# A Survey on Semantic Encoder Technology Trends for 6G Task-Oriented Semantic Communications

Chang gyo Jeong
*Department of Computer Science and Engineering*
*Kongju National University*
Cheonan City, Republic of Korea
0322brad@smail.kongju.ac.kr

Woongsoo Na
*Department of Software*
*Kongju National University*
Cheonan City, Republic of Korea
wsna@kongju.ac.kr

***Abstract*—In the era of 6G communications, massive data traffic and ultra-low-latency service requirements expose the limitations of conventional Shannon-based communication approaches that prioritize bit-level accuracy. Semantic communication has emerged as a key enabling technology to address these limitations, and the semantic encoder plays a central role by extracting task-relevant semantic information from source data and generating a compact semantic representation. This paper analyzes technology trends of semantic encoders, focusing on their architectures, learning paradigms, and extensibility, based on nine representative recent studies. Our analysis shows that semantic encoders are evolving along multiple directions, including knowledge-driven encoders, multimodal semantic relay structures, self-supervised (low-label) semantic encoders, LLM-enabled knowledge-augmented encoders, generative-model-based semantic encoding, metasurface-based physical-layer semantic encoding, and multi-user semantic separation approaches.**

***Index Terms*—Semantic encoder, semantic communication, task-oriented communication, self-supervised learning, LLM, 6G.**

## I. Introduction

6G mobile communications aim to simultaneously achieve ultra-low latency, ultra-reliability, and ultra-broadband performance. These requirements expose fundamental limitations of conventional Shannon-based communication, which focuses on bit-level accurate reconstruction. In many emerging applications, such as autonomous driving or robotic teleoperation, reconstructing raw data is unnecessary; instead, only task-relevant semantic information needs to be delivered. Figure 1 illustrates the conceptual difference between conventional bit-level communication and semantic communication, highlighting the shift from signal reconstruction to task-oriented meaning delivery.

To address this paradigm shift, semantic communication has emerged as a promising alternative, where the semantic encoder plays a central role. A semantic encoder extracts task-essential representations $Z$ from input data $X$, which can be formulated using the Information Bottleneck (IB) principle:

$$\min_{p(z|x)} I(X;Z) - \beta I(Z;Y), \tag{1}$$

where $Y$ denotes the target task and $\beta$ is a hyperparameter that balances semantic preservation and compression. This formulation theoretically supports the idea that a semantic encoder should "discard task-irrelevant information and preserve only semantically important information."

The notion of semantic encoding is also aligned with several research streams in machine learning and computer vision. The Neural Semantic Encoder proposed by Farsad *et al.* [1] designed the transmitter and receiver as an end-to-end neural architecture, demonstrating that semantic representations can be delivered robustly under noisy channels.

From a representation learning perspective, the marginalized latent semantic encoder (MLSE) by Ding *et al.* [2] provides an important foundation. MLSE optimizes the following objective to learn generalized latent semantic representations even under noise or limited samples:

$$\min_{W,Z,S} \|W\tilde{X} - Z\|_F^2 + \alpha\|Z - AS\|_F^2 + \beta\|S - H\|_1, \tag{2}$$

which matches the requirements of semantic communication where meanings must be preserved under distortion and incomplete observations.

Meanwhile, classical encoder–decoder architectures such as SegNet [3] effectively extract hierarchical semantics (feature hierarchies) while removing unnecessary pixel-level details. This structural advantage naturally connects to the philosophy that a semantic encoder should extract "meaning-centered representations rather than fine-grained reconstruction."

In addition, the Context Encoding Network (EncNet) proposed by Zhang *et al.* [4] leverages global contextual semantics to modulate class-dependent features. Its key process can be expressed via context residual computation as:

$$e_{ik} = \frac{\exp(-s_k\|x_i - d_k\|^2)}{\sum_j \exp(-s_j\|x_i - d_j\|^2)}(x_i - d_k), \tag{3}$$

followed by semantic modulation:

$$\gamma = \sigma(We), \tag{4}$$

which is conceptually equivalent to emphasizing task-relevant semantic factors and suppressing irrelevant information.

