



Research
6G Requirements, Vision, and Enabling Technologies—Article

Toward Wisdom-Evolutionary and Primitive-Concise 6G: A New Paradigm of Semantic Communication Networks



Ping Zhang^{a,b}, Wenjun Xu^{c,*}, Hui Gao^d, Kai Niu^c, Xiaodong Xu^a, Xiaoqi Qin^a, Caixia Yuan^e, Zhijin Qin^f, Haitao Zhao^g, Jibo Wei^g, Fangwei Zhang^h

^aState Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

^bPeng Cheng Laboratory, Shenzhen 518066, China

^cKey Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

^dKey Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

^eSchool of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

^fThe School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK

^gCollege of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China

^hSchool of Humanities, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:

Received 4 January 2021

Revised 31 May 2021

Accepted 1 September 2021

Available online 17 November 2021

Keywords:

6G

Semantic information

Semantic communication

Intelligent communication

ABSTRACT

The sixth generation (6G) mobile networks will reshape the world by offering instant, efficient, and intelligent hyper-connectivity, as envisioned by the previously proposed Ubiquitous-X 6G networks. Such hyper-massive and global connectivity will introduce tremendous challenges into the operation and management of 6G networks, calling for revolutionary theories and technological innovations. To this end, we propose a new route to boost network capabilities toward a wisdom-evolutionary and primitive-concise network (WePCN) vision for the Ubiquitous-X 6G network. In particular, we aim to concretize the evolution path toward the WePCN by first conceiving a new semantic representation framework, namely semantic base, and then establishing an intelligent and efficient semantic communication (IE-SC) network architecture. In the IE-SC architecture, a semantic intelligence plane is employed to interconnect the semantic-empowered physical-bearing layer, network protocol layer, and application-intent layer via semantic information flows. The proposed architecture integrates artificial intelligence and network technologies to enable intelligent interactions among various communication objects in 6G. It features a lower bandwidth requirement, less redundancy, and more accurate intent identification. We also present a brief review of recent advances in semantic communications and highlight potential use cases, complemented by a range of open challenges for 6G.

© 2021 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

Human society is undertaking the fourth industrial revolution, driven mainly by the convergence of the digitalization of everything, information and communication technologies (ICTs), and artificial intelligence (AI) innovations. ICT plays a vital role in the evolution of society toward an intelligent and digital era. With the rolling out of the fifth generation (5G) mobile networks, 5G is opening up a new paradigm for the Internets of humans, machines, and things, enabled by the orchestration of ubiquitous

communications, computing, and control (UC³) capabilities [1]. The vision of the sixth generation (6G) mobile networks is to reshape the world by offering instant, efficient, and intelligent hyper-connectivity between the physical world and the digital world. This trend opens up a new era for mobile communications, in which the scope of communications will span both the physical and digital worlds.

Research on 6G began recently, with a focus on innovative network architectures and key technologies [2]. It is notable that the hyper-massive and cross-world connectivity envisioned by 6G presents tremendous challenges in network operation and management, as we mentioned in our proposed Ubiquitous-X 6G framework [1]. In addition to the conventional human-machine-

* Corresponding author.

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

thing architecture, a new type of communication object, the *genie*, is introduced in this framework to bridge the physical and digital worlds. As the AI-empowered super assistant for physical communication objects, the genie can accurately identify intents and handle complicated information processes beyond the experience and capacity of physical communication objects. Moreover, it aggregates and extracts valuable information to enable efficient intent-oriented interactions among communication objects. The features of communication objects are listed in Table 1.

The extensive spread of the Ubiquitous-X 6G is hindered by certain critical challenges, including an explosive growth of connections, a deficiency of rigid layered network protocols, and the emergence of innovative applications with diversified service requirements. Consider a scenario in which multiple communication agents interact and collaborate to complete a particular task. The interactions among these agents involve real-time sensing data exchange, information fusion, and collaborative decision-making. The massive data exchange may scale up the network complexity in terms of signaling cost and protocol overhead. We believe that the intelligence of the communication agents can be fully exploited to identify task-related information, so that the interactions among agents become concise and efficient. As illustrated in Fig. 1, the intelligence-oriented interconnection among four types of communication objects relies on semantic communications, which improve the efficiency of conventional data exchange in the 0–1 bit-stream by transmitting key semantic elements. Communication objects with situation awareness and background knowledge will extract semantic elements from their intents in order to improve transmission efficiency.

A considerable amount of literature on semantic communications has been published. Since Shannon’s masterpiece was published in 1948 [3], the Shannon limit has served as guidance in communication system design for more than seven decades. With the aid of AI, semantic-aware communication techniques are emerging. In physical layer processing, recent advances in natural language processing (NLP) and computer vision enable semantic-enhanced coding strategies [4] and end-to-end semantic transmission schemes [5,6] which boost the transmission quality of different types of sources. In the media access control/link/network protocol layers, the semantic-filtering mechanism [7] is initiated to reduce the redundancy of the layered protocols. In the application layer, semantic-based user intent identification [8] is used to automate network configurations and simplify network management. Although we are witnessing tremendous success in implementing AI-empowered and semantic-aware technologies in each protocol layer, a systematic framework is still missing. Therefore, it is critical to conceive an initial architecture to embed semantic intelligence (SI) across multiple layers for 6G.

Based on the discussion above, we introduce an intelligent and efficient semantic communication (IE-SC) architecture toward a wisdom-evolutionary and primitive-concise network (WePCN), which aims to improve the network intelligence level to enable a more efficient and concise network. Unlike the traditional network design philosophy, which upgrades network capability mainly by stacking more spectra, computation modules, denser access points, and antennas with ever-increasing complexity in multiple domains, we boost network capabilities with concise signaling originating from accumulated network wisdom. The core of IE-SC architecture is a novel SI plane, which implements semantic environment representation, background knowledge management, semantic deduction, and decision-making. Moreover, three new semantic-empowered abstract protocol layers are designed to reshape the existing protocol layers—namely, the semantic-empowered physical-bearing (S-PB) layer, the semantic-empowered network protocol (S-NP) layer, and the semantic-empowered application-intent (S-AI) layer. The SI plane coordinates the three layers via the semantic information flow (S-IF), which carries the application intent and semantic information across the network. Upon receiving the S-IF, the S-NP layer can orchestrate the intent-related semantics to generate flexible and concise protocols. Working coherently with the S-NP layer, the S-PB layer can adopt appropriate joint semantic-and-syntactic coding strategies to improve the physical resource utilization toward high intent-accomplishing efficiency. In this way, the high-complexity issue confronting the Ubiquitous-X network can be resolved by the proposed IE-SC architecture. Moreover, the proposed architecture can comprehensively upgrade the network capability toward the future WePCN vision: to build an ordered, efficient, and intelligent Ubiquitous-X network for future applications and services.

This paper is organized as follows. In Section 2, we discuss related works in semantic information and communication. In Section 3, we present the IE-SC framework and the technical content of the S-PB layer with initial simulation results, and conceive the concepts and roadmaps for the S-NP and S-AI layers. We further present three promising application scenarios for the IE-SC in Section 4 and identify a range of future work ideas in Section 5. Section 6 concludes this paper.

2. Recent advances in semantic information and communications

Since it was first proposed, the concept of semantic information has been continuously refined. Early works on semantic communication follow the Shannon probability measure framework,

Table 1
Human–machine–thing–genie architecture: orchestrating the physical and digital worlds.

Communication-object	Definition/concept	Attribute	Intelligence level	Characteristic/role	Capability	Usage
Human	The main entity of the physical world	The user of communication services	High	Play a prime role in the human–machine–thing–genie loop	Cognition	<ul style="list-style-type: none"> Human-centered communications Human-X API
Machine	A device that can actively initiate the communication process	The provider/user of communication services	Moderate	Extend the physical/mental powers of humans	Communication	<ul style="list-style-type: none"> Machine-centered communications Machine-X API
Thing	An object that is passively involved in communication processes	The provider of communication services	Low	Complement the sensing power of humans	Sensing	<ul style="list-style-type: none"> Internet of Things Perceptual computing
Genie	The main entity of the digital world	The provider/user of communication services	Super	The fusing of human–machine–thing	Intelligence	<ul style="list-style-type: none"> AI-empowered super-agent of physical communication objects

API: application programming interface.

