Engineering 56 (2026) 34–44

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
6G: From Theory to Practice—Article

Explicit Semantic-Base-Empowered Communications for 6G Mobile Networks

Fengyu Wang [a], Yuan Zheng [b], Wenjun Xu [b,*], Junxiao Liang [b], Ping Zhang [b], Zhu Han [c]

[a] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[c] Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004, USA

ARTICLE INFO

Article history:
Received 9 December 2024
Revised 13 July 2025
Accepted 25 August 2025
Available online 9 September 2025

Keywords:
Semantic communications
Semantic bases
Knowledge base
Unequal error protection

ABSTRACT

Increasing demands for massive data transmission pose significant challenges to communication systems. Compared with traditional communication systems that focus on the accurate reconstruction of bit sequences, semantic communications (SemComs), which aim to deliver information connotation, are regarded as a key technology for sixth-generation (6G) mobile networks. Most current SemComs utilize an end-to-end (E2E) trained neural network (NN) for semantic extraction and interpretation, which lacks interpretability for further optimization. Moreover, NN-based SemComs assume that the application and physical layers of the protocol stack can be jointly trained, which is incompatible with current digital communication systems. To overcome those drawbacks, we propose a SemCom system that employs explicit semantic bases (Sebs) as the basic units to represent semantic connotations. First, a mathematical model of Sebs is proposed to build an explicit knowledge base (KB). Then, the Seb-based SemCom architecture is proposed, including both a communication mode and a KB update mode to enable the evolution of communication systems. Sem-codec and channel codec modules are designed specifically, with the assistance of an explicit KB for the efficient and robust transmission of semantics. Moreover, unequal error protection (UEP) is strategically implemented, considering communication intent and the importance of Sebs, thereby ensuring the reliability of critical semantics. In addition, a Seb-based SemCom protocol stack that is compatible with the fifth-generation (5G) protocol stack is proposed. To assess the effectiveness and compatibility of the proposed Seb-based SemComs, a case study focusing on an image-transmission task is conducted. The simulations show that our Seb-based SemComs outperform state-of-the-art works in learned perceptual image patch similarity (LPIPS) by over 20% under varying communication intents and exhibit robustness under fluctuating channel conditions, highlighting the advantages of the interpretability and flexibility afforded by explicit Sebs.

© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Over the past few decades, the rapid development of mobile communications has profoundly revolutionized human society, shifting the world from the Internet of Things (IoT) to the Internet of Everything (IoE). A variety of innovative services, such as extended reality (XR) and digital twins (DTs), have emerged, with escalating requirements for data throughput and ubiquitous intelligence, catalyzing the development of sixth-generation (6G) mobile networks.

Current state-of-the-art communication frameworks are mainly designed based on Shannon's classical information theory (CIT) [1], which characterizes information as random variables and assumes that semantics are irrelevant to communications, for the sake of the generality of the theory [2]. However, 6G envisions the integration of ubiquitous artificial intelligence (AI) terminals to build a network that incorporates not only communications but also perception and intelligence in order to achieve ubiquitous connectivity and intelligence. Consequently, CIT-based communication systems are struggling to meet the grand vision of 6G, necessitating a fundamental shift in the communication paradigm [3].

* Corresponding author.
E-mail address: wjxu@bupt.edu.cn (W. Xu).

https://doi.org/10.1016/j.eng.2025.08.039
2095-8099/© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Thanks to recent success in AI, user equipment (UE) such as IoT sensors and smartphones can understand the meaning of raw data and use it to perform intelligent tasks [4]. As a result, semantic communications (SemComs), which perform in an "understanding-before-transmission" manner [5], have been rapidly developing in recent years and are considered a key technology in achieving the vision of 6G [6].

SemComs are capable of adaptively constructing scenario-specific services by processing the essential connotations of messages according to the intent of intelligent tasks (Fig. 1), thereby increasing the intelligence and efficiency of the system. More specifically, the meaning of messages is extracted and sent in a form that is understandable for both source and destination, with a knowledge base (KB) being introduced in SemComs to ensure transceivers share the same background knowledge. A prevalent approach employed in recent works involves the establishment of an implicit KB based on an end-to-end (E2E) trained neural network (NN) for SemComs [7–9], also known as deep joint source and channel coding (DeepJSCC). To ensure transceivers can understand each other, NNs at both the transmitter and receiver are jointly trained in semantic encoding/decoding (codec), where the parameters of the NNs serve as the implicit KB of SemComs.

Although techniques such as DeepJSCC have demonstrated the superiority of SemComs in terms of communication efficiency, the implicit KB-based architecture presents several challenges for practical implementation. A primary issue is that these systems are trained under the premise that the environment (i.e., the statistics of messages and channels) is static [10]. If channel conditions or the knowledge within messages deviate from the scope of the training data, the entire system needs to be retrained, resulting in considerable additional overheads [11,12]. To overcome this issue, a task-unaware SemCom system for image transmission is proposed in Ref. [13], where transfer learning is employed to evolve the KB, adapting the system under dynamic data settings. Update mechanisms for the encoder–decoder NN and the KB module are proposed in Ref. [14], in which KB is modeled as an explicit module independent of the encoder–decoder structure. However, both studies [13,14] indirectly treat the training dataset as the KB, making the system's utilization of knowledge indirect and necessitating an additional fine-tuning process. Moreover, the aforementioned approaches struggle with issues of interpretability, as the lack of interpretability makes it difficult to measure the effectiveness of systems and make relevant updates. To theoretically clarify the interpretability of SemComs, a mathematical theory of SemComs is proposed in Ref. [15], which models SemComs from the perspective of synonymous mapping. However, the question of how to implement SemComs following the proposed explicit and controllable synonymous mapping has not yet been comprehensively studied.

In addition, thanks to transceivers' ability to understand the meaning of the source message in SemComs, implementing unequal error protection (UEP) based on varying levels of message importance for a specific task has become an effective way to increase the efficiency of SemCom [16,17]. Semantic importance can be detected by entropy modeling [18], saliency detection [19], gradient-weighted class activation mapping (Grad-CAM) [20], and so forth. However, most studies apply UEP based on an implicit KB; as a result, their importance modeling is mainly limited to the individual messages themselves. Additionally, most current joint source and channel coding (JSCC)-based SemComs assume that the communication systems support full-resolution constellation points [21] and therefore directly map semantics to continuous complex constellations. As a result, it is difficult to make these JSCC-based SemComs compatible with current digital communication systems. To achieve compatibility, DeepJSCC-Q-type techniques [21,22] set constellation points as finite and use an NN to learn constellation mapping. However, the modulation scheme cannot change once the model is trained, resulting in a lack of flexibility in practical use. Moreover, SemComs based on DeepJSCC assume that the application layer and physical layer

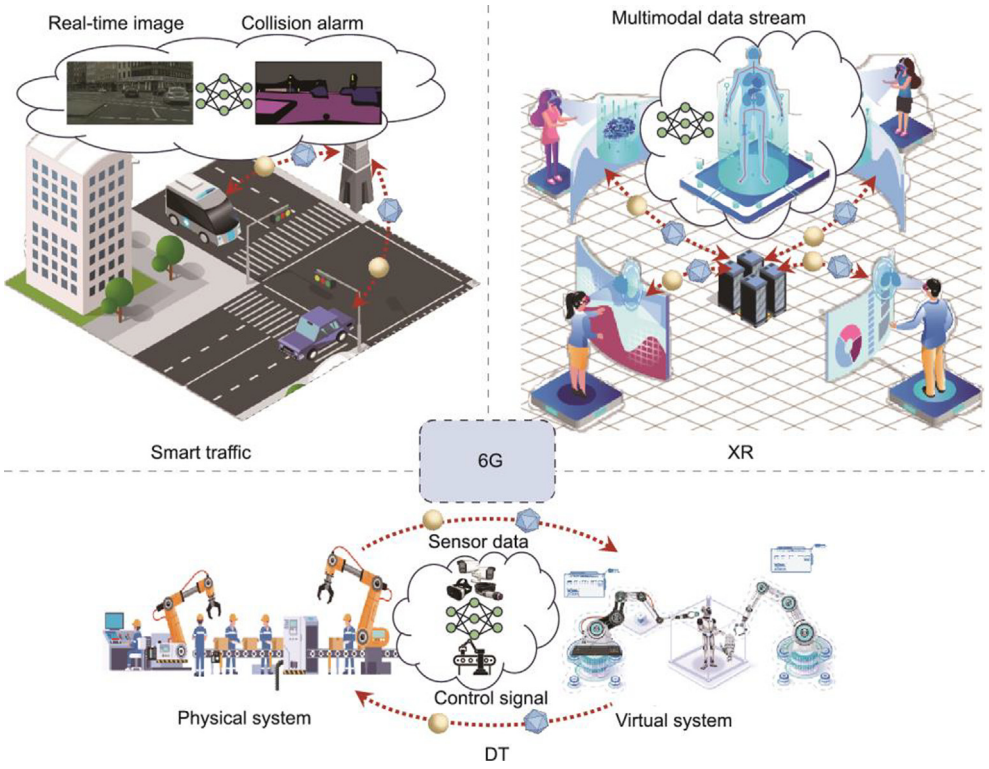

Fig. 1. SemCom-empowered services for typical 6G applications.

can be jointly trained, ignoring the constraints proposed by current protocol stacks. Therefore, the question of how to design compatible SemCom systems with a higher degree of interpretability and flexibility is an open yet important topic for 6G.