Overall, these prior studies highlight core elements necessary for semantic encoder design: (1) semantic-centric optimization rather than bit-level accuracy, (2) selective extraction of task-related features, (3) utilization of global semantic structures, and (4) robustness to noise and domain shifts. Building

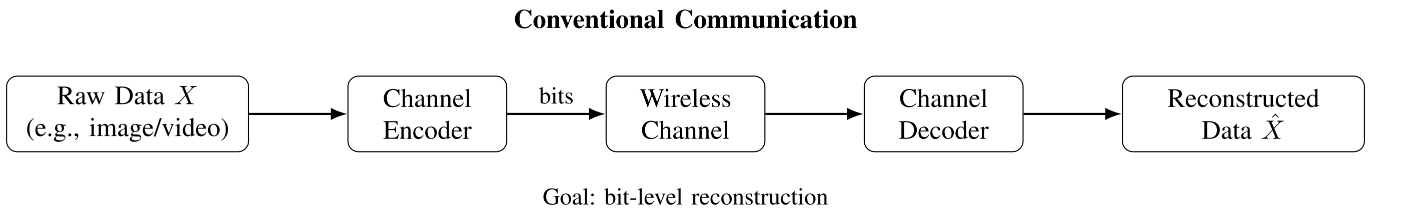


Fig. 1. Comparison between conventional bit-level communication and semantic communication.

on these foundations [1]–[4], this paper systematically analyzes nine recent semantic communication studies [5]–[13] to examine structural evolutions, learning methods, system-level scalability, and future research directions for semantic encoders.

## II. Definitions and Evaluation Metrics

This section clarifies key definitions and evaluation metrics that recur across recent semantic communication studies. In particular, we highlight the need for consistent measurement protocols to enable fair comparison across heterogeneous tasks, modalities, and system settings.

### *A. Semantic Fidelity, Semantic Noise, and Task Performance*

Unlike conventional communications that mainly focus on bit-level fidelity (e.g., BER/BLER), semantic communication emphasizes *semantic fidelity*, i.e., how accurately task-relevant meaning is delivered for downstream task execution. We categorize distortions as follows:

**(1) Channel noise:** Distortions introduced by the physical channel (e.g., additive noise, fading), which can be captured by SNR or BER/BLER under a given modulation and coding setting.

**(2) Semantic noise:** Distortions in *meaning* that cause task-level degradation even if symbol-level recovery is partially successful. Semantic noise includes (i) misinterpretation of the intended semantics, (ii) task mismatch between transmitter and receiver, (iii) hallucination or irrelevant knowledge injection in LLM-enabled pipelines, and (iv) misalignment of multimodal semantics in distributed settings.

To quantify semantic noise, a practical approach is to measure the gap between transmitted and recovered semantics using task-level and representation-level metrics (e.g., task accuracy drop, semantic similarity reduction). For example, semantic noise can be operationally characterized as:

$$\Delta_{\text{sem}} = 1 - \text{Sim}(Z_{\text{tx}}, Z_{\text{rx}}), \tag{5}$$

where $\text{Sim}(\cdot,\cdot)$ can be cosine similarity or other task-relevant similarity measures, and $Z_{\text{tx}}, Z_{\text{rx}}$ are semantic representations at the transmitter and receiver, respectively. In practice, $\text{Sim}(\cdot,\cdot)$ should be instantiated according to the modality and task (e.g., embedding cosine similarity for text, CLIP/BERTScore-style alignment for vision-language, or class-wise similarity for recognition tasks).