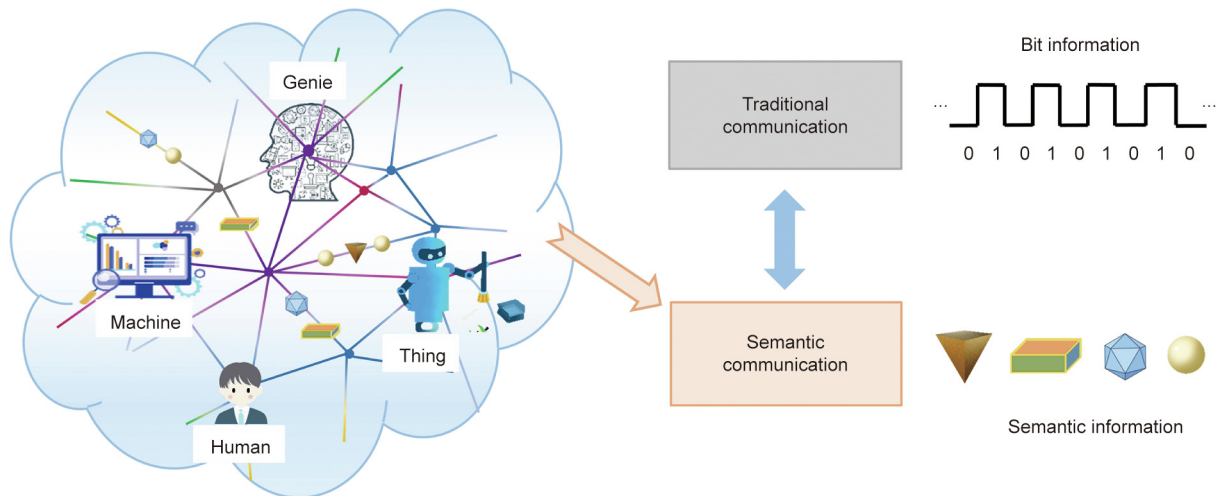


Fig. 1. Illustration of the intelligence-oriented semantic interconnection among humans, machines, things, and genies.

complemented with logical and fuzzy transformation. The recent blooming of AI-based applications offers new opportunities for the design of semantic communication systems. In this section, we review the development of the semantic information concept and recent advances in semantic communication systems.

2.1. Semantic information

The development of semantic information theory can be roughly divided into two major phases. The classic semantic information theory originated from the pre-Shannon era, is featured by its connection to Shannon-information measure and primitive natural language (NL). Meanwhile, the modern semantic information theory is mainly developed in the recent decade with more diverse views on the essence of semantic information.

2.1.1. Classic semantic information theory

The concept of semantics was initially proposed by Morris [9], who introduced the triple-definition of syntactics, semantics, and pragmatics in the theory of signs. Weaver [10] proposed a three-level communication framework and further characterized the syntactic, semantic, and pragmatic features of communications. Carnap and Bar-Hillel [11,12] outlined a theory of semantic information with propositional logic in 1953. They also used the probability measure for semantic information. Barwise and Perry [13] extended semantic information theory [11,12] to situational logic, and Floridi [14] solved the problem that contradictions cannot be measured correctly. D’Alfonso [15] employed the concept of truth-likeness to quantify semantic information to support a broader range of use cases.

2.1.2. Modern semantic information theory

Over the last decade, the concept of semantic information theory has gone beyond Carnap’s framework. For example, Zhong [16] proposed a theory of semantic information by introducing the information trinity and proved that semantic information is the unique representative of the trinity. Kolchinsky and Wolpert [17] defined semantic information as the syntactic information between a system and its environment that causally contributes to the continued existence of the system from the physical perspective. More recently, Kountouris and Pappas [18] gave a multi-granularity definition of semantic information at different levels of a communication system and used Rényi entropy [19] to measure semantic information. Jiang et al. [4] pointed out some

limitations of the current communication systems that are attributed to the lack of semantics awareness and suggested that AI can boost semantic information technology. It has been observed that modern semantic information theories offer a comprehensive view of semantic information and demonstrate its great potential to empower communication systems with AI.

2.2. Semantic communications

The core of semantic communication is to ensure the successful delivery of the meaning of information. Due to the broad nature of semantics, semantic communication techniques span multiple protocol layers in communication networks. Some recent works are briefly reviewed below.

2.2.1. Semantic-based physical-layer transmission

The classical model-based source coding and channel coding in the Shannon framework aim to recover the syntactic information accurately at the receiver—that is, ensuring accurate symbol reception. In contrast, the semantic communication process aims to accurately recover the semantic information at the destination, focusing on the information contents beyond the symbols, which introduces new features in terms of coding purposes and coding methods. Existing works [4,6,20–22] have proved that physical layer transmission efficiency can be improved by employing semantic encoding and decoding in representative application scenarios. Since semantic communication still lacks a comprehensive and unified mathematical formulation [4], existing semantic encoding and decoding modules are mainly realized with model-free machine learning methods [20–22]. Moreover, existing solutions can be roughly classified into two categories: modular design and integrated design. Modular design adds semantic encoding and decoding modules to the existing communication systems with block-wise segmentation. The semantic encoding and decoding modules realize the mutual transformation between syntactic and semantic information to support the efficient transmission of text, speech, or image. For example, a context-based decoder was integrated into a conventional communication system to reduce the decoding overhead of text [4]. As another example, a long short-term memory (LSTM) network was employed in Ref. [20] to extract the meaning of the text for semantic encoding and decoding, which further improves the performance of text transmission. The integrated design takes the route of semantic-enhanced joint source-channel coding. The semantic encoding

and decoding module and other channel coding or joint source–channel coding modules are optimized with a common objective. For example, an end-to-end semantic communication architecture was proposed in Ref. [21], which combines semantic reasoning and physical layer communication to eliminate semantic errors. A joint source–channel coding algorithm based on autoencoder architecture was proposed in Ref. [22], in which convolutional neural networks replace the conventional source/channel coding blocks. Lite semantic communication systems were proposed in Refs. [5,6], which combined joint source–channel coding with pruning strategies, achieving efficient image/text transmission and classification in Internet of Things (IoT) networks. Moreover, end-to-end semantic communication has been proposed in Refs. [5,23] for text/speech transmission, which exhibits significant performance improvement under various channel conditions.

2.2.2. Application-aware communication protocols

Existing lower-layer communication protocols are designed to support various upper-layer applications, often providing a relatively wide range of functions that may not always be relevant to specific applications. Some application-aware protocol designs have recently emerged with reduced physical resource consumption to improve goal-achieving efficiency. The dominant approach is the cross-layer protocol design. For example, efficient routing protocols are proposed to address high-mobility and dynamic topology challenges in an unmanned aerial vehicle (UAV) *ad hoc* network [24] and inter-vehicle network [25], in which the information in the lower protocol layers is directly integrated into the routing protocol to reduce the end-to-end delay. Recently, a more ambitious application-aware cross-layer protocol framework was proposed [7], wherein an application-aware semantic-filtering mechanism is conceived to support flexible protocol-function orchestration in order to reduce redundancy. Moreover, the application-aware protocol design can deal with multi-agent communications, which sheds light on new protocol designs in mobile communication networks. For example, Sukhbaatar et al. [26] proposed a learning-based multi-agent communication scheme in which the interaction strategy or protocol is autonomous via neural networks. Such a human-like communication protocol can be considered as a prototype of an autonomous protocol for intelligent communication agents with reduced resource consumption.

2.2.3. Semantic-based intent-driven networks

Current network management and control cannot autonomously capture the user's business intents and agilely generate a network to meet the user's quality of experience in fine granularity. Semantic-related techniques are essential in identifying users' intents and implementing these intents across the network to achieve intent-driven intelligent networks. To be specific, by utilizing the recent advances from AI and NLP, some progress has been made toward accurately identifying and understanding intents. For example, a stream of work in the contextual spoken language understanding method is proposed by Refs. [8,27], which can simultaneously identify intents and informative slots by capturing contextual semantics. An intent-based cloud service management framework is proposed in Ref. [28], which understands users' intents in NL and translates them into the network's resource-management language. More recently, new architectures and enabling technologies for intent-based networks have been reviewed in Ref. [29], which shows that the users' business intents can be captured with domain-specific language and an appropriate user–computer interface. Inspired by the studies described above, the design and implementation of semantic-empowered intent-driven networks demonstrate attractive potential. Accordingly, advanced semantics processing techniques are required to achieve accurate intent identification, decomposition, and representation

in general scenarios. Moreover, the intent-driven design should be implemented comprehensively across network protocol layers, combining all intent-related elements toward an integrated and agile intent-driven network.