To this end, in this article, we propose a new SemCom architecture based on explicit semantic bases (Sebs), where the generation of Sebs serves as the direct realization of synonymous mapping [23]. Moreover, two levels of UEP—that is, Seb-wise and message-wise UEP—are proposed to improve the transmission efficiency of SemComs. Protocol stacks compatible with CIT-based systems are further proposed. The contributions of this work are summarized as follows:

- A mathematical model of Sebs is proposed, laying the foundation for building an explicit KB for SemComs. Based on this, a SemCom architecture with an explicit KB is proposed, consisting of both a communication mode and an update mode, enabling the intelligent evolution of the system.
- Criteria for designing a semantic encoder/decoder (Sem-codec) and channel codec are illustrated, where the communication intent is incorporated through the Seb importance level and messages to provide UEP for critical semantics.
- To facilitate the practical implementation of SemComs, a Seb-based SemCom protocol stack consistent with the current 3rd Generation Partnership Project (3GPP) protocol stack is proposed, bridging the gap between theory and application with respect to SemComs for 6G mobile networks.

The remainder of the paper is organized as follows: Section 2 introduces the architecture of the Seb-based SemComs, including Sebs, the KB, the Sem-codec, and the channel codec. Next, the protocol of Seb-based SemComs is further discussed in Section 3, followed by a case study to demonstrate the effectiveness of the proposed Seb-based SemComs in Section 4. The work is concluded in Section 5.

2. Architecture of SemComs with explicit Sebs

In this section, we provide an overview of the proposed Seb-based SemCom architecture. As shown in Fig. 2, the architecture involves two modes—namely, the communication mode and the KB update mode—enabling the proposed SemCom to provide intelligent services while being capable of self-evolution and optimization. During communications, the KB of the access point (AP) and that of the UEs are synchronized, facilitating the extraction and interpretation of message semantics. Notably, the underlying intent of communications may change over time, making a KB obsolete and resulting in inefficient semantic representation (SR). Consequently, updating is necessary to enable the evolution of the network. The KB update mode is triggered by requests from the UEs, during which the AP generates and transmits a KB update message by analyzing the recently sent messages. To enable flexible SR and efficient KB updates, the KB is composed of explicit Sebs, represented by the diamond-marked cubes in Fig. 2. In the remainder of this section, the mathematical model of Sebs will be introduced in Section 2.1, based on which the criteria for designing the KB, Sem-codec, and channel codec will be further illustrated in Sections 2.2, 2.3, and 2.4, respectively.

2.1. Semantic base

The introduction of Sebs in the KB grants SemCom systems great interpretability and flexibility compared with implicit KB-based approaches. As the basic unit for expressing semantic information, Sebs serve as an analogous concept to bits in CIT [23]. Nonetheless, the granularity of Sebs can be adjusted to align with the intent of communication for intelligent services, thereby meeting diverse representational demands.

We regard the generation of Sebs as a concrete form of synonymous mapping [15], where messages sharing the same meaning in multiple manifestations can be represented by the same Seb. In other words, the relationship between syntactic information and its corresponding semantics is typically many-to-one mapping [15], as shown in Fig. 3. In practice, clustering algorithms could be employed to realize the aforementioned process of synonymous mapping. By clustering the raw semantic features of message segments with similar semantics, the corresponding cluster centers can effectively summarize the emerged "synonyms." Given a set of accessible messages $\mathscr{X} = \{\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, ..., \tilde{\boldsymbol{x}}_l, ...\}$, where $\tilde{\boldsymbol{x}}_l$ denotes the lth message and $l \in \{1, 2, ...\}$ is the message index, this process can be mathematically expressed as follows:

$$\mathscr{S}^1 = f_{\text{cluster}}(f_{\text{SE}}(\mathscr{X})) = f_{\text{cluster}}(\mathscr{F}) \tag{1}$$

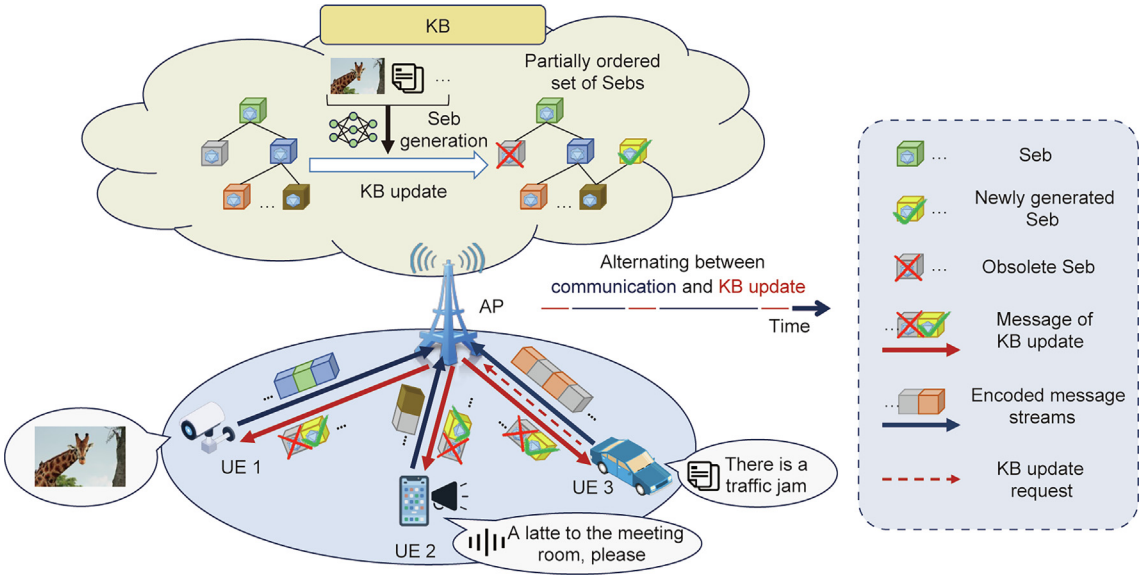

Fig. 2. Architecture of the Seb-based SemComs. The system performs in two modes: the communication mode and the KB update mode. During communications, messages are interpreted based on a synchronized KB. A KB update is triggered through requests from the UEs to evolve the system. AP: access point.

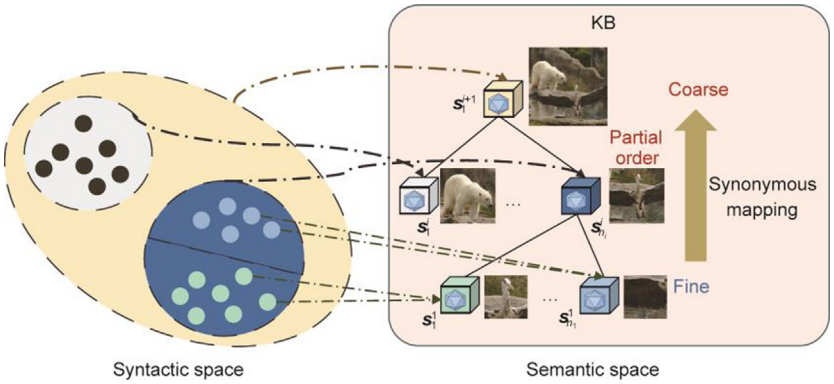

Fig. 3. Illustration of a partially ordered set of Sebs, where the semantics of the original message at various granularities are depicted by Sebs at different levels. $\boldsymbol{s}_{n_i}^{i}$: the n_ith Seb at the ith level; n_i: the total number of Sebs in $\mathscr{S}^i$; $i \in \{1, 2, \ldots\}$: the level index; $\mathscr{S}^i$: the generated set of Sebs at the ith level.

where $f_{\mathrm{SE}}(\cdot)$ denotes the semantic extraction module, which can be implemented using a pretrained NN to output a collection of raw semantic features $\mathscr{F}$. $f_{\mathrm{cluster}}(\cdot)$ denotes the clustering algorithm employed (e.g., K-means). The variable $\mathscr{S}^1 = \{\boldsymbol{s}_1^1, \boldsymbol{s}_2^1, \ldots, \boldsymbol{s}_{n_1}^1\}$ represents the generated set of Sebs at the first level, where n_1 represents the total number of Sebs in $\mathscr{S}^1$.