### *B. Semantic Capacity and Semantic QoS*

In addition to Shannon capacity, recent studies motivate the notion of *semantic capacity*, which informally refers to the maximum achievable task performance (or semantic fidelity) under limited communication resources. Accordingly, semantic capacity is often operationalized as an achievable utility region (task performance versus resource cost), rather than a single closed-form scalar as in classical Shannon theory. Because semantic communication objectives vary across tasks, semantic capacity is typically expressed in terms of task-level utility under resource constraints:

$$C_{\text{sem}}(\mathcal{R}) = \max_{\pi \in \Pi} \mathcal{U}\big(\text{Task}(Z_{\text{rx}})\big) \quad \text{s.t.} \quad \text{Cost}(\pi) \leq \mathcal{R}, \tag{6}$$

where $\mathcal{R}$ denotes a resource budget (e.g., power, bandwidth, latency), $\pi$ is a semantic encoding/communication policy, and $\mathcal{U}(\cdot)$ denotes task utility (e.g., detection accuracy, success rate, or a weighted QoS score).

In practice, *semantic QoS* can be reported as a composite measure that jointly reflects semantic QoS, latency, and resource usage:

$$Q_{\text{sem}} = \alpha \cdot \mathcal{U}_{\text{task}} + \beta \cdot \text{Sim}_{\text{sem}} - \gamma \cdot \text{Latency}, \tag{7}$$

where $\alpha, \beta, \gamma$ are application-dependent weights.

### *C. Need for Consistent Measurement Protocols*

A key challenge in surveying semantic encoders is the lack of unified evaluation protocols. Existing works report diverse metrics (e.g., mAP, task success rate, semantic similarity, BER/BLER), making cross-paper comparison difficult. To promote consistent evaluation, we recommend reporting results along three complementary axes:

**(A) Task-level metrics:** application-specific performance (e.g., mAP for object detection, classification accuracy, trajectory success rate).

**(B) Representation-level metrics:** semantic similarity between recovered and reference semantics (e.g., cosine similarity between embeddings, CLIP score for vision-language alignment, BERTScore for text).

**(C) Resource-level metrics:** communication cost such as latency, bandwidth usage, energy/power, and compute overhead.

Moreover, we recommend that papers explicitly state (i) dataset/task definition, (ii) channel model and SNR range, (iii) ablation settings (w/ or w/o knowledge/LLM modules, w/ or w/o relay/edge), and (iv) a common semantic QoS score (e.g., (7)) to enable fair comparisons across different semantic encoder designs.

## III. Technology Trends of Semantic Encoders

Semantic encoders are evolving along the following five perspectives: (1) knowledge-driven and task-oriented semantic encoding, (2) distributed and multimodal semantic processing, (3) LLM/modular/physical-layer extensions, (4) multi-user and resource management, and (5) self-supervised semantic representation learning. We analyze the nine papers based on this taxonomy.

### A. Knowledge-Driven and Task-Oriented Semantic Encoding

In Guo *et al.* [5], object-level semantic units are extracted from UAV imagery using YOLO-World and SAM. The semantic encoder adjusts the transmitted semantic representation according to the importance weight $w_i$ of each object:

$$Z = \sum_{i=1}^{N} w_i f_i(X). \tag{8}$$

This provides high efficiency for object-centric tasks such as military reconnaissance and emergency rescue.

Tian *et al.* [6] further leverage a conditional GAN-based generative model to synchronize semantic models between transmitter and receiver without sharing a common dataset. This enables semantic encoding without data sharing under privacy-preserving scenarios.

These knowledge-driven, task-oriented studies [5], [6] represent practical applications of IB-style objectives in real networks and significantly improve semantic resource utilization efficiency.

### B. Distributed, Multimodal, and Edge-Based Semantic Encoding

The DTCN architecture by Guo *et al.* [8] proposes a multimodal semantic relay structure in which the device, relay, and edge jointly transmit and align different semantic features. Semantic alignment can be conceptually modeled as:

$$Z_{\text{edge}} = F_{\text{align}}(Z_{\text{dev}}, Z_{\text{relay}}), \tag{9}$$

which reduces noise and modality mismatch by performing semantic complement across modalities. This suggests important design directions for distributed semantic encoding in autonomous driving, smart cities, and multi-robot collaboration.