3. IE-SC architecture

Recent progress in semantic information theory and communication techniques is relatively fragmented. Applications in different network layers are independently studied, yet a systematic design is still missing. In this section, we will further deepen the technical implications of the Ubiquitous-X 6G [1] by integrating IE-SC toward the WePCN. More specifically, we will first conceive the semantic base (Seb) and propose an architecture for the semantic-enhanced Ubiquitous-X 6G. Then we will introduce advanced semantic communications and information-processing technologies, infrastructures, and modules to enhance network capabilities and reformulate the network protocol hierarchy. We will also introduce mechanisms to enable intelligent communications among various communication objects in the Ubiquitous-X 6G.

3.1. Semantic base

It may be recalled that the concept of “Bit,” as proposed in Shannon's classical information theory, is not a unit but a representation and measurement framework of information entropy. In line with this view, we propose the concept of “Seb” as a representation framework for semantic information. In particular, Seb provides a modularized and highly abstractive method to represent semantic information, thus making semantic communication more efficient. To clarify the differences between Seb and Bit more intuitively, we refer to the process of constructing an architecture, as illustrated in Fig. 2. The original message from the transmitter is the outline of the construction. The communication system delivers this message using its predefined base. Traditional communication systems can be regarded as building the architecture or reconstructing the message at the receiver in a brick-by-brick manner. Bit serves as brick and concrete, which gives a precise representation of the original message. In contrast, the semantic communication system uses Seb, which is similar to reconstructing the architecture/message using a laminboard/integrated window or door. Such a construction is highly modulated with the aid of a material warehouse dimensioned by the Seb. Thus, the message delivery is much more efficient using the common knowledge of architecture construction/decomposition and the warehouse.

Seb may provide a new perspective to describe the complex nature of semantic information involving application intent and the form of information. From an abstract perspective, Seb serves as a representation framework. It contains multi-level transformations that extract the information's multimodal characteristics and eventually transform them into semantic elements. More specifically, Seb can contain user intent-related background knowledge, intent-knowledge mapping mechanisms, semantic-elements extraction, and expressions; the input of Seb can be the intent of communication, and the output of Seb can be a bit sequence carrying intent-related semantic elements. The intent-related background knowledge within Seb can be considered as a particular knowledge graph or other organized expressions. Taking the knowledge graph as an example, each vertex represents a semantic element, and each edge represents the correlation between two semantic elements. Moreover, semantic-elements extraction finds all possible paths corresponding to the intent-accomplishing process. Finally, semantic-elements expression is an appropriate bit

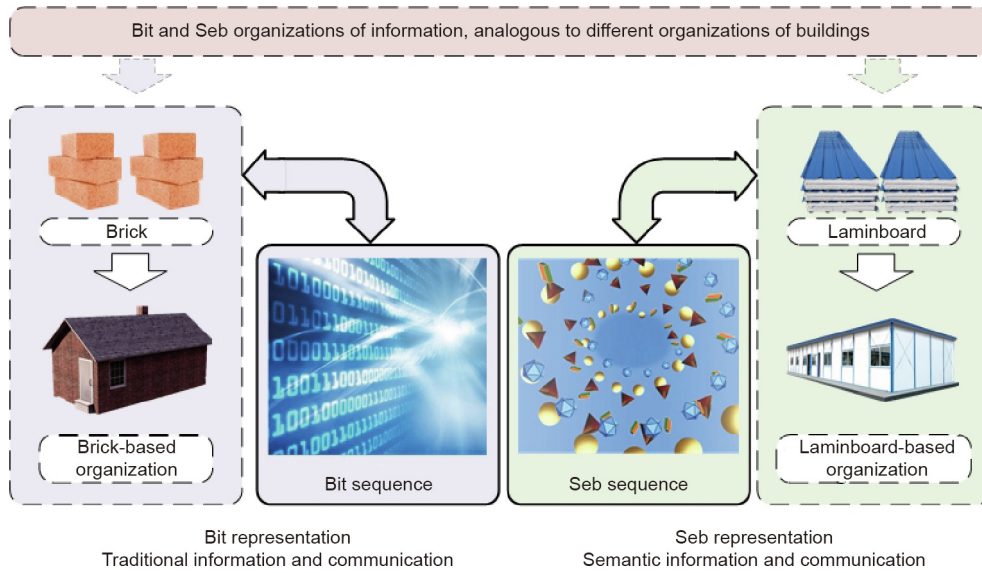


Fig. 2. A comparison between traditional communication systems and semantic communication systems.

sequence that can uniquely identify intent-oriented semantic information.

Furthermore, Seb may be further developed as a measurement framework in line with Bit. It will contain the Bit framework as a particular case and offer a multi-perspective measurement of semantic information. For example, when viewed from the physical form or the syntactic perspective of the information, the number of bits needed to carry the information can be obtained from Seb. When viewed from the application intent or the semantic perspective of the information, the semantic elements can be obtained from Seb. Therefore, Seb can be a representation and measurement framework to embrace the multimodal and multi-perspective characteristics of information. The general principles and mechanisms of Seb demand further investigation.

3.2. IE-SC architecture

3.2.1. Architecture design

In this subsection, we present a novel IE-SC architecture, which features one plane, three layers, and a set of flows, as shown in Fig. 3. More specifically, the SI plane is responsible for semantic representation, knowledge management, semantic decision, and deduction. The SI plane coordinates with the three layers:

- The S-AI layer, which identifies and decomposes user intents;
- The S-NP layer, which implements a semantic-empowered interaction protocol to support the intelligent network;
- The S-PB layer, which implements semantic-empowered message transmission in the physical layer.

The S-IF is a high-level representation of environmental information and internal information. The former includes information about the physical environment, spectrum environment, electromagnetic environment, and so forth. The latter includes network layer information, decision-making information, and other related intelligence information. The SI plane and the three semantic-empowered layers interact with each other via a set of S-IFs. A brief comparison between the proposed IE-SC and the conventional syntactic communication architecture is given in Table 2. It is notable that the IE-SC architecture requires the modification, enhancement, or replacement of existing network modules. However, these efforts can result in an advanced network with improved information transmission efficiency, management-control efficiency, and intent-achieving efficiency. It is also notable that the bottleneck

of existing networks stems from the explosive growth of communication links and data. IE-SC provides a new approach to address this challenge. In particular, with the advances of specialized chipsets and hardware, the awareness of semantics at transceivers can significantly improve communication efficiency by reducing the number of transmitted data bits while maintaining the intent of the communication.

3.2.2. SI plane

The SI plane spans all layers in the IE-SC architecture, and has the following functions:

(1) **Semantic environment representation.** The internal and external environmental information is processed by filtering and semantic extraction; then, the information is aggregated in the SI plane. After the semantic classification, the environment representation is formed. The semantic information is then embedded into the S-IF, which can flow through the SI plane and interfaces at different layers.

(2) **Background knowledge management.** Different background knowledge of different network elements and layers, such as context and environment, will affect the performance of the S-AI, S-NP, and S-PB layers. Therefore, the SI plane is responsible for coordinating the exchange of background knowledge. The SI plane can classify, integrate, and store knowledge after semantic extraction. It then shares the knowledge via S-IF.

(3) **Semantic decision and deduction.** The SI plane is capable of evaluating network capability and synthesizing user intents. More specifically, the S-AI layer feeds the decomposed user intents to the SI plane via S-IF. The SI plane then synthesizes the intents and network function to evaluate the achievable performance. It then performs decision-making for all network layers. Finally, the decisions are transmitted to the control plane to enable intent-driven transmission and networking.

3.2.3. S-PB layer

The purpose of semantic communication differs from that of conventional data communication, because semantic communication delivers meaning. To achieve this goal, the S-PB layer is responsible for carrying semantic information from the upper layers with physical signals; the following modules in the S-PB layer should be designed carefully.