A synonymous mapping correlation also exists between fine- and coarse-grained semantics. For instance, in scenarios requiring coarser Seb granularity, such as higher-level intelligent tasks including classification or object detection, multiple fine-grained Sebs can be further mapped into a single coarse-grained Seb based on synonymous mapping, thereby leading to a more abstract representation of information for transmissions (shown in Fig. 3 as Sebs in higher levels). We can iteratively cluster the Sebs obtained from Eq. (1) to derive higher-level Sebs; that is,

$$\mathscr{S}^{i+1} = f_{\mathrm{cluster}}(\mathscr{S}^i) \tag{2}$$

Here, $\mathscr{S}^i$ with superscripts $i \in \{1, 2, \ldots\}$ represents the set of Sebs at the ith level. In other words, Sebs in different granularities exhibit a hierarchical structure. Accordingly, we model the set of Sebs as a partially ordered set $\langle \mathscr{S}, \underline{\prec} \rangle$, where Sebs $\boldsymbol{s}_1^1, \boldsymbol{s}_2^1, \ldots, \boldsymbol{s}_{n_1}^1, \boldsymbol{s}_1^2, \ldots, \boldsymbol{s}_{n_i}^i$ are elements in the set of Sebs $\mathscr{S}$, $\boldsymbol{s}_{n_i}^i$ represents the n_ith Seb at the ith level, n_i is the total number of Sebs in $\mathscr{S}^i$, and $\underline{\prec}$ denotes the synonymous mapping relationship, indicating the process of mapping fine-grained Sebs into coarse-grained Sebs, as shown in the Hasse diagram in Fig. 3.

Fig. 3 gives an exemplary process of Seb generation for an image, where syntactic information that records pixel-level details corresponds to the finest granularity points in the feature space, shown as dots in the left part. Fine-grained Sebs aggregate groups of pixel-level details sharing the same meaning, corresponding to the lowest row of Sebs in the KB. Coarse-grained Sebs, shown in the first and the second rows in the KB, aggregate sets of similar fine-grained Sebs, forming more abstract semantic concepts corresponding to messages that share the same meaning in a broader perspective (i.e., the dots in the dashed circles in the left part of the figure). Overall, by leveraging the many-to-one correlation between the source messages and their corresponding semantics across different granularities, Sebs can condense multiple manifestations of the same meaning into a unified SR, thereby improving efficiency.

The encoding process utilizing Sebs will be described further in Section 2.3. In practical systems, Seb granularity is determined by both user intent and the available transmission resources. For certain intelligent tasks, we can assign varying levels of importance to different fractions of the transmitted message, considering their differentiated representational needs. More specifically, coarse-grained Sebs excel at capturing abstract semantics, which are suitable for resource-constrained scenarios to support the transmission of essential task-related semantics. In comparison, fine-grained Sebs, which incur higher communication overheads, are more accurate in SR and are ideal for conveying detailed semantics according to the communication intent. The combination of Sebs with different granularities offers a flexible and efficient methodology for the execution of intelligent tasks. Moreover, it should be noted that Sebs are not limited to representing messages within the same modality. By customizing feature-extraction methods for different modalities, a shared feature space can be built that jointly aligns semantics from different modalities. Consequently, Sebs can be generated following the same procedure shown in Fig. 3, providing a shared SR for data in various modalities. In other words, Sebs can be used to facilitate the alignment of multimodal data, which would be significant for the realization of ubiquitous connectivity and intelligence for novel applications in 6G networks.

2.2. Knowledge base

The KB is a distinctive component of SemComs that has no counterpart in CIT-based networks. By synchronizing knowledge across transceivers, the KB enables messages to be interpreted with Sebs that serve as shared prior knowledge, thus ensuring the effectiveness of SemComs.

It should be noted that different fractions of the transmitted message may have varying levels of importance, and the importance of Sebs should differ accordingly. For example, in a scenario where a set of images are transmitted for object recognition, Sebs that mainly represent the images' foreground should be more important than those mainly representing the background. Moreover, communication intents may often change dynamically in practical scenarios, leading to a mismatch between statically generated Sebs and dynamic intents and decreasing the accuracy and efficiency of SR. Therefore, dynamic updates of Sebs are essential and should further consider Seb importance levels, which are shown as the red bars beside the Sebs in Fig. 4.

To be specific, the importance levels of Sebs are first initialized based on the importance of the Sebs' corresponding raw semantic features or lower-level Sebs; that is,

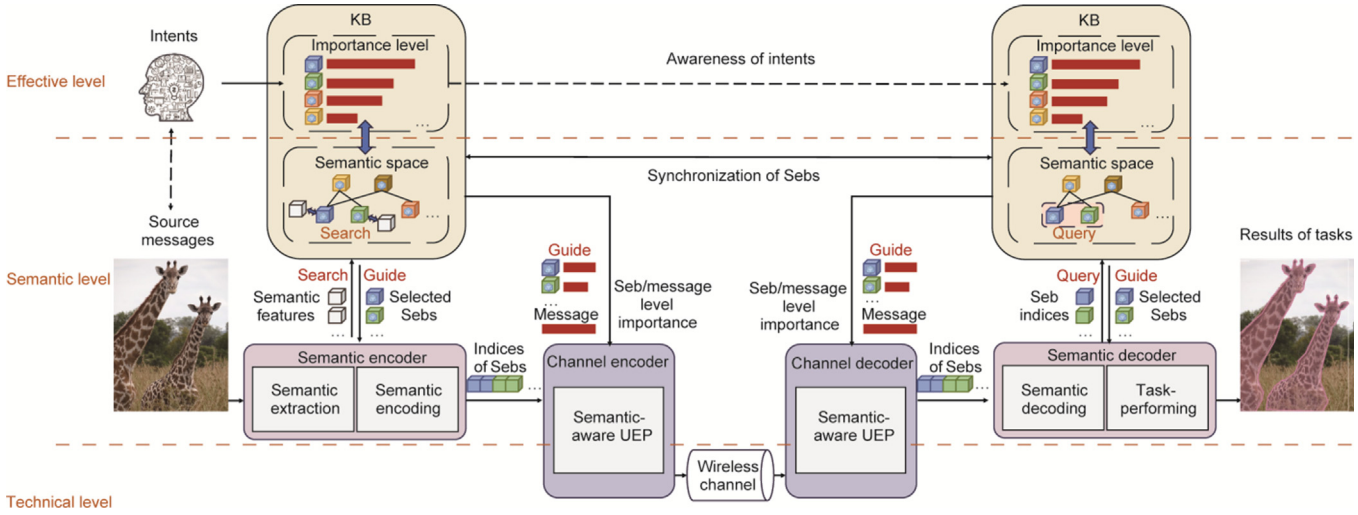

Fig. 4. Architecture of the proposed SemCom system. Sebs are regarded as the basic unit for expressing semantic information and are synchronized in the KB to support the implementation of the Sem-codec and channel codec.

$$I_k^1 = \frac{1}{|\mathscr{F}_k|} \sum_{f \in \mathscr{F}_k} I_f \tag{3}$$

where I_k^1 denotes the importance of $\boldsymbol{s}_k^1$. $\mathscr{F}_k \subseteq \mathscr{F}$ is the set of features clustered into $\boldsymbol{s}_k^1$, $|\cdot|$ denotes the cardinality of the set, and I_f is the importance of the raw semantic feature $\boldsymbol{f}$ that can be obtained through methods such as Dino [24] or Grad-CAM [20]. Additionally, we have

$$I_k^{i+1} = \frac{1}{|\mathscr{S}_k^i|} \sum_{\boldsymbol{s}_m^i \in \mathscr{S}_k^i} I_m^i \tag{4}$$

for higher-level Sebs with $i \in \{2, 3, \ldots\}$. Similarly, $\mathscr{S}_k^i \subseteq \mathscr{S}^i$ is the set of Sebs clustered into $\boldsymbol{s}_k^{i+1}$, I_m^i denotes the importance of the mth Seb in the ith level ($m \in \{1, 2, \ldots, n_i\}$), which is further clustered into $\boldsymbol{s}_k^{i+1}$. I_k^i is the importance of the kth Seb in the ith level ($k \in \{1, 2, \ldots, n_i\}$).