### C. LLM-Based, Modular, and Metasurface Semantic Encoding

Rachwan *et al.* [9] propose a modular semantic encoding framework that dynamically selects processing pathways according to input context and task requirements, enabling adaptive LLM-based semantic processing.

Hu *et al.* [7] introduce a semantic-similarity-based hallucination suppression mechanism when fusing LLM-generated knowledge with original representations:

$$S(x, g) = \cos\big(f(x), f(g)\big), \tag{10}$$

and semantic fusion is performed only when similarity exceeds a threshold:

$$z = \lambda z_x + (1-\lambda) z_{\text{LLM}}. \tag{11}$$

Huang *et al.* [10] further propose a metasurface-based semantic encoding paradigm that directly maps semantic tags to radiation patterns,

$$E(\theta, \phi) = G(\Phi_{\text{tag}};\, \Theta), \tag{12}$$

enabling semantic transmission at the physical layer without intermediate digital encoding.

### D. Multi-User Semantic Separation and Power Allocation

Ma *et al.* [11] formulate semantic interference in multi-user scenarios using an information-theoretic framework, where the received representation jointly reflects user semantics, transmitted symbols, and interference:

$$\min_{p(y|x,s,u)} -I(U;Y) - \lambda I(X;Y) - I(S;Y), \tag{13}$$

with $\lambda$ controlling the trade-off between semantic preservation and signal fidelity.

The received signal is modeled as

$$y_i = g_{i,i} x_i + \sum_{j \neq i} g_{i,j} x_j + n_i, \tag{14}$$

indicating that co-channel transmission naturally induces semantic interference. From this perspective, SFDMA can be interpreted as a semantic-domain multiple access mechanism.

Xu *et al.* [12] further study semantic-aware power allocation for generative semantic communication. The received signal for the $i$-th semantic stream is

$$y_i = \sqrt{q_i}\, h_i z_i + n_i, \tag{15}$$

where $q_i$ denotes the allocated power. The semantic value is defined as

$$L_i = 1 - P^i, \tag{16}$$

with $P^i$ representing perceptual accuracy.

The power allocation problem is formulated as

$$\min_{\{q_i\}} \sum_i K_i q_i \tag{17}$$

$$\text{s.t.} \quad \hat{L}_i(\psi_i) \geq \bar{L}_i, \ \forall i, \tag{18}$$

which enforces semantic QoS constraints rather than conventional SNR-based criteria. This formulation highlights a shift from Shannon-capacity-centric design toward semantic-aware resource management.

### E. Self-Supervised Semantic Encoding

The SLSCom study by Gu *et al.* [13] considers label-scarce real-world settings and proposes a contrastive-learning-based self-supervised framework to learn semantic representations, followed by fine-tuning with a small amount of labels. A representative InfoNCE loss is defined as:

$$\mathcal{L}_{\text{contrast}} = -\log \frac{\exp\big(\text{sim}(z, z^{+})/\tau\big)}{\sum_{z' \in \mathcal{N}} \exp\big(\text{sim}(z, z')/\tau\big)}, \tag{19}$$

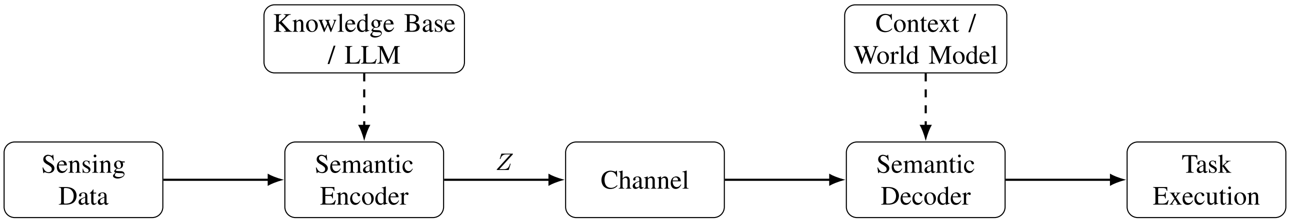


Fig. 2. End-to-end semantic communication pipeline integrating knowledge/LLM and context for task-oriented meaning transfer.

where $z$ and $z^+$ form a positive pair (same semantics), $\mathcal{N}$ is the set of negative samples, and $\tau$ is the temperature parameter. This strengthens the representation power of semantic encoders by bringing embeddings of the same semantics closer while pushing different semantics apart.