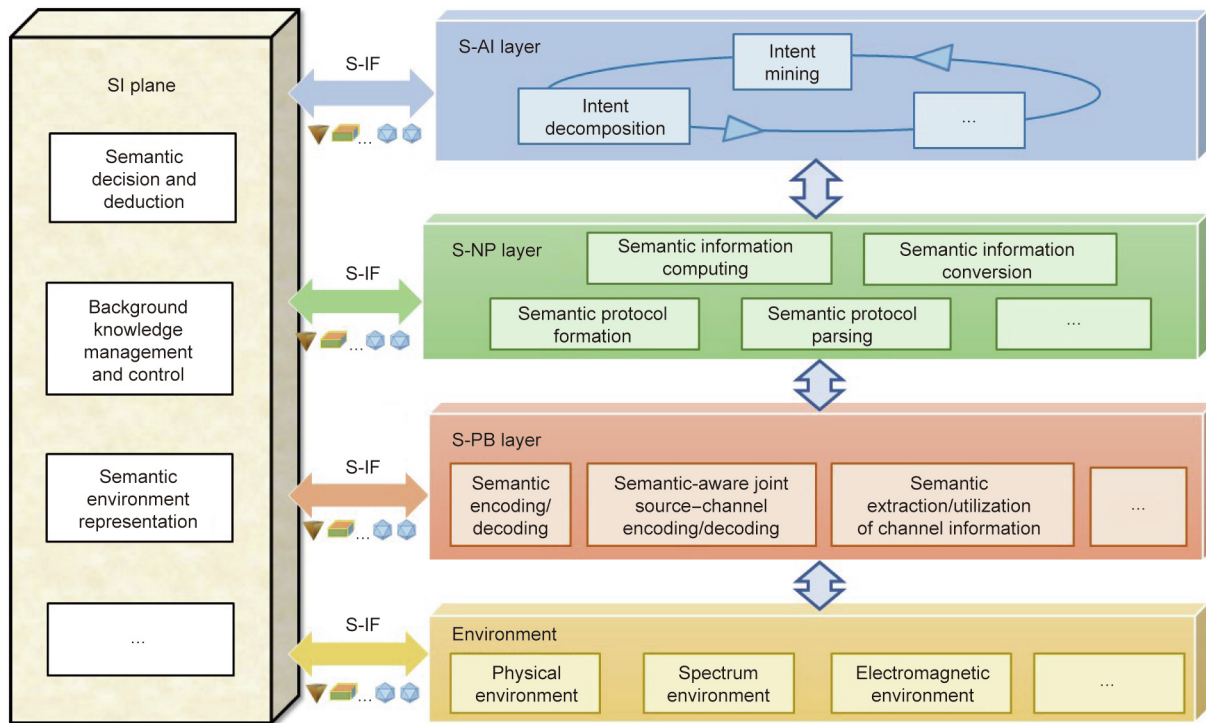


Fig. 3. The intelligent semantic communication-empowered Ubiquitous-X 6G framework, featuring an SI plane, three layers including the S-AI layer, S-NP layer, and S-PB layer, and S-IF.

Table 2
Comparison between semantic communication architecture and syntactic communication architecture.

Performance metric\architecture	Semantic communication architecture	Syntactic communication architecture
Complexity of computation and implementation	High	Low
Delay	Low	High
Bandwidth	Low	High
Accuracy	High	Low
Throughput	High	Low

(1) **Semantic encoding/decoding.** According to the modular design method, the encoding and recovering process of information is realized at the semantic level, independent of other modules in the system, such as channel coding.

(2) **Semantic-aware joint source–channel encoding/decoding.** According to the integrated design method, source encoding/decoding and channel encoding/decoding can be jointly designed to perform semantic encoding/decoding.

(3) **Semantic extraction/utilization of channel information.** Channel state information, such as fading, interference, and signal-to-noise ratio (SNR), is extracted and integrated to facilitate semantic information transmission.

The architecture of the semantic communication link is illustrated in Fig. 4. It is notable that the background knowledge of the source and the destination may differ in general. As a result, semantic information extracted at the source may be understood in a different way by the destination, which imposes significant challenges on semantic communications. In the following discussion, we provide three cases of semantic transmission in the S-PB layer with fully synchronized source–destination background knowledge. The specific solutions are also discussed for semantic encoding/decoding and semantic-aware joint source–channel encoding/decoding. Here, we mainly consider the data-driven method. Model-driven semantic encoding/decoding and semantic-aware joint source–channel encoding/decoding are left for future study.

Case 1: context-based semantic encoding/decoding for text

Context-based semantic encoding/decoding is designed according to the modular design method, as shown in Fig. 4. The transmitter’s background knowledge and the receiver are implemented with part-of-speech (POS) tagging, semantic similarity, and context. Besides the item-occurrence probability distribution, the POS-based encoding method considers that one codeword can be assigned to several items with different POS tagging—that is, nouns, verbs, and so forth—to reduce the number of transmitted bits. The decoding method can distinguish these semantically distant items with the same codeword based on specific context information. The encoding process can be described as four steps, taking Google’s Brown corpus [30] as an example. First, all words in the corpus are supposed to be divided into P classes ($P \geq 2$) according to their POS tags, where P is the number of classes. Second, each class is sorted in descending order of the occurrence frequency of its words to form its occurrence frequency ranking list. Third, the i th word of the list is obtained from each class to place into a coding node A_i , $i = 1, 2, \dots, M$, where M denotes the maximum number of items in these P classes. Note that each coding node contains P words $|\{A_i\}| = P$ and the weight of A_i is the sum of the frequencies of occurrence of these P words. Fourth, a Huffman tree is built with all coding nodes, in which each A_i corresponds to a leaf node. The Huffman coding works from leaves to root to minimize the probability-weighted mean of the code length.

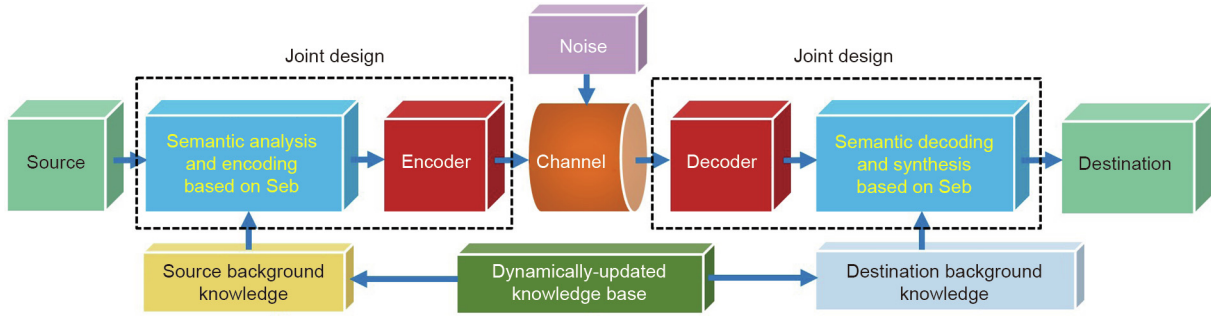


Fig. 4. Illustration of semantic communication in the S-PB layer.

In the context-based decoding method, the sequence \mathbf{s} is represented as $\mathbf{s} = (s_1, s_2, \dots, s_n)$, $n \in \mathbb{Z}^+$ is the size of the sequence \mathbf{s} or the context window in the context-based dynamic programming algorithm. Here, \mathbf{s} can be modeled as a Markov chain, and the context can be modeled as a state transition probability. Thus, the decoding involves finding the sequence \mathbf{s}^* with the maximum probability in the set \mathcal{S} that collects all the possible realizations of \mathbf{s} . Then, the N -gram model [31] and dynamic programming algorithm are adopted to solve the problem as $\mathbf{s}^* = \operatorname{argmax}_{\mathbf{s} \in \mathcal{S}} \Pr(\mathbf{s})$, where \mathcal{S} is a set containing all possible realizations of \mathbf{s} , and $N \in \mathbb{Z}^+$ is the size of a context window. The continuous bag of words (CBOW) [32] is utilized to extract the context associated with $\Pr(\mathbf{s})$, and f is the feature window's size to extract the context features. We combine the CBOW with LSTM [33] to further improve the decoding performance in order to extract context features.

Fig. 5(a) shows that the dynamic average codeword length of the POS-based semantic coding method is shorter than that of the baseline—that is, the standard Huffman coding method. This observation suggests that the proposed encoding method can reduce the number of bits for transmission. In Fig. 5(b), four standard semantic similarity scores regarding the transmitted and recovered character all increase with the growth of the context window size n when the size of the feature window f is set as 4. Moreover, when n is not smaller than f , the proposed context-based decoding method achieves and maintains a high semantic similarity score.