To ensure the efficiency of the proposed SemCom system, a KB update is triggered when the UE detects inaccuracy in the SR, as indicated by the dashed line in Fig. 2. In implementation, the accuracy of the SR can be measured by the distortion D between the raw semantic features of a message to be transmitted $\boldsymbol{x}$ and the Sebs closest to them—that is, the Euclidean distance:

$$D(f_{\mathrm{SE}}(\boldsymbol{x}), \mathscr{S}) = \| f_{\mathrm{SE}}(\boldsymbol{x}) - f_{\mathrm{quant}}(f_{\mathrm{SE}}(\boldsymbol{x}), \mathscr{S}) \|_2 \tag{5}$$

where $f_{\mathrm{quant}}(\cdot)$ is the semantic encoding module that identifies the closest Sebs in a predefined level based on the importance of the raw semantic features. When the accuracy falls below the predefined threshold, the UE will regard the current SR as inaccurate. The update process includes the following strategies to capture emerging topics:

- **Add new Sebs.** New candidate Sebs are generated from recent messages, along with their corresponding levels of importance. These candidates are integrated into the KB if they exhibit significant differences from the existing Sebs (e.g., may also be measured by the Euclidean distance).
- **Adjust Seb importance.** The levels of importance of existing Sebs are further adjusted according to their age of information (AoI) [4], which underscores the importance of freshness. Specifically, the level of importance of a Seb decreases over time but will be refreshed upon usage; that is,

$$I_k^i(t) = I_k^i \cdot \Delta(t - t_0), \quad t \geq t_0 \tag{6}$$

$\Delta(t - t_0)$ is a monotonically decreasing function satisfying $\Delta(0) = 1$ and $\Delta(+\infty) = 0$. Here, t_0 denotes the most recent time at which Seb $\boldsymbol{s}_k^i$ was used and t is the current time stamp.

- **Remove obsolete Sebs.** Sebs with levels of importance falling below a predefined threshold are considered obsolete and are removed from the KB, thereby reducing KB redundancy and improving encoding efficiency.

The KB update messages are synchronized from the AP to the UEs, as shown in Fig. 2.

2.3. Semantic encoder/decoder

The Sem-codec is deployed to extract and interpret the source message utilizing Sebs, which can be realized by an NN. The semantic encoder mainly comprises a semantic extraction module $f_{\mathrm{SE}}(\cdot)$ and a semantic encoding module $f_{\mathrm{quant}}(\cdot)$, as shown in Fig. 4. The source message is first mapped into the semantic space through semantic extraction, yielding the corresponding hierarchical semantic features. Then, the semantic features are quantized by corresponding Sebs within the KB. Based on their levels of importance, these features are quantized using Sebs with different granularities. To be specific, more important features are quantized using finer Sebs, while less important features are quantized using coarser Sebs. The semantic encoder $f_{\mathrm{sem}}^{\mathrm{en}}(\cdot)$ can be expressed as follows:

$$\boldsymbol{s}^{\mathrm{idx}} = f_{\mathrm{sem}}^{\mathrm{en}}(\boldsymbol{x}, \mathscr{S}) = f_{\mathrm{quant}}(f_{\mathrm{SE}}(\boldsymbol{x}), \mathscr{S}) \tag{7}$$

where the output $\boldsymbol{s}^{\mathrm{idx}}$ records the indices of the selected Sebs. The semantic decoder consists of a semantic decoding module $f_{\mathrm{query}}(\cdot)$ and a downstream task-performing module $f_{\mathrm{TP}}(\cdot)$. The received $\widehat{\boldsymbol{s}}^{\mathrm{idx}}$ are first interpreted by $f_{\mathrm{query}}(\cdot)$ with the assistance of the KB, which retrieves the semantic features according to the Seb indices. Then, the semantic features are directly used by the task-performing module to execute downstream tasks; that is,

$$\boldsymbol{z} = f_{\mathrm{sem}}^{\mathrm{de}}(\widehat{\boldsymbol{s}}^{\mathrm{idx}}, \mathscr{S}) = f_{\mathrm{TP}}(f_{\mathrm{query}}(\widehat{\boldsymbol{s}}^{\mathrm{idx}}, \mathscr{S})) \tag{8}$$

where $\boldsymbol{z}$ is the task-performing result and $f_{\mathrm{sem}}^{\mathrm{de}}(\cdot)$ represents the semantic decoder.

Since the accuracy and efficiency of SR vary with the granularity of Sebs, by considering the specific requirements of the

communication intent, a balance can be achieved through the comprehensive application of Sebs with different granularities. Fig. 4 illustrates a SemCom system in which images are transmitted for classification. Fine-grained Sebs can be adopted to represent the key features necessary for distinguishing between categories, such as the facial and body patterns of a giraffe, which ensures overall accuracy. Meanwhile, coarse-grained Sebs can be used for less critical features, such as the trees and the sky in the background, thereby increasing the efficiency of the communication process.

2.4. Channel encoder/decoder

It should be noted that, while Sem-codec determines the encoding efficiency of the source, the transmission accuracy needs to be further considered in practical applications. In general, the transmission process of the encoded symbols $\boldsymbol{s}^{\text{idx}}$ through the channel codec can be expressed as follows:

$$\begin{aligned} \boldsymbol{y} &= f_{\text{chan}}^{\text{en}}(\boldsymbol{s}^{\text{idx}}, \mathscr{I}) \\ \hat{\boldsymbol{s}}^{\text{idx}} &= f_{\text{chan}}^{\text{de}}(\hat{\boldsymbol{y}}, \mathscr{I}) \end{aligned} \tag{9}$$

where $f_{\text{chan}}^{\text{en}}(\cdot)$ and $f_{\text{chan}}^{\text{de}}(\cdot)$ denote the channel encoding and decoding modules, respectively; $\mathscr{I} = \{I_1^1(t), I_2^1(t), \ldots\}$ represents the set of importance levels of Sebs; $\boldsymbol{y}$ denotes the channel encoding result of $\boldsymbol{s}^{\text{idx}}$; and $\hat{\boldsymbol{y}}$ is the impaired version of $\boldsymbol{y}$ after transmission over the physical channel. Similar to the design in the Sem-codec, semantic information with different levels of importance have differing requirements for accuracy: Errors in important semantics must be minimized, whereas less-important semantics have a certain level of error tolerance in transmission. More specifically, elements in $\boldsymbol{s}^{\text{idx}}$ that are assigned different importance levels should be encoded with different channel-coding strategies, where the UEP mechanism can be naturally introduced. For Seb-based SemCom systems, UEP can be implemented at two levels—namely, message-wise and Seb-wise UEP.

Fig. 5 illustrates the merits of Seb-wise and message-wise UEP, taking image transmission as an example. Suppose an image contains multiple messages, such as a bear, a bird, and the background in the image. These messages are first mapped to the semantic feature space and represented as symbols by multiple Sebs. Then, these symbols are mapped to the code space, and the mapping process reflects two UEP mechanisms. The message-wise UEP considers the case in which messages have varying importance based on downstream tasks. As shown in the importance heatmap, symbols corresponding to the bear and bird are mapped to larger encoding spheres in the code space, as they are more important than the messages corresponding to the background. The Seb-wise UEP further considers the case in which different Sebs representing a message have distinct levels of importance for a certain downstream task. As illustrated in Fig. 5, the blue/orange regions marked in rectangles are distinctive and crucial for classification. Accordingly, the corresponding blue/orange Sebs are assigned a higher level of importance, necessitating prioritized protection. More specifically, the code space is divided into two subspaces, where the Seb-wise UEP mechanism ensures that the Seb corresponding to the bear (i.e., the region marked in the blue rectangle) only appears in the upper half of the code space, while the Seb corresponding to the bird (i.e., the region marked in the orange rectangle) only appears in the lower half of the subspace, ensuring a large distance to avoid confusion between them. Notably, the two levels of UEP can be implemented simultaneously, as they protect semantic information that is crucial for downstream tasks from different perspectives, thereby increasing the robustness of the entire SemCom system.

By employing explicit Sebs and UEP mechanisms, the proposed framework establishes a connection between messages and the knowledge within the KB, which ensures the accurate transmission of critical semantic information, thereby meeting the demands of various intents with greater flexibility. The entire process of KB initialization and communication is summarized in Algorithm 1.