### *F. Discussion: Toward Unified Evaluation for Semantic Encoders*

Although the reviewed works demonstrate rapid progress, cross-paper comparison remains challenging due to heterogeneous tasks, modalities, and evaluation criteria. To reduce ambiguity, we encourage future studies to (i) report both task-level and representation-level metrics, (ii) explicitly define semantic noise sources beyond channel noise (e.g., hallucination, task mismatch, multimodal misalignment), and (iii) adopt a unified semantic QoS score that jointly considers semantic QoS, latency, and resource costs. Such standardized reporting will facilitate reproducibility and enable more meaningful comparisons across semantic encoder architectures and learning paradigms.

## IV. Conclusion

Figure 2 summarizes an end-to-end semantic communication pipeline integrating perception, knowledge, and task execution. This paper reviewed recent advances in semantic encoder design for 6G task-oriented communications. We showed that semantic encoders are evolving from bit-level compression modules into intelligent components that integrate communication, perception, and reasoning.

Key trends include knowledge-driven encoding, multimodal fusion, LLM-enabled semantic reasoning, and semantic-aware resource management. These developments indicate a paradigm shift from Shannon-centric metrics toward semantic QoS–oriented system design.

Future research should focus on unified evaluation frameworks, robust semantic noise modeling, and real-world validation across diverse application domains. Such efforts will be essential to realizing practical and scalable semantic communication systems for next-generation networks.


## Acknowledgment

This research was supported in by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ITRC(Information Technology Research Center)(IITP-2025-RS-2022-00156353) grant funded by the Korea government(Ministry of Science and ICT) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2019-NR040074)


## References


[1] N. Farsad, S. O. Rao, and A. Goldsmith, "Deep Learning-Based Semantic Communication Systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 701–714, 2018.

[2] Z. Ding and M. Shao, "Marginalized Latent Semantic Encoder for Zero-Shot Learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 6191–6199, 2019.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder–Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

[4] H. Zhang *et al.*, "Context Encoding for Semantic Segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 7151–7160, 2018.

[5] C. Guo *et al.*, "A Knowledge Base Driven Task-Oriented Image Semantic Communication Scheme," 2024.

[6] Z. Tian *et al.*, "Model-Enabled Task-Oriented Semantic Communications Through Knowledge Synchronization," 2025.

[7] W. Hu *et al.*, "Task-Oriented Semantic Communication with Large Language Model Enabled Knowledge Base," ICC 2025.

[8] J. Guo *et al.*, "Distributed Task-Oriented Communication Networks with Multimodal Semantic Relay and Edge Intelligence," 2024.

[9] J. Rachwan *et al.*, "Dynamic Neural Pathway Synthesis for LLM Processing Using Modular Semantic Encoding," 2024.

[10] W. HuangFu *et al.*, "Semantic Encoding Metasurface for Traffic Signage," ICMMT 2025.

[11] S. Ma *et al.*, "Semantic Feature Division Multiple Access," IEEE TWC, 2024.

[12] C. Xu *et al.*, "Semantic-Aware Power Allocation for Generative Semantic Communications with Foundation Models," 2024.

[13] R. Gu *et al.*, "Task-Oriented Low-Label Semantic Communication With Self-Supervised Learning," IEEE TWC, 2025.