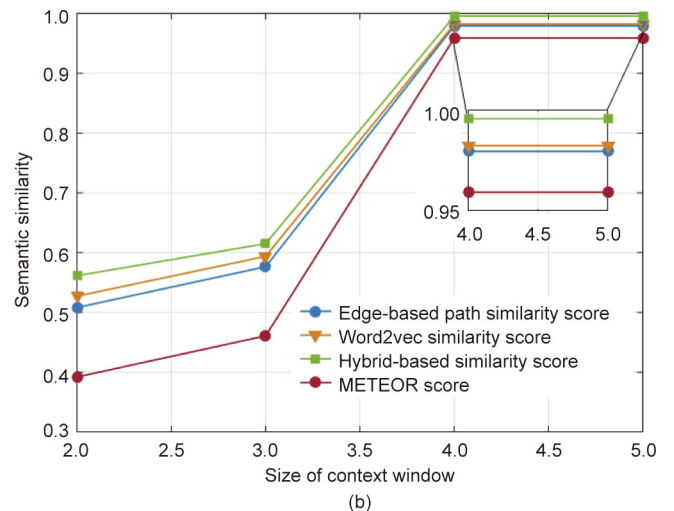
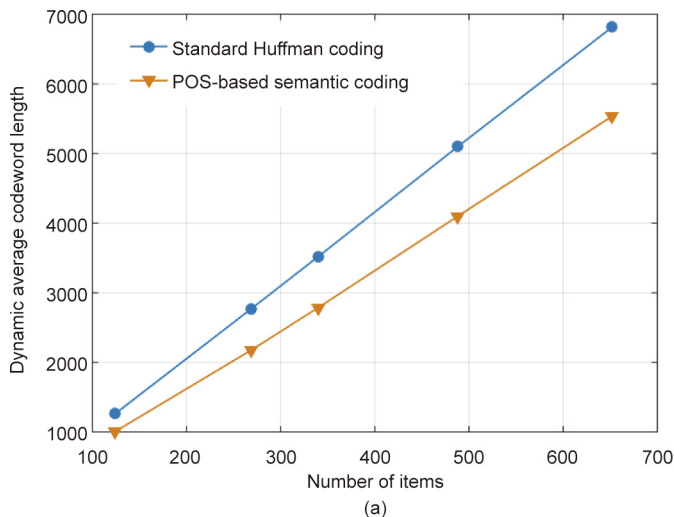


Fig. 5. (a) Dynamic average codeword length; (b) semantic similarity with a feature window of size 4. METEOR: metric for evaluation of translation with explicit ordering; word2vec: word to vector.

Case 2: semantic encoding/decoding neural network for industry images

In this case, we present a semantic image encoding/decoding scheme in the specific scenario, as shown in Fig. 6. In the considered model, an input image \mathbf{x} is compressed as a semantic vector \mathbf{w} by the semantic source encoder $g_E(\mathbf{x}; \theta_E)$, where θ_E is the parameter set of $g_E(\cdot)$. After that, the semantic vector \mathbf{w} should be quantized as $\hat{\mathbf{w}}$, coded, and modulated as discrete symbols for transmission. The channel decoder and the semantic source decoder $g_D(\hat{\mathbf{w}}; \theta_D)$ parameterized by θ_D generate the reconstruction image $\hat{\mathbf{x}}$ from the noisy symbols at the receiver. The source encoder/decoder parameters are jointly optimized as $\{\theta_E^*, \theta_D^*\} = \operatorname{argmin}_{\theta_E, \theta_D} E_{\mathbf{x}}[\lambda H(Q(g_E(\mathbf{x}; \theta_E))) + d(\mathbf{x}, g_D(Q(g_E(\mathbf{x}; \theta_E))))]$, where $H(\cdot)$ is Shannon entropy, $d(\cdot)$ is the distortion function, $Q(\cdot)$ is the quantization function, $E_{\mathbf{x}}(\cdot)$ denotes the expectation operator for \mathbf{x} , λ is the hyper-parameter, and g_D is a discriminator. We use $\lambda > 0$ to balance the distortion term against the entropy term. The commonly used distortions (e.g., mean square error (MSE)), sometimes fail to describe humans' semantic/perceptual distortion. Therefore, we use MSE to measure the pixel-wise distortion and a neural discriminator to learn the semantic/perceptual distortion. The distortion loss is defined as $d(\mathbf{x}, \hat{\mathbf{x}}) = E_{\mathbf{x} \sim p(\mathbf{x})} [\alpha \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \beta g_D(\hat{\mathbf{x}}; \theta_D)]$, where $p(\cdot)$ denotes the probability distribution of a random variable, α and β are two controlling factors balancing the related terms, and $g_D(\hat{\mathbf{x}}; \theta_D)$ is a

discriminator parameterized by θ_D that forms a generative adversarial networks (GAN) structure with $g_E(\mathbf{x}; \theta_E)$.

Our training set consists of a large set of images collected from the industrial cameras. The model is trained with a resolution of 256×256 and fine-tuned with 1920×1080 images, where the resolution of $h \times w$ denotes a image with h pixels height and w pixels weight. Adam [34] is chosen as the optimizer with a constant learning rate of 0.0002 for 50 000 iterations during the training phase. The detailed simulation settings are given in Table 3.

For a fair comparison, H.264 encoding is set to the frame-by-frame encoding mode—that is, the intra-frame-only mode. Pixel-wise metrics such as the peak signal-to-noise ratio (PSNR) or structural similarity (SSIM) are sometimes far from human aesthetic perception. Therefore, we adopt the learned perceptual image patch similarity (LPIPS) metric [35] for evaluation. Visual examples are shown in Fig. 7. Case 2 presents a semantic image transmission scheme with a finite image training dataset in a specific industry

scenario. For generalized image sources, more research efforts are needed to develop universal coding and transmission schemes.

Case 3: deep learning-based end-to-end semantic encoding/decoding

Unlike the modular design, source coding and channel coding can be jointly designed and represented by neural networks. In this case, the semantic transceiver can be regarded as an end-to-end communication system, which merges the typical communication blocks to represent and transmit semantic information [5], as shown by the dotted box in Fig. 4. A deep learning-enabled semantic communication (DeepSC) [5] and its variants, named L-DeepSC [36] and DeepSC-S [23], have been proposed for texts and speech transmission. The source information is directly mapped to the transmitted symbols through the semantic transmitter, which contains the semantic encoder and channel encoder, represented by neural networks. At the receiver, noisy information is recovered by the corresponding semantic receiver. More specifically,