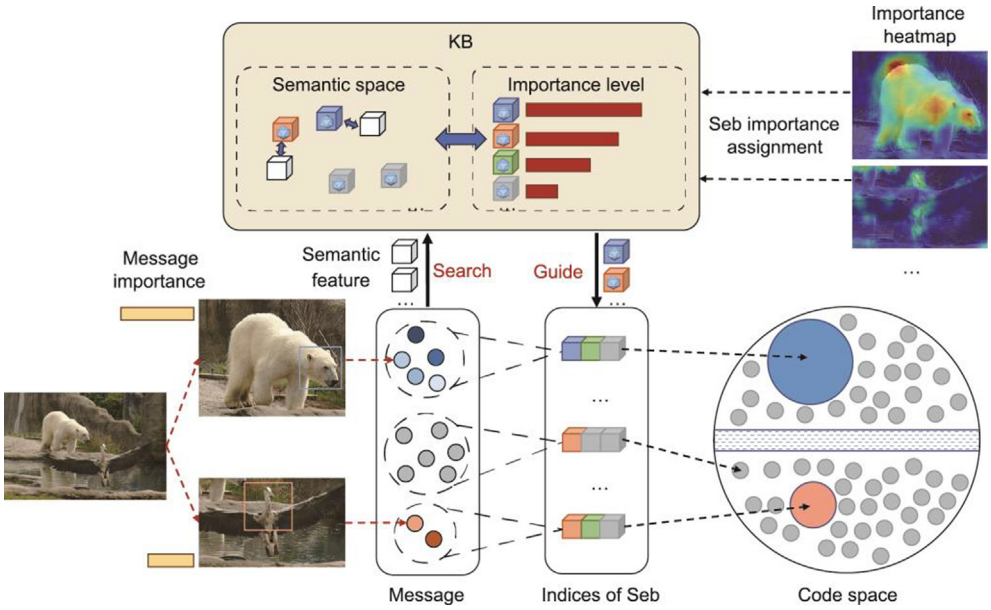

Fig. 5. The Seb-wise and message-wise UEP strategies.

Algorithm 1. KB initialization and communication process in the Seb-based SemCom framework.

KB initialization:
Input: Accessible message set $\mathscr{X}$
1. Initialize the KB with hierarchical Sebs $\langle \mathscr{S}, \prec \rangle$ according to Eqs. (1) and (2), and attach the set of importance levels of Sebs $\mathscr{I}$ according to Eqs. (3) and (4)
2. Synchronize the KB among the UEs

Communication:
Input: Accessible message set $\mathscr{X}$**, message to be transmitted at UE** $\boldsymbol{x}$**, indicator of update request** $R = 0$
3. UE estimates accuracy of SR according to Eq. (5) for $\boldsymbol{x}$ and sets update request $R = 1$ if SR accuracy $<$ predefined threshold
4. UE performs semantic/channel encoding according to Eqs. (7) and (9) and sends the message
5. **If** update request $R = 1$ **do**
6. Generate new candidate Sebs $\boldsymbol{s}$ from $\mathscr{X}$ according to Eqs. (1) and (2)
7. **If** $D(\boldsymbol{s}, \mathscr{S}) >$ predefined threshold **do**
8. Integrate new Sebs $\boldsymbol{s}$ into the KB
9. **End if**
10. Adjust $\mathscr{I}$ according to Eq. (6)
11. Remove obsolete Sebs
12. Synchronize the KB among the UEs
13. **End if**

3. Protocol of SemComs with explicit Sebs

In this section, a protocol stack of the proposed Seb-based SemCom is proposed, which is designed to be highly compatible with the existing 3GPP protocol stack, as shown in Fig. 6. To ensure the system can achieve both intelligence and high efficiency, a semantic intelligence (SI) plane integrating the KB is introduced, where Sebs and their corresponding importance information (as discussed in Section 2) are incorporated into the KB. The introduction of the SI plane establishes interactions between the application and physical layers of the protocol stack, thereby extending the functionality of the 3GPP protocol stacks to process information at the semantic level.

In the uplink process, the message $\boldsymbol{x}$ is first processed by the semantic encoder at the UE's application layer, where the elements of the output $\boldsymbol{s}^{\text{idx}}$ are then assigned varying levels of importance. During the semantic encoding process, the SI plane provides Sebs for quantifying raw semantics and synchronizes the importance information with the semantic encoder, enabling the partitioning of $\boldsymbol{s}^{\text{idx}}$ based on Seb or message importance. Despite being implemented via an NN, the output $\boldsymbol{s}^{\text{idx}}$ can be represented by bit sequences corresponding to the indices of the Sebs. As a result, the encoded bit sequences can be directly processed by existing protocols such as the service data adaptation protocol (SDAP), packet data convergence protocol (PDCP), and media access control (MAC) for purposes of transmission—including quality-of-service (QoS) mapping, encryption, and scheduling—before the final generation of physical signals. To achieve UEP for semantic flows according to their level of importance, the interaction between the SI plane and the physical layer is taken into consideration. The SI plane provides importance-related information for the upper-layer packets, thereby matching each packet with an appropriate modulation and coding scheme (MCS) and prioritizing the transmission of more critical semantics.

Meanwhile, the next-generation NodeB (gNB), which shares the same underlying protocol structure as the UE, decodes the received signals and further forwards the decoded packets to the target destinations. The uplink transmission procedure with the proposed Seb-empowered protocol is summarized in Algorithm 2 for clarity.

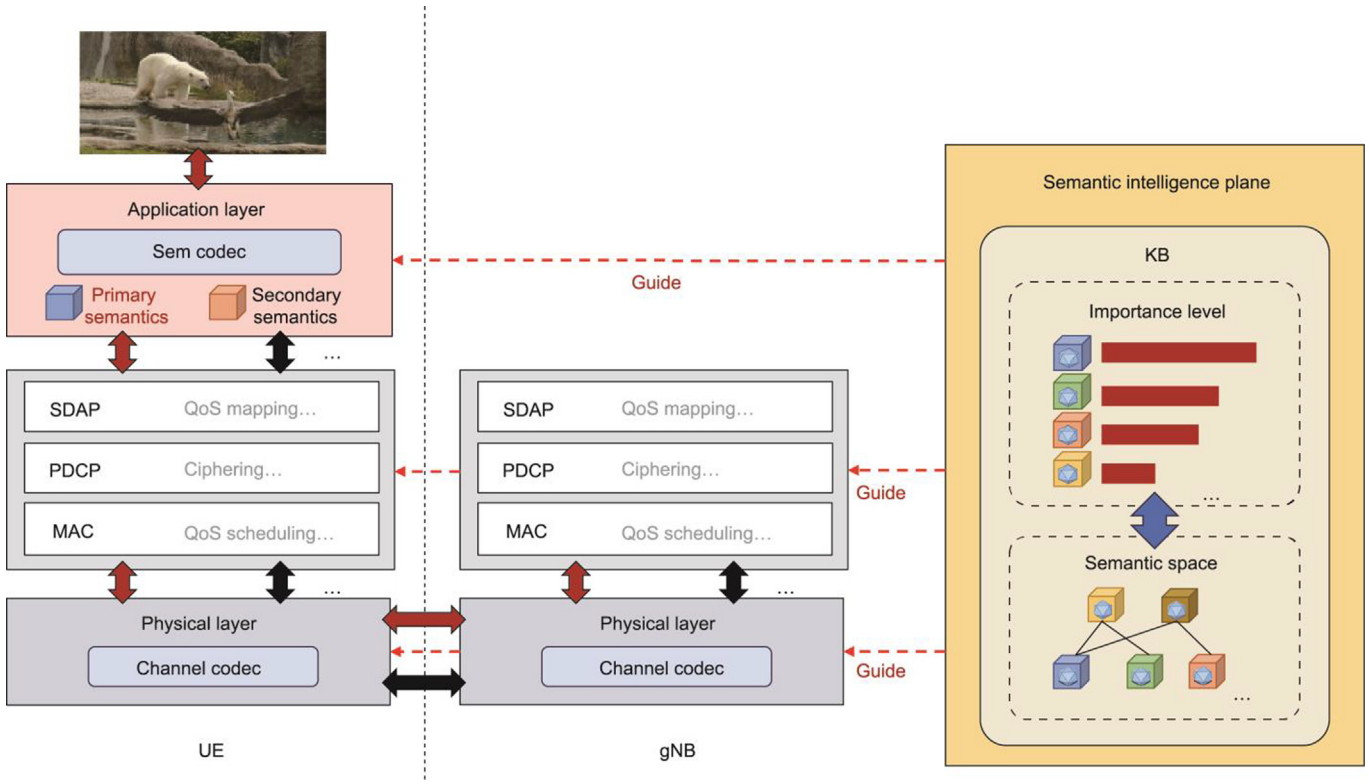

Fig. 6. Protocol of SemComs with explicit Sebs. SDAP: service data adaptation protocol; PDCP: packet data convergence protocol; MAC: media access control; QoS: quality-of-service; gNB: next-generation NodeB.

Similarly, for downlink scenarios, the SI plane provides Sebs and the corresponding importance-related information for both the gNB and the UE, enabling semantic interpretation of the received signals for the specific intelligent task.

Algorithm 2. Uplink transmission procedure with the proposed SemCom protocol stack.

