receiver. Consequently, classification accuracy tends to deteriorate more rapidly at low SNR or low rate. These observations suggest that transmitting task-oriented representations is more effective in preserving discriminative information that directly supports classification under constrained and noisy links. Figure 3 presents the Top-1 accuracy curves as a function of the retention ratio $r$under multiple SNR levels.

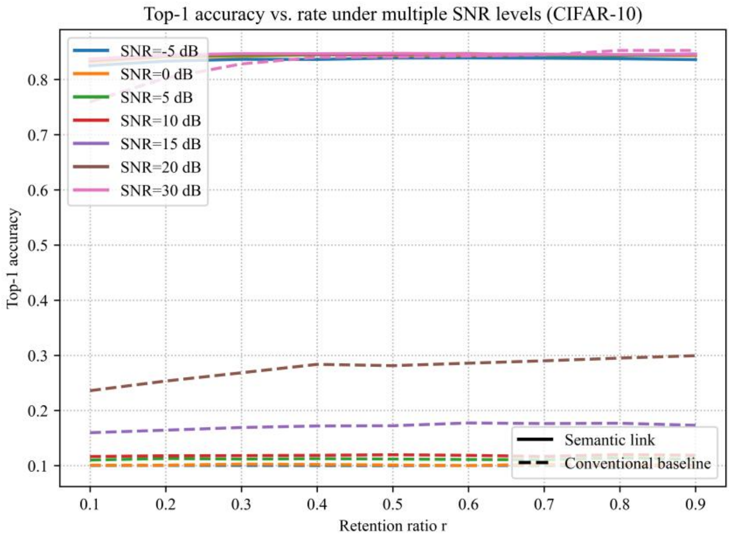

Figure 3 Top-1 Accuracy Versus Rate under Multiple Snr Levels on CIFAR-10.

From the perspective of perceptual quality, both PSNR and SSIM generally increase as r becomes larger, which is consistent with the intuition that retaining more information improves reconstruction fidelity. However, the semantic link is not necessarily superior to the conventional baseline in PSNR/SSIM across all operating points. This behavior is expected because the semantic link is optimized not only for reconstruction but also for task performance via an end-to-end objective; therefore, it may allocate limited rate resources preferentially to preserving task-critical semantic structures rather than uniformly enhancing pixel-level fidelity. As a result, there exist conditions where perceptual metrics are not dominant while classification accuracy remains higher. Figures 4 and 5 report PSNR and SSIM as functions of $r$under multiple SNR levels, respectively, to further illustrate this trade-off.

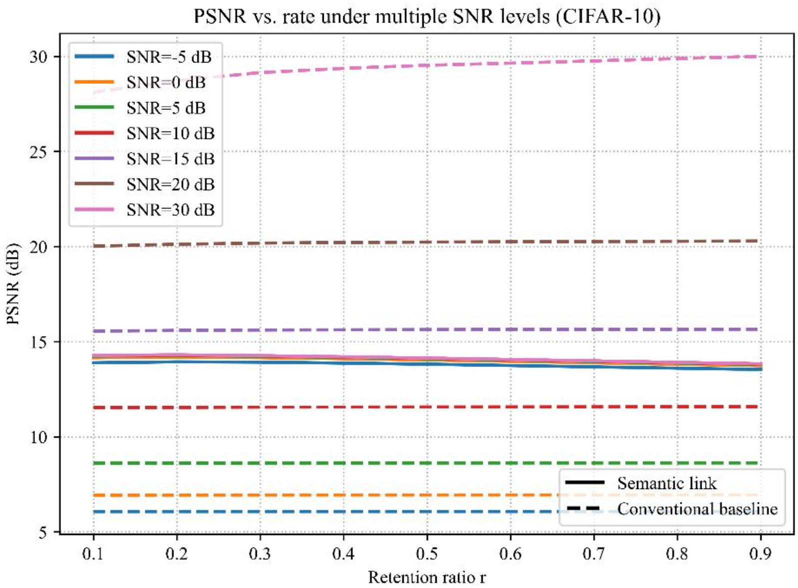

Figure 4 PSNR Versus Rate under Multiple SNR Levels on CIFAR-10.