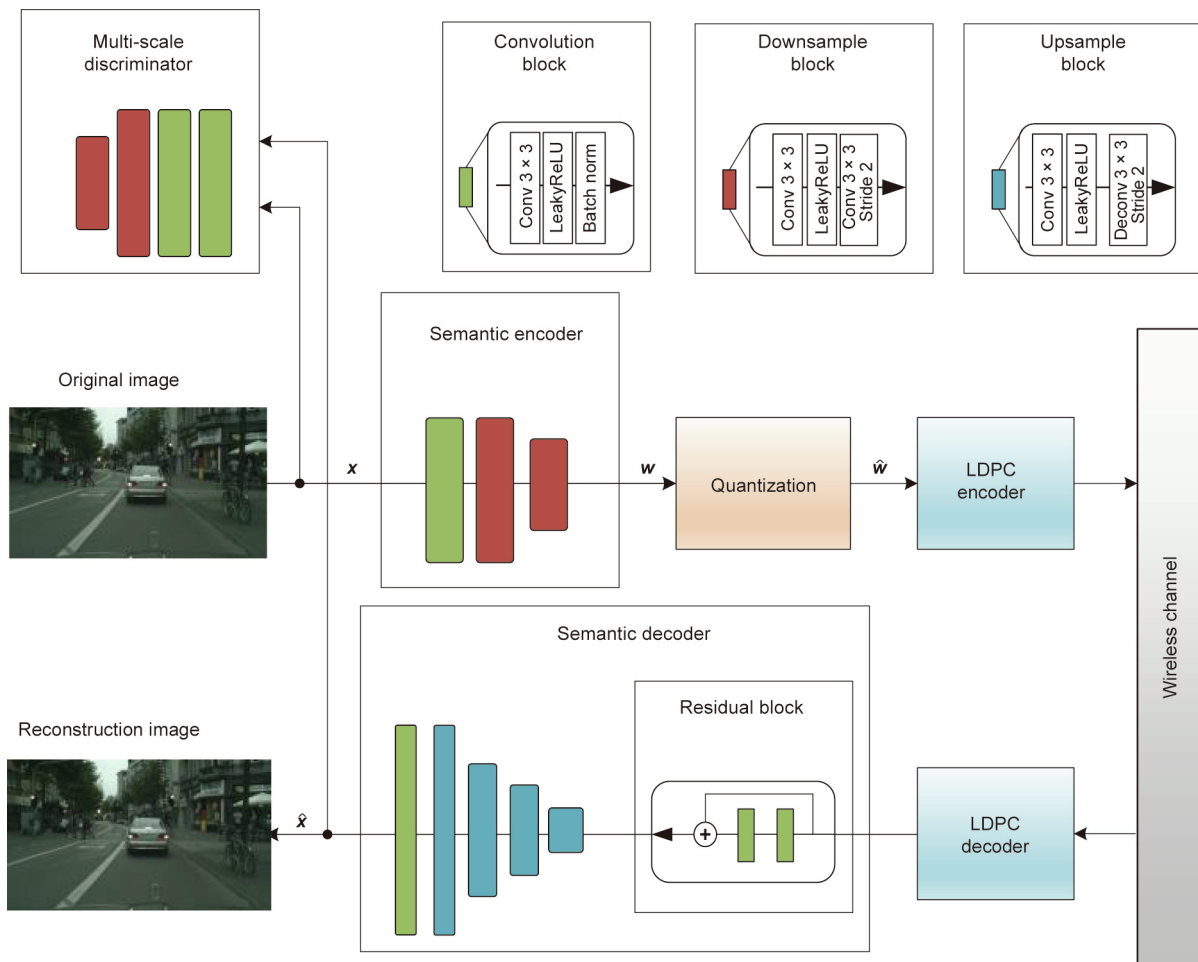


Fig. 6. Illustration of the semantic encoding/decoding neural network for industry images. \mathbf{x} : input image; \mathbf{w} : semantic vector; $\hat{\mathbf{w}}$: the quantized \mathbf{w} ; $\hat{\mathbf{x}}$: reconstructed image; LDPC: low density parity check code; LeakyReLU: leaky rectified linear unit; Conv: convolution; Decov: deconvolution.

Table 3
Simulation settings.

Source	Resolution	Frame rate (fps)	Channel	Channel coding	Code rate (Mbps)	LPIPS
Original video	1920 × 1080	25	3 dB AWGNC	(660, 440) LDPC	415	—
H.264 encoding	1920 × 1080	25	3 dB AWGNC	(660, 440) LDPC	12.44	0.15
Semantic encoding	1920 × 1080	25	3 dB AWGNC	(660, 440) LDPC	2.49	0.14

fps: frame per second; dB: decibel; AWGNC: additive white Gaussian noise channel; LDPC: low density parity check code; Mbps: megabit per second; LPIPS: learned perceptual image patch similarity.

Transformer [37] is utilized to extract the semantic information. Channel coding is achieved by a fully connected layer.

Take text as an example, where the input of the neural network is a sentence. The total loss function to train the whole neural network is the weighted sum of the cross-entropy (CE) and the estimated mutual information (MI), which recovers the transmitted sentences at the semantic level, and maximizes the data rate [31]. Furthermore, the denoising neural network estimates the channel state information during training [36]. In addition, since the transmitter directly maps the original sentence to the sending symbols, the learned constellation will not be limited to just a few points, placing additional burdens on the hardware. In this regard, a two-stage method has been proposed to compress the learned constellation point inspired by the network quantization for the trained model. By doing so, the compressed constellation points only need eight bits for quantization without performance degradation. Thus, the size of the neural network model is lowered, making the proposed method applicable for IoT scenarios.

Fig. 8 [5,38] shows a performance comparison of the proposed DeepSC and the following benchmarks: ① the traditional communication link with Huffman coding and Reed–Solomon (RS) code (5, 7) in 64-quadrature amplitude modulation (QAM); ② the

traditional communication link with five bits coding with RS (7, 9) in 64-QAM; and ③ deep learning-enabled joint source–channel coding [38]. The bilingual evaluation understudy (BLEU) score is used to measure the sentence similarity; this is a commonly used indicator in machine translation [39]. As shown in Fig. 8 [5,38], the proposed DeepSC method performs better than the benchmarks, especially in the low SNR regime.

3.2.4. S-NP layer

The primary function of the S-NP layer is to efficiently serve the upper-layer application intents with intelligent network protocols. The intent is related to the purpose of communication. For example, the purpose of the real-time interaction between the terminal and the monitor is to detect abnormal situations, so the interaction is intended to “inform” about an abnormality or “inquire” about the abnormal parameters. As another example, the purpose of communication between two agents in the industrial network is to collaboratively fulfill a specific task, so the interaction is intended to “exchange” data or “confirm” a fact. Therefore, the design of the S-NP layer mainly focuses on semantic interaction mechanisms and strategies. More specifically, the S-NP layer includes several key modules:

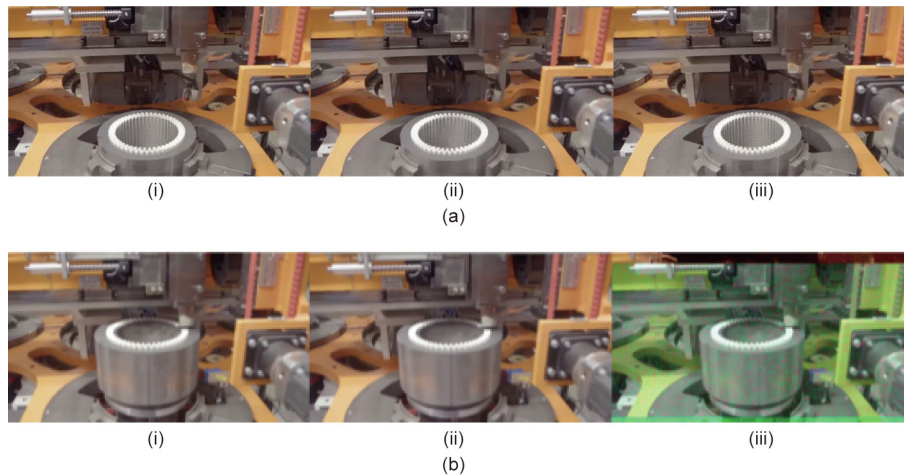


Fig. 7. (a) Visual comparison results of (i) original image, (ii) semantic encoding method, and (iii) H.264 coding. (b) Visual comparison results of (i) original image, (ii) semantic encoding method, and (iii) H.264 coding with error propagation.

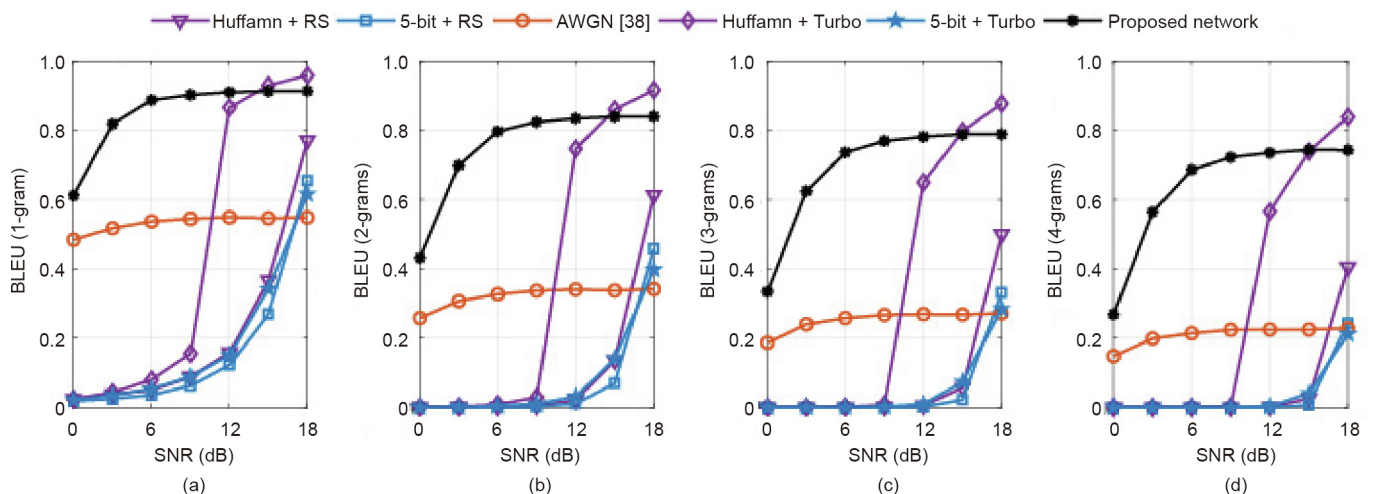


Fig. 8. Bilingual evaluation understudy (BLEU) score versus SNR for the same total number of transmitted symbols, with benchmarks and our DeepSC trained under the AWGN channel [5]: (a) 1-gram, (b) 2-grams, (c) 3-grams, and (d) 4-grams. RS: Reed–Solomon; AWGN: additive white Gaussian noise; dB: decibel.