UE processing:

1. **Application layer:** Encode message $\boldsymbol{x}$ following Section 2.3 with Sebs provided by SI plane, obtaining $\boldsymbol{s}^{\text{idx}}$ according to Eq. (7)
2. **SDAP layer:** Perform QoS mapping according to importance levels provided by SI plane
3. **PDCP layer:** Cipher upper-layer packets
4. **MAC layer:** Perform mapping between logical channels to transport channels
5. **Physical layer:** Encode upper-layer packets into transmitting data steams using channel encoder proposed in Section 2.4 according to Eq. (9), where different MCS are adopted for packets with different importance levels

gNB processing:

6. **Physical layer:** Decode received signals following Section 2.4, where packets with different importance levels are correspondingly decoded according to Eq. (9)
7. **MAC layer:** Perform mapping from transport channels to logical channels
8. **PDCP layer:** Decipher data packets
9. **SDAP layer:** Resolve data packets of different QoS flows and prepare the forwarding packet to destination gNB

4. Case study: Seb-based SemCom for image transmission

In this section, a case study of an image-transmission task is carried out to illustrate the performance of the proposed Seb-based SemCom system. The implementation methodology is first introduced; then, the simulation settings and simulation results are presented.

4.1. Implementation of the SemCom system

We consider a scenario in which images are transmitted by UEs to the AP, while the AP maintains and synchronizes the KB with the UEs by analyzing the previously received images. The NN of the SemCom system is first trained as an initialization process. Specifically, the semantic extraction module and the task-performing module are implemented by a convolutional neural network (CNN). For semantic extraction, each image is first divided into patches and then processed by four convolution layers with generalized divisive normalization (GDN) [25] as the activation function to obtain a collection of raw semantics. For task performance, the recovered semantics are inversely processed by four deconvolution layers with inverse GDN (IGDN) [25] to obtain patch-wise image reconstructions, which are then concatenated [26]. The system is jointly trained on the ImageNet training dataset [27] with 50 000 images being randomly sampled. The learning rate is set to 10^{-4} and 10^{-5} each for one epoch and is then fixed at 10^{-6} until the NN converges.

To realize a hierarchical representation, we consider two different levels of patch size—namely, $\{32 \times 32, 16 \times 16\}$—to generate coarse and fine-grained Sebs, respectively. For semantic encoding, the top 20% of the most important regions of the transmitted image are encoded by fine-grained Sebs, and the remaining 80% of the image regions are encoded by coarse-grained Sebs. For channel encoding, we adopt message-wise UEP. Sebs that represent the top 50% of the most important regions of the image are transmitted at a lower rate (i.e., using 1/2 rate low-density parity-check (LDPC) and 4-quadrature amplitude modulation (QAM)) to ensure higher accuracy. Meanwhile, Sebs representing the remaining 50% of image regions are transmitted at a higher rate (i.e., using 2/3 rate LDPC and 4-QAM). The test is conducted on additive white Gaussian noise (AWGN) channels. The engineered codec better portable graphics (BPG), JPEG2000, and the SemCom methods [7] (referred to as "JSCC") and [28] (referred to as "VQSemCom") are adopted for comparison. To demonstrate the impact of the UEP in our proposed method, all comparison schemes except JSCC directly adopt a high-rate transmission scheme (i.e., using 2/3 rate LDPC and 4-QAM), and the JSCC method is directly trained under an AWGN channel with the target signal-to-noise ratio (SNR; i.e., 4 dB). The channel-bandwidth ratio (CBR) is set equally across all methods by adjusting the source coding rate accordingly for fair comparison. The learned perceptual image patch similarity (LPIPS) metric [29] is primarily adopted to evaluate the quality of the reconstructed images, being defined as the average Euclidean distance between feature maps. The LPIPS effectively reflects human perception, with lower values indicating greater perceptual similarity. The peak signal-to-noise ratio (PSNR) and multi-scale structural similarity (MS-SSIM) [30] results are also provided for reference, where higher values indicate better results.

Notably, the proposed Seb-based SemCom system performs well when the testing dataset differs from the training dataset without updating the parameters of the NN. To verify this property, a validation dataset is constructed by sampling images from various distributed datasets. The dataset consists of 30 image subsets, with the first ten subsets randomly sampled from the Cityscapes [31] test set and the middle and last ten subsets randomly sampled from the Bosphorus and YachtRide sequences from the Ultra Video Group (UVG) dataset [32], respectively. Each subset contains 30 −50 images, each with 2048×1024 pixels. The subsets are transmitted sequentially to mimic cases in which the intent of image transmission changes over time. According to the methodology in Section 2.2, the KB of our method is updated after the transmission of the first, 11th, and 21st subsets; the performance will be validated in Section 4.2.

4.2. Simulation results

A visual comparison of the proposed Seb-based SemCom system with the baselines is shown in Fig. 7, where the first row shows the transmitted image and its corresponding heatmap. The second row shows the visualization and LPIPS results under channel conditions that ensure error-free transmission. The third row shows the corresponding results under poor channel conditions, when transmission errors are inevitable. The CBR of the proposed method, the BPG method, the JSCC method, and the VQSemCom method are set equally.

More specifically, the second row in Fig. 7 demonstrates the results when the SNR is 4 dB, allowing all methods to ensure error-free transmissions. The proposed method (referred to as "ours") utilizes fine-grained Sebs to more precisely represent the visually important foreground areas, such as the ship and the splashing waves, while coarse-grained Sebs ensure efficient representation of the background elements, such as the mountains in the transmitted image. Consequently, the proposed method achieves superior perceptual quality compared with the BPG, JSCC, and VQSemCom methods. The "ours w/o update" scheme refers to the Seb-based framework in the absence of KB updates. Under this

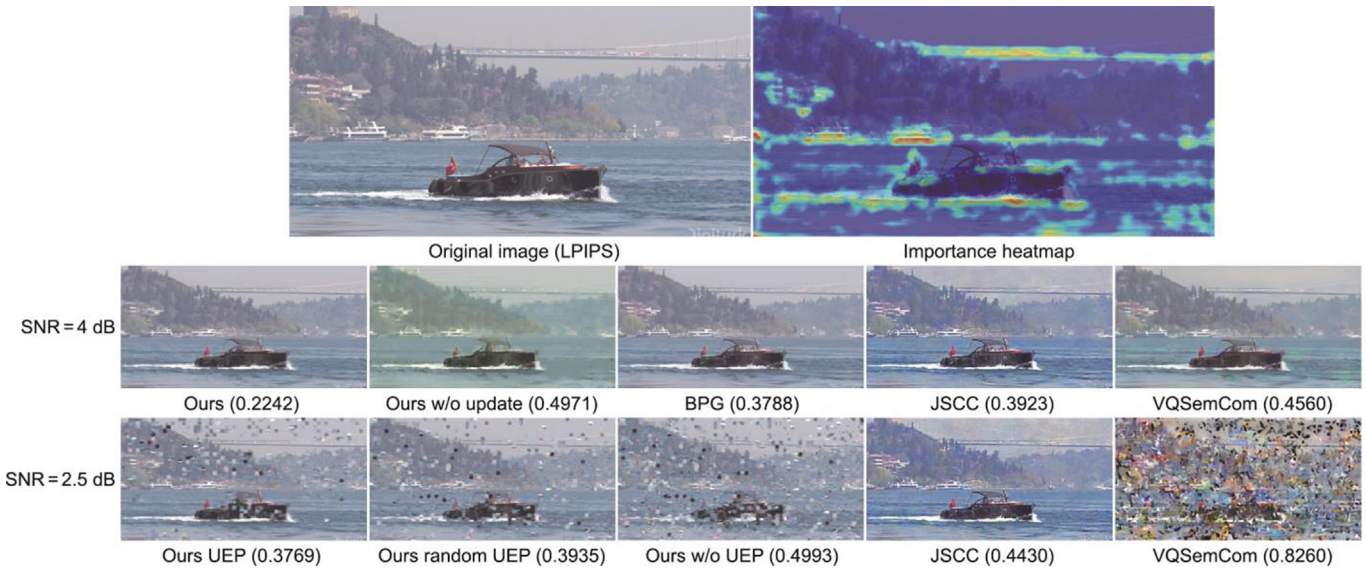

Fig. 7. A visual comparison of image reconstruction by different methods. w/o: without.

condition, distortions in color and detail can be observed due to the mismatch between the Sebs and the current communication intent. It should be noted that we study an extreme case, in which there is a large distinction between the transmission intents, to elucidate the impact of KB synchronization on system performance. In practical applications, intents generally do not undergo such distinct changes, resulting in only a gentle degradation of performance, which will be further illustrated in Section 4.3.