TABLE I
COMPARISON OF REPRESENTATIVE RECENT STUDIES ON SEMANTIC ENCODERS (TASK, MODALITY, METRICS, PROS/CONS).

| **Work** | **Task** | **Modality** | **Learning / Key Idea** | **Evaluation Metrics (Reported / Typical)** | **Strengths / Limitations** |
|---|---|---|---|---|---|
| Guo *et al.* [5] | UAV image semantic comm. (object-centric) | Vision (UAV imagery) | Knowledge-base driven object extraction (YOLO-World + SAM), importance weighting $w_i$ | Task utility (e.g., detection performance such as mAP/precision/recall), bandwidth/latency reduction; semantic QoS under resource constraint | **+** Efficient object-level semantics for mission-critical tasks.<br>**-** Performance depends on detector/segmenter quality; limited beyond object-centric tasks. |
| Tian *et al.* [6] | Knowledge synchronization for semantic comm. | Vision / generic data | Conditional GAN-based model synchronization without sharing common datasets | Semantic similarity / task accuracy under privacy constraints; convergence stability indicators | **+** Avoids raw data sharing; privacy-preserving semantic alignment.<br>**-** GAN training instability; synchronization overhead not always negligible. |
| Hu *et al.* [7] | Task-oriented semantic comm. with LLM knowledge | Text (and knowledge base) | LLM-enabled knowledge augmentation with hallucination suppression via similarity thresholding | Semantic similarity (e.g., cosine over embeddings), task success/accuracy, hallucination rate/consistency checks, latency/compute | **+** Knowledge boosts robustness under missing information.<br>**-** Risk of hallucination and extra compute; metric choices vary across tasks. |
| Guo *et al.* [8] | Distributed task-oriented comm. w/ multimodal relay | Multimodal (vision+sensor) | Device/relay/edge semantic relay and alignment $F_{\text{align}}(\cdot)$ | Task success rate, semantic alignment score, latency (edge/relay), bandwidth saving | **+** Robust to modality mismatch via semantic complement.<br>**-** Coordination/synchronization complexity; requires edge/relay availability. |
| Rachwan *et al.* [9] | LLM processing with dynamic modular pathways | Text (LLM) | Dynamic pathway synthesis (modular semantic encoder) depending on input/context/task | Downstream task metrics, inference cost (FLOPs/latency), routing accuracy/efficiency | **+** Adaptive compute-resource trade-off; modular extensibility.<br>**-** Routing/controller overhead; reproducibility depends on module design. |
| HuangFu *et al.* [10] | Physical-layer semantic encoding for traffic signage | RF / semantic tags | Metasurface maps semantic tags to radiation/phase patterns (direct physical-layer encoding) | Semantic error rate (SER), detection/recognition rate, latency/energy efficiency | **+** Ultra-low latency semantic transmission at physical layer.<br>**-** Scenario specificity; hardware constraints limit general applicability. |
| Ma *et al.* [11] | Multi-user semantic separation (multiple access) | Signals / multi-user | Semantic Feature Division Multiple Access using mutual-information objectives | Semantic rate / mutual information terms, interference robustness, task-level utility | **+** Formalizes semantic interference; enables semantic-aware multiple access.<br>**-** Implementation complexity; mapping MI objectives to practical systems remains open. |
| Xu *et al.* [12] | Semantic-aware power allocation for generative semantic comm. | Multi-stream (semantic symbols) | Power minimization under semantic-performance constraints (semantic QoS) | Power/energy, BER/BLER (proxy), semantic QoS / task accuracy constraints, latency | **+** Resource management centered on semantic QoS, not Shannon capacity.<br>**-** QoS definition varies; may rely on BER as imperfect proxy for semantics. |
| Gu *et al.* [13] | Low-label semantic comm. (label-scarce) | Multimodal / generic | Self-supervised contrastive learning (InfoNCE) + low-label fine-tuning | Representation quality (contrastive), downstream task accuracy, label efficiency | **+** Strong under limited labels; improved generalization.<br>**-** Needs careful positive/negative pair design; task transfer may vary. |