- **Semantic information computation.** This module extracts intent information from the S-IF and obtains knowledge from the counterparts.
- **Semantic protocol parsing.** This module analyzes the available functions of the current protocols.
- **Semantic protocol formation.** This module optimizes the original protocol or forms a new protocol to meet the application's intents.
- **Semantic information conversion.** This module encapsulates semantic information according to the generated protocol.

Since network intelligence will be strengthened continuously toward 6G, the interaction protocol among elements/terminals in the 6G network will generate SI. Taking the routing scheme as an example, the SI plane determines a routing policy with high-level SI in order to directly fulfill the users' intents. Moreover, the routing policy is dynamic and autonomous, so it is not fully preset by the network administrator, as is done in conventional routing schemes.

Recent advances in intelligent multi-agent communication offer new insights into the autonomous protocol for future networks. A deep-learning-based multi-agent communication scheme is proposed in Ref. [26], wherein a learning-based interaction strategy is proposed. Such an intelligent interaction can be considered as a semantic application-layer protocol built on top of existing protocol layers. In this direction, a goal-oriented reference expression generation method is proposed in Ref. [40], which investigates intelligent interaction schemes to accomplish the application intents. The agent utilizes the learned model to generate reference expressions based on interactive dialogue clues, significantly improving goal-achieving efficiency. The scheme in Ref. [40] offers some clues to inspire intent-grounded semantic interaction among genies. Fig. 9 outlines a pipelined method for such semantic interaction, which contains three main steps:

- **Complex interaction and learning.** In this step, two genies learn to recognize the intents of the counterpart through multiple rounds of dialogues to accumulate experience for more efficient communications.
- **New knowledge update.** In this step, the genies update their knowledge about the communication goals and interaction strategies.
- **Simplified semantic interaction.** In this step, the genies can refine and optimize interaction strategies to achieve efficient semantic interaction after knowledge accumulation.

Designing lower-layer network protocols to support efficient semantic interaction is challenging. Aiming at this goal, we conceive a roadmap to evolve the S-NP layer within the IE-SC architecture. In general, we can modify existing layered protocols according to the interaction intent with semantics awareness, and then gradually integrate these protocols toward a new autonomous protocol. More specifically, guided by specific intents, the genie in the Ubiquitous-X 6G framework can use semantic information computation and protocol parsing modules to sense and filter the intent-related protocol functions in order to reduce non-intent-related protocol redundancy. Then, based on intent-related protocol functions, the genie rearranges the protocol functions with appropriate priorities determined by specific intent. The semantic protocol formation module generates a modified protocol with clear semantics awareness. As the genie is intelligent, various protocol modification policies can be learned and accumulated. Employing the experience from such policies, new intent can be quickly identified by matching it with an appropriate semantic-aware protocol in the S-NP layer. Finally, advanced AI tools can be implemented to promote an integrated and autonomous protocol to align with general intents by using the accumulated knowledge. Protocol functions will be automatically orchestrated or even generated, and the boundaries between the layers will also be blurred.

3.2.5. S-AI layer

In a broad sense, each communication user or object has an intent that is related to specific network services. A user's intent is often decomposed and translated into the network's specific deployment, configuration, or control policies. Semantics play a double role in the expression of a user's intent and in the network's understanding of that intent. In our proposed architecture, the S-AI layer can mine, understand, and decompose intents. It can transfer the sub-intent set to the SI plane via the S-IF in order to drive intelligent network management. As shown in Fig. 10, the S-AI layer has three main functions:

- **Intent mining.** After receiving the original expressions of intent from the users/applications, the S-AI layer extracts, analyzes, aggregates, and synthesizes these intents for further processing.
- **Intent decomposition.** The extracted intent is decomposed into a set of sub-intents that can guide the execution of each layer's functions.
- **Semantic representation.** The S-AI layer performs intent representation on the sub-intent set, facilitating the SI plane's decision-making.

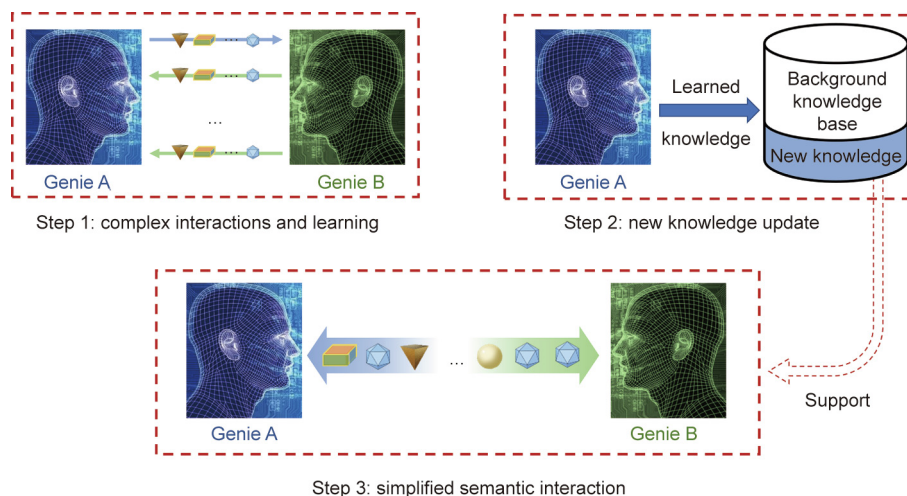


Fig. 9. Multi-agent interactive dialogue organizes the protocols on demand to support specific communicating intent as the grounded knowledge accumulates incrementally.

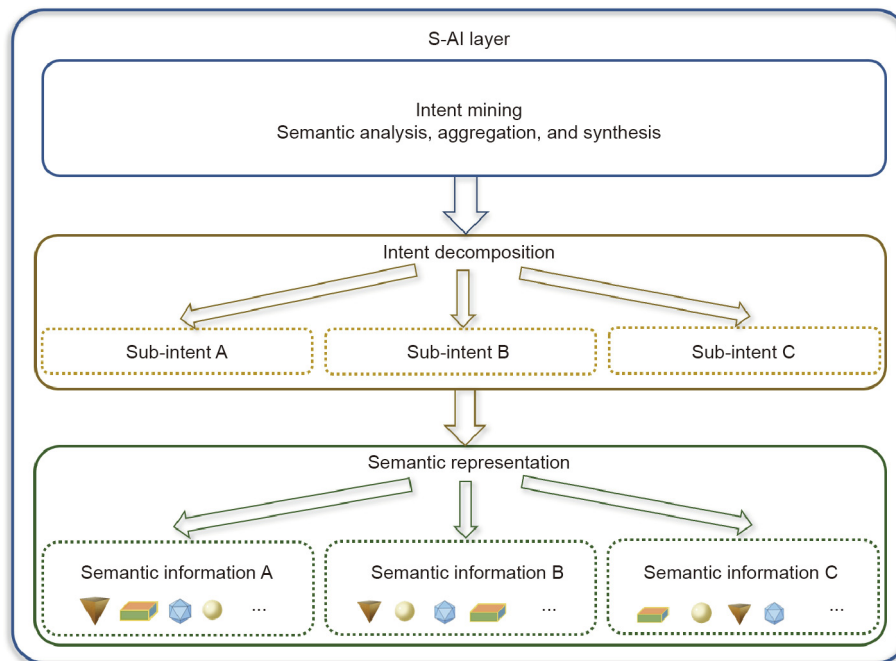


Fig. 10. The proposed generalized intent resolution process. The S-AI layer is used to mine intent through semantic analysis, aggregation, and synthesis; the mined intent is then decomposed into multiple sub-intents. Finally, the sub-intents are represented in the semantic embeddings.

Based on the intent information from the S-AI layer, the SI plane generates corresponding semantic instructions and maps them to the functions of the S-NP and S-PB layers; in this way, SI is embedded into the network.

Due to the diversity and complexity of intents, there are several challenges in achieving an intent-driven network, especially in understanding intents and in the cross-layer implementation of an intent. First, intent understanding is related to complex semantic processing. Taking NL as an example, users might express more than one intent in an utterance, or the intent might be inexplicitly embedded in dialogue. To address such issues, statistical tools and machine learning tools are often employed for intent inference. In addition, it is necessary to implement the intent across multiple protocol layers and domains in order to fully exploit the potential of semantic communication. To achieve this goal, we extend our early work [41] and conceive the concept of S-IF. In particular, S-IF can flow across layers in the overall framework, enabling the cross-layer implementation of intent. For example, S-IF can inform the S-PB layer to adopt scalable semantic coding to meet a specific application intention with different information granularity, such as a high-definition video with subtle details or autonomous robot collaboration with only feature details. Moreover, S-IF can promote efficient interaction among heterogeneous network elements to achieve mutual understanding.