The third row in Fig. 7 illustrates the results when SNR is set as 2.5 dB, where transmission errors become inevitable in this case. The BPG method fails to reconstruct the image under this condition and thus is not shown in the figure. In contrast, the SemCom methods demonstrate better robustness against transmission errors. Notably, the JSCC method exhibits the highest robustness, with an LPIPS degradation of 0.05. This robustness is attributed to the joint design of source and channel coding. The VQSemCom method, which employs separate source and channel coding, suffers more severe noise impact, barely revealing the outline of the scene. Our method (referred to as "ours UEP") effectively mitigates noise impact by emphasizing the protection of the critical image regions highlighted in the heatmap; thus, it maintains an advantage over the JSCC and VQSemCom methods in LPIPS. In contrast, randomly protecting regions (referred to as "ours random UEP") or transmitting the entire image at a higher rate (referred to as "ours w/o UEP") results in a more severe noise impact, indicating the effectiveness of the explicit UEP strategies proposed in Section 2.4.

Fig. 8 shows the impact of the SNR on the proposed Seb-based SemCom system. It can be observed that, when the SNR reaches 4 dB and above, all schemes achieve error-free transmission. The proposed method provides more desirable image perceptual quality, with an average improvement of more than 20% in LPIPS compared with both traditional and SemCom counterparts at the same transmission cost. When the SNR falls below 3.5 dB, the high-rate transmission scheme fails to ensure error-free transmission, leading to a performance decline of differing degrees across all methods. Notably, the proposed Seb-based method with UEP applied (denoted as "ours UEP") generally outperforms the counterpart that applies UEP randomly to the image (denoted as "ours random UEP"), demonstrating the effectiveness of the importance-driven UEP mechanism. The reason for this effectiveness is that the UEP mechanism enables the Seb-based method to provide additional protection to pivotal parts of the message through the low-rate transmission scheme, further increasing the robustness against channel noise. Moreover, it can be seen that the performance of the BPG and JPEG2000 methods deteriorate more rapidly, showing a "cliff effect" in terms of LPIPS. The VQSemCom method exhibits better robustness compared with traditional methods, as it does not use entropy coding in its encoding process, thereby avoiding further performance degradation brought by error propagation.

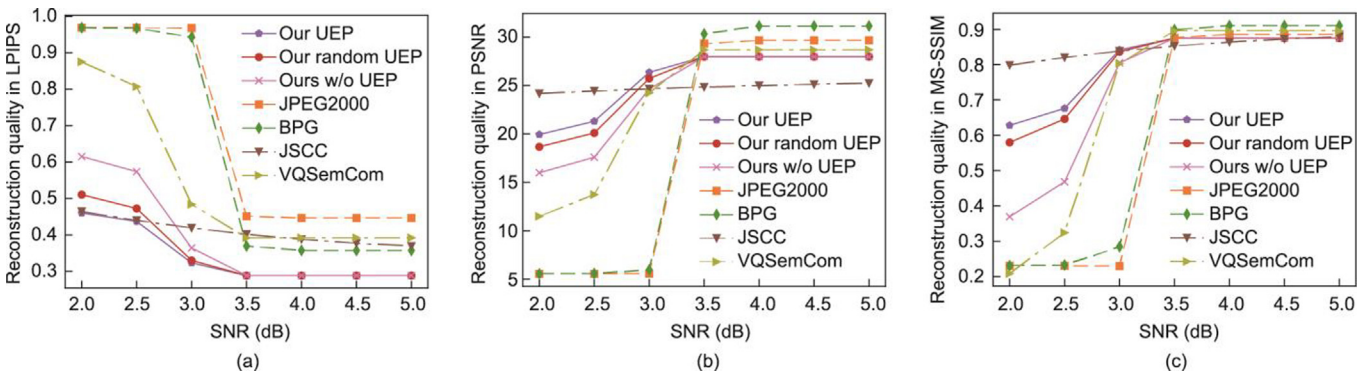

Fig. 8. Overall performance under different SNRs in an AWGN channel: (a) LPIPS under different SNRs, (b) PSNR under different SNRs, and (c) MS-SSIM under different SNRs.

However, its performance still shows a significant decline. Notably, the Seb-based method, which also uses the same high-rate transmission scheme (denoted as "ours w/o UEP"), despite experiencing severe transmission errors, exhibits better performance than the SemCom baselines, highlighting the advantage of the Seb-based Sem-codec. The JSCC method shows the best robustness against noise, which is attributed to its joint design methodology. However, the upper-bound of its performance under good channel conditions is less competitive. The performance is also evaluated in PSNR and MS-SSIM, which show similar trends as in LPIPS. However, it is interesting to observe that the traditional image coding method BPG generally performs better than all SemCom methods in terms of PSNR and MS-SSIM in a high-SNR regime. This result may occur because the two evaluation metrics focus more on the syntactic aspects of images. In the low-SNR regime, the SemCom systems show better performance, indicating their advantage of robustness under harsh transmitting conditions.

4.3. Impact of KB updating

Fig. 9 exhibits the effectiveness of KB updating. The test is conducted under an SNR of 4 dB, ensuring error-free transmissions for the ablation study. The transmission intents change at the 11th and 21st subsets, as depicted in Section 4.1. For comparison, the Seb-based system without a KB update is also evaluated, shown as the solid line marked with X's. It is obvious that the performance degrades once the communication intents change. This occurs because the mismatch between the Sebs in the KB and the changed communication intent results in a poor SR performance in semantic encoding. However, it should be noted that a KB update will be triggered once the SR performance falls below a predefined threshold, as discussed in Section 2.2; the Seb-based SemCom can then quickly evolve without retraining the NN, as shown in Fig. 9, providing an easier deployment strategy in practical settings.

4.4. Impact of Seb granularity

To demonstrate the impact of Seb granularity on the performance of intelligent tasks, an additional experiment is conducted on the Seb-based SemCom system with respect to an image classification task. As shown in Fig. 10, different ratios of fine-grained Sebs are adopted to represent the image semantics. The experiment is conducted using the ImageNet [27] validation dataset, which contains 50 image samples per class. Within the dataset, 25 images are randomly selected for Seb generation, while the remaining 25 are used for testing. A ResNet-50 classifier [33] is introduced to further perform the image classification tasks based on the reconstructed images. The basic setting of the framework follows Section 4.1, except that a uniform 1/2 rate LDPC and 4-QAM are adopted for channel coding to ensure reliable transmission. The ratios of fine-grained Sebs are set as {0, 0.2, 0.5, 0.8, 1.0}, with the corresponding CBR being {0.32, 0.51, 0.79, 1.08, 1.27}, represented by the bars from left to right in Fig. 10. As the proportion of fine-grained Sebs increases, higher encoding and transmission overheads are incurred.

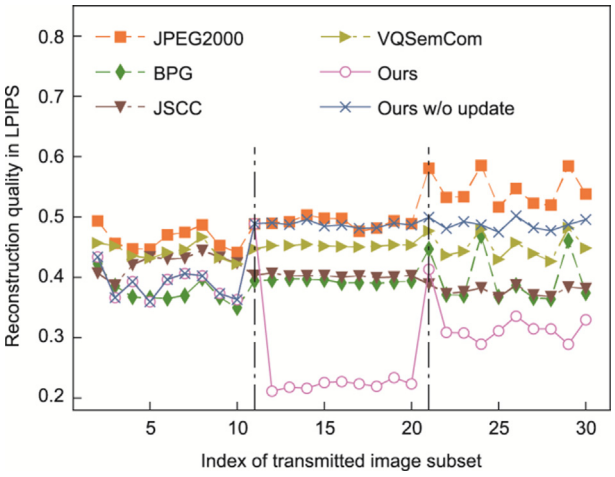

Fig. 9. Impact of KB updating, where the communication intent change at the 11th and 21st subsets.

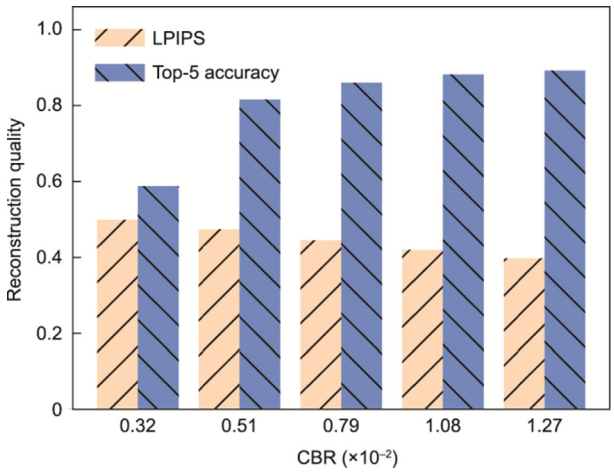

Fig. 10. Performance of the Seb-based SemCom with varying ratios of fine-grained Sebs.