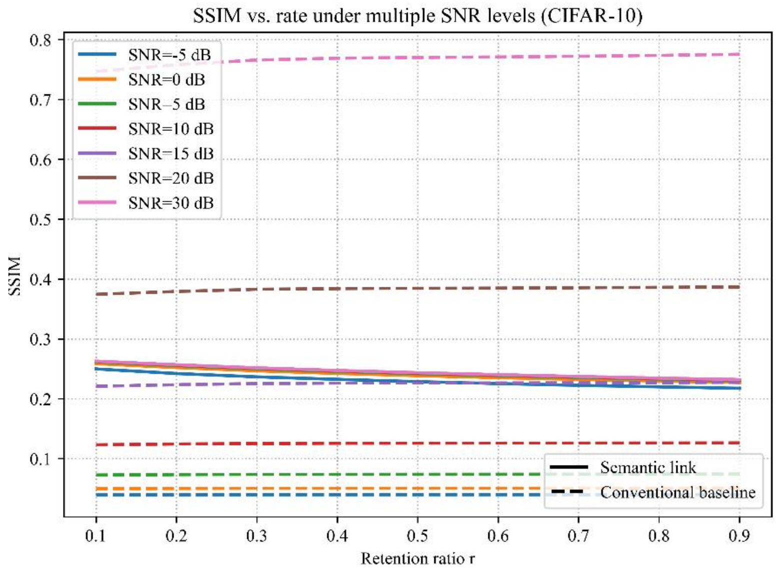

Figure 5 SSIM Versus Rate under Multiple SNR Levels on CIFAR-10

Taken together, the results indicate that the semantic link's advantage is most pronounced in the low-to-medium SNR and low-rate region, while perceptual quality does not necessarily improve in a uniformly consistent manner. This leads to an observable task-perception mismatch at certain operating points: task accuracy can remain high even when PSNR/SSIM is not superior. For smart-classroom educational visual analytics, where stable and reliable task outputs are often the primary objective, relying solely on perceptual metrics such as PSNR/SSIM is insufficient to characterize end-to-end utility. A more informative evaluation should jointly report task metrics and perceptual metrics on a unified SNR-rate grid, thereby revealing the trade-off boundary among task performance, perceptual quality, and communication resource consumption in a reproducible and deployment-relevant manner.

5 CONCLUSION

This paper targets the visual analytics requirements of education big data in smart classrooms and develops a task-oriented semantic communication link for image transmission. Under a unified channel model and a matched rate scale, we conduct a systematic comparison against a conventional DCT-based baseline. Extensive grid-based evaluations on CIFAR-10 across multiple SNR and rate operating points show that the proposed semantic link maintains substantially higher Top-1 accuracy in the low-to-medium SNR and low-rate regimes, demonstrating improved robustness to channel noise and severe rate constraints. At the same time, we observe that perceptual metrics do not consistently correlate with task accuracy, indicating that pixel-level or structural fidelity alone is insufficient for characterizing end-to-end utility in task-driven educational visual transmission. These findings underscore a central implication for smart-classroom deployments: link design and benchmarking should be guided by task-level objectives and assessed under deployment-relevant operating grids, rather than being dominated by reconstruction-centric criteria.

Looking forward, we will extend this framework to classroom-relevant vision tasks and end–edge collaborative inference workflows, where temporal dynamics, multi-camera views, and heterogeneous devices introduce additional constraints and opportunities. An important future direction is the investigation of adaptive rate-control mechanisms under more realistic channels and time-varying resource budgets, toward achieving a principled and scalable balance among task utility, robustness, latency, and privacy in practical smart-classroom systems.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Shao J, Mao Y, Zhang J. Learning Task-Oriented Communication for Edge Inference: An Information Bottleneck Approach. IEEE Journal on Selected Areas in Communications, 2022, 40(1): 197–211. DOI: 10.1109/JSAC.2021.3126087.

[2] Shao J, Mao Y, Zhang J. Task-Oriented Communication for Multidevice Cooperative Edge Inference. IEEE Transactions on Wireless Communications, 2023, 22(1): 73–87. DOI: 10.1109/TWC.2022.3191118.

[3] Wang Y, Guo S, Deng Y, Zhang H, Fang Y. Privacy-Preserving Task-Oriented Semantic Communications Against Model Inversion Attacks. IEEE Transactions on Wireless Communications, 2024, 23(8): 10150–10165. DOI: 10.1109/TWC.2024.3369170.

[4] Xie H, Qin Z, Li GY, Juang BH. Deep Learning Enabled Semantic Communication Systems. IEEE Transactions on Signal Processing, 2021, 69: 2663–2675. DOI: 10.1109/TSP.2021.3071210.

[5] Yang W, Du H, Liew ZQ, Lim WYB, Xiong Z, Niyato D, Chi X, Shen X, Miao C. Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges. IEEE Communications Surveys & Tutorials, 2023, 25(1): 213–250. DOI: 10.1109/COMST.2022.3223224.

[6] Liu Y, Wang X, Ning Z, Zhou MC, Guo L, Jedari B. A survey on semantic communications: Technologies, solutions, applications and challenges. Digital Communications and Networks, 2024, 10(3): 528–545. DOI: 10.1016/j.dcan.2023.05.010.

[7] Iyer S, Khanai R, Torse D, Pandya RJ, Rabie KM, Pai K, Khan WU, Fadlullah Z. A Survey on Semantic Communications for Intelligent Wireless Networks. Wireless Personal Communications, 2023, 129(1): 569–611. DOI: 10.1007/s11277-022-10111-7.

[8] Zhang W, et al. Predictive and Adaptive Deep Coding for Wireless Image Transmission in Semantic Communication. IEEE Transactions on Wireless Communications, 2023, 22(8): 5486–5501. DOI: 10.1109/TWC.2023.3234408.

[9] Wu H, Shao Y, Bian C, Mikolajczyk K, Gündüz D. Deep Joint Source-Channel Coding for Adaptive Image Transmission Over MIMO Channels. IEEE Transactions on Wireless Communications, 2024, 23(10): 15002–15017. DOI: 10.1109/TWC.2024.3422794.

[10] Feng J, Cao G, Liu J, Wang Z, Liu Z, Ding Y. Deep joint source-channel coding for wireless image transmission with auxiliary semantics. Physical Communication, 2025, 73: 102836. DOI: 10.1016/j.phycom.2025.102836.