4. Promising application scenarios of semantic communication networks

In this section, we outline three promising application scenarios for the IE-SC-empowered Ubiquitous-X 6G network: the air–space–ground–ocean integrated network (ASGO-IN), the Industrial Internet of Things (I-IoT), and the intelligent unmanned machine network (IUMN), as shown in Fig. 11. All these promising application scenarios will converge in the WePCN vision of the Ubiquitous-X 6G.

4.1. ASGO-IN

The ASGO-IN integrates terrestrial networks with satellite, ocean, and aerial networks, as shown in Fig. 11(a). It is a widely recognized solution to achieve global coverage and on-demand service for 6G networks. One technical challenge in realizing ASGO-IN is flexible and efficient interconnection among heterogeneous networks covering a large span of space and time. Currently, gateways are used to bridge different networks with complicated protocol translations. Our proposed IE-SC architecture provides an agile and concise solution to support ASGO-IN. To be specific, the S-AI layers at heterogeneous nodes in ASGO-IN can identify their intent of integration. Using the preinstalled common knowledge, the SI plane coordinates the S-NP layer to orchestrate the intent-related semantic elements to form a concise integration-oriented protocol that is implementable among the nodes. Without an additional gateway, these nodes can directly interact with the S-PF layer through a unified air interface to perform the integration.

4.2. Industrial Internet of Things

The I-IoT introduces advanced information and communication technologies to connect various elements, such as humans, machines, and things, to serve the purpose of industrial manufacturing with collaboration and interoperability, as shown in Fig. 11(b). However, existing data communication networks cannot efficiently integrate manufacturing intents into the interactions among these elements, which results in low collaboration efficiency and heavy signaling overhead. Under our proposed IE-SC architecture, the S-AI layer can identify the manufacturing purpose and generate semantic communication strategies toward collaboration and interoperability. Furthermore, intent-related semantic information can be transmitted efficiently in a highly compressed manner by using efficient joint semantic-source–channel coding

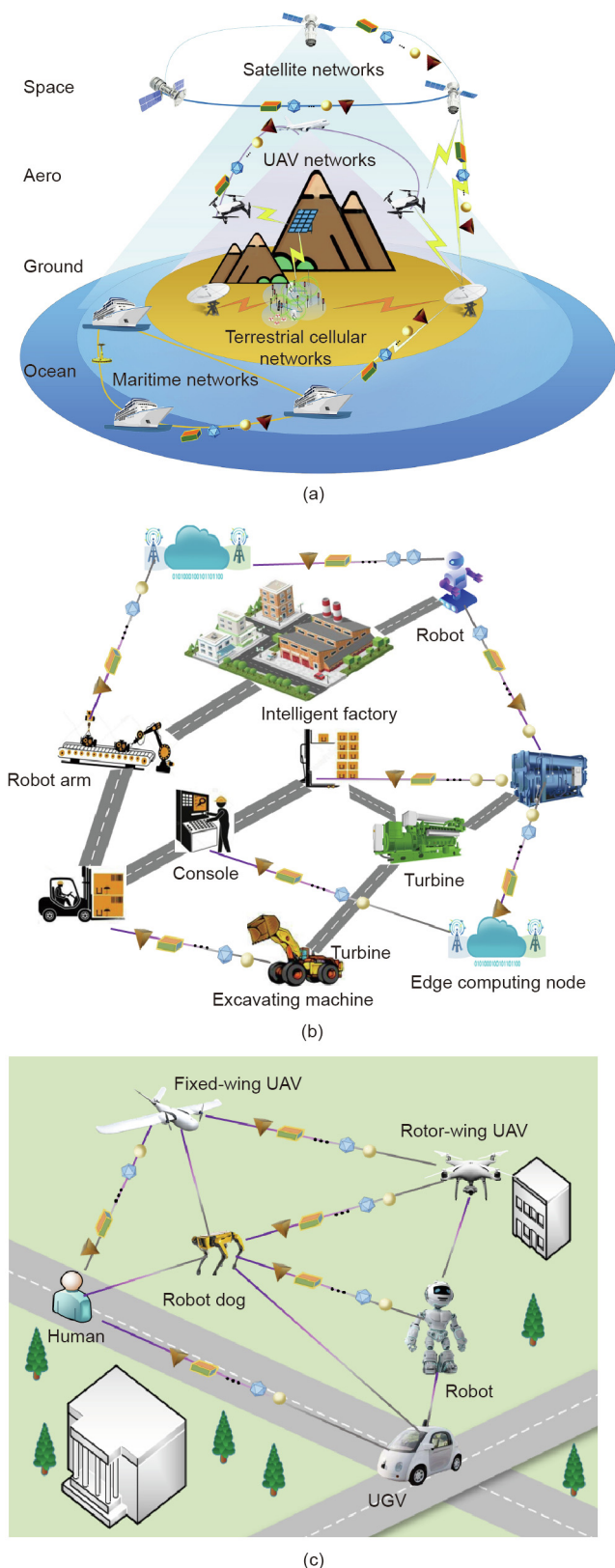


Fig. 11. (a) A semantic-empowered ASGO-IN; (b) a semantic-driven I-IoT; (c) a semantic-based IUMN with unmanned ground vehicle (UGV).

at the S-PB layer. Thus, the overall efficiency of the I-IoT will be boosted significantly to support intelligent manufacturing with intent-oriented networking and collaboration under our IE-SC architecture.

4.3. Intelligent unmanned machine network

The IUMN, which includes autonomous vehicles, robotics, and drone swarms, as shown in Fig. 11(c), represents an expansion of networking from data-centric communication to machine-centric control and collaboration [42]. Equipped with advanced sensing and communication modules, each intelligent unmanned machine (IUM) can sense the environment and interact with others to accomplish specific tasks. Our proposed IE-SC architecture can support autonomous and task-driven networking among IUMs. The S-AI layer can comprehensively analyze the sensing data to extract the task-related semantic information, which is then fed into the SI plan to generate semantic-based decision policies and networking strategies to accomplish the task. Following the SI plan decision and strategy, the S-NP layer can dynamically control link configurations, network topology, and routing strategy to assure a robust and agile task-driven IUMN.

5. Main challenges and future directions

(1) **Further development of Seb representation.** The exact connotation behind a piece of information is affected by the understanding of communication agents. Moreover, different agents may have different syntactic forms for the same connotation, just like synonyms or multilingual phenomena in NL. Therefore, the Seb framework should be further studied to enable unified and generalized semantic information extraction and representation for multimodal information. It should be noted that the current semantic extraction/representation relies on AI and neural networks, which involve extensive computation. Therefore, Seb is expected to be an essential building block for a more comprehensive semantic information-processing framework that integrates semantic communication and computation.

(2) **Fundamental limits of semantic communications.** In the physical layer, the design objective of semantic communications is to optimize semantic information transmission over different types of channels with relevant background knowledge. Therefore, the fundamental limits of semantic communications are determined by physical and contextual constraints. Furthermore, the degree of mutual understanding between a pair of communication agents may determine the interaction or signaling strategies and the volume of semantic communications. Thus, an appropriate measure of intent-achieving efficiency should be established in order to answer this question: What is the most efficient semantic communication strategy to achieve the intent? The measurement framework is generally abstract and complex; strategies may include the semantic-related processing in higher layers and semantic-aware joint source-channel coding in the physical layer. Therefore, some theories and coding schemes should first be established in order to concretize the new measure framework, offering certain achievable bounds toward the limits of semantic communications.

(3) **General semantic-based intent-driven networks.** Currently, intent-driven networks mainly target human-centric network configuration and management. NL serves as the carrier for human intents. The NLP module serves as a middle-ware to translate human intents into predefined network configuration policies. However, NL is not necessarily the best tool for non-human-centric networks in the new era of human-machine-thing-genie networking. In addition, the concept of intent-driven networks may go beyond conventional ICT data networks; it is applicable in various emerging functional networks, including military and manufacturing networks, among others. To this end, a generalized semantic-based intent-driven network should be established, wherein Seb-based coding and interaction may replace NL to

enable efficient cross-species semantic communications, and intent-driven capabilities should become native capabilities for both data and functional networks.

6. Conclusions

In this paper, we proposed a systematic design for semantic communication networks to support intelligent interactions among various communication agents in Ubiquitous-X 6G networks. First, we initiated the Seb concept, which serves as the representation of semantic information. Next, we presented the IE-SC architecture, which consists of the SI plane, S-PB layer