It can be observed that, as the ratio of fine-grained Sebs increases, the perceptual quality gradually improves, corresponding to the decreases in LPIPS values. This observation is consistent with the analysis in Section 2.1, where fine-grained Sebs provide more precise SR to support higher perceptual quality, at the cost of higher channel bandwidth consumption. The top-5 accuracy with respect to the classification task also increases with the incremental ratio of fine-grained Sebs. However, there is an evident saturation floor with respect to the granularity, which is distinct from the reconstruction task. To be specific, the top-5 accuracy rises from 60% to 80% as the proportion of fine-grained Sebs grows from 0 to 20%, and the performance is gradually saturated, such that additional increments in the proportion of fine-grained Sebs lead to more moderate gains.

The distinct patterns regarding the impact of granularity for the two different tasks are in line with the analysis in Section 2.1. For the classification task, critical regions within the image contain the key features of objects. The precision of the representation of these critical regions directly determines the accuracy of the classification task. Once the proportion of fine-grained Sebs effectively meets the representation requirements of these regions, the additional increment in fine-grained Sebs no longer contributes to the intelligent task (i.e., classification in this case). As a result, the performance of the intelligent task shows a saturation floor with respect to Seb granularity, where finer-grained Sebs lead to a higher CBR, as shown in Fig. 9. In contrast, the reconstruction tasks require detailed semantics across the image; thus, the usage of increasingly finer-grained Sebs leads to consistent improvement in perceptual quality.

Overall, it is obvious that Seb granularity impacts the performance of intelligent tasks, where finer-grained Sebs lead to a better performance. However, finer-grained Sebs also lead to a larger communication overhead—that is, a larger CBR in the experiment. Moreover, the choice of Seb granularity is related to the communication intent—that is, downstream tasks in the system. As a result, the efficiency and overall performance of the Seb-based SemCom

can be adaptively controlled by adjusting the Seb granularity in practical use, which is significant for the flexible deployment of the Seb-based SemComs.

5. Conclusions

In this article, we proposed an explicit Seb-based SemCom architecture for interpretable and flexible semantic-level communications for 6G. First, a mathematical model of Sebs was introduced, followed by descriptions of the designed criteria for the KB, Sem-codec, and channel-codec. A novel Seb-based protocol stack was further proposed, considering the compatibility of current CIT-based networks, with the introduction of an SI plane to incorporate the KB for ubiquitous intelligence and self-evolution for 6G networks. A case study of an image transmission task was conducted, in which the proposed architecture functioned robustly regarding changes in communication intents, obtaining an average improvement of 20% in the LPIPS compared with the BPG, JPEG2000, JSCC, and VQSemCom baselines. An ablation study showed that the efficiency and overall performance of the proposed Seb-based SemComs can be effectively controlled by adjusting the Seb granularity, demonstrating great flexibility in deployment for practical use.

CRediT authorship contribution statement

Fengyu Wang: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Yuan Zheng:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Wenjun Xu:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Junxiao Liang:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis. **Ping Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Zhu Han:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62293485 and 62301069) and the Fundamental Research Funds for the Central Universities (2025TSQY11).

References

[1] Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–423.
[2] Yang W, Du H, Liew ZQ, Lim WYB, Xiong Z, Niyato D, et al. Semantic communications for future internet: fundamentals, applications, and challenges. IEEE Commun Surv Tutor 2023;25:213–50.
[3] Niu K, Dai J, Yao S, Wang S, Si Z, Qin X, et al. A paradigm shift toward semantic communications. IEEE Commun Mag 2022;60:113–9.
[4] Yang W, Liew ZQ, Lim WYB, Xiong Z, Niyato D, Chi X, et al. Semantic communication meets edge intelligence. IEEE Wirel Commun 2022;29:28–35.
[5] Pokhrel SR, Choi J. Understand-before-talk (UBT): a semantic communication approach to 6G networks. IEEE Trans Veh Technol 2023;72:3544–56.
[6] Zhang P, Xu W, Liu Y, Qin X, Niu K, Cui S, et al. Intellicise wireless networks from semantic communications: a survey, research issues, and challenges. IEEE Commun Surv Tutor 2025;27(3):2051–84.
[7] Bourtsoulatze E, Kurka DB, Gündüz D. Deep joint source–channel coding for wireless image transmission. IEEE Trans Cogn Commun Netw 2019;5:567–79.
[8] Xie H, Qin Z, Li GY, Juang BH. Deep learning enabled semantic communication systems. IEEE Trans Signal Process 2021;69:2663–75.
[9] Li A, Wei X, Wu D, Zhou L. Cross-modal semantic communications. IEEE Wirel Commun 2022;29:144–51.
[10] Tong W, Li GY. Nine challenges in artificial intelligence and wireless communications for 6G. IEEE Wirel Commun 2022;29:140–5.
[11] Ren J, Zhang Z, Xu J, Chen G, Sun Y, Zhang P, et al. Knowledge base enabled semantic communication: a generative perspective. IEEE Wirel Commun 2024;31:14–22.
[12] Yi P, Cao Y, Kang X, Liang YC. Deep learning-empowered semantic communication systems with a shared knowledge base. IEEE Trans Wirel Commun 2023;23:6174–87.
[13] Zhang H, Shao S, Tao M, Bi X, Letaief KB. Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data. IEEE J Sel Areas Commun 2023;41:170–85.
[14] Luo X, Chen HH, Guo Q. Semantic communications: overview, open issues, and future research directions. IEEE Wirel Commun 2022;29:210–9.
[15] Niu K, Zhang P. A mathematical theory of semantic communication. J Commun 2024;45:7–59.
[16] Dai J, Zhang P, Niu K, Wang S, Si Z, Qin X. Communication beyond transmitting bits: semantics-guided source and channel coding. IEEE Wirel Commun 2022;30:170–7.
[17] Fu Y, Cheng W, Zhang W, Wang J. Scalable extraction based semantic communication for 6G wireless networks. IEEE Commun Mag 2023;62:96–102.
[18] Ballé J, Chou PA, Minnen D, Singh S, Johnston N, Agustsson E, et al. Nonlinear transform coding. IEEE J Sel Top Signal Process 2020;15:339–53.
[19] Judd T, Ehinger K, Durand F, Torralba A. Learning to predict where humans look. In: Proceedings of the IEEE International Conference on Computer Vision; 2009 Sep 29–Oct 2; Kyoto, Japan. New York City: IEEE; 2009. p. 2106–13.
[20] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision; 2017 Oct 22–29; Venice, Italy. New York City: IEEE; 2017. p. 618–26.
[21] Zhang G, Yang P, Cai Y, Hu Q, Yu G. From analog to digital: multi-order digital joint coding-modulation for semantic communication. 2024. arXiv:2406.05437.
[22] Tung TY, Kurka DB, Jankowski M, Gündüz D. DeepJSCC-Q: constellation constrained deep joint source–channel coding. IEEE J Sel Areas Inf Theory 2022;3(4):720–31.
[23] Zhang P, Xu W, Gao H, Niu K, Xu X, Qin X, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks. Engineering 2022;8:60–73.
[24] Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al. DINOv2: learning robust visual features without supervision. 2024. arXiv:2304.07193.
[25] Ballé J, Laparra V, Simoncelli EP. Density modeling of images using a generalized normalization transformation. 2016. arXiv:1511.06281.
[26] Zheng Y, Wang F, Xu W, Pan M, Zhang P. Semantic communications with explicit semantic base for image transmission. In: Proceedings of the IEEE Global Communications Conference; 2023 Dec 4–8; Kuala Lumpur, Malaysia. New York City: IEEE; 2023. p. 4497–502.
[27] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large-scale visual recognition challenge. Int J Comput Vis 2015;115:211–52.
[28] Fu Q, Xie H, Qin Z, Slabaugh G, Tao X. Vector quantized semantic communication system. IEEE Wirel Commun Lett 2023;12:982–6.
[29] Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–22; Salt Lake City, UT, USA. Piscataway: IEEE; 2018. p. 586–95.
[30] Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: Proceedings of the 37th Asilomar Conference on Signals, Systems & Computers; 2003 Nov 9–12; Pacific Grove, CA, USA. New York City: IEEE; 2003. p. 1398–402.
[31] Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, et al. The Cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. New York City: IEEE; 2016. p. 3213–23.
[32] Mercat A, Viitanen M, Vanne J. UVG dataset: 50/120 fps 4K sequences for video codec analysis and development. In: Proceedings of the 11th ACM Multimedia Systems Conference; 2020 Jun 8–11; Istanbul, Türkiye. New York City: Association for Computing Machinery; 2020. p. 297–302.
[33] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. New York City: IEEE; 2016. p. 770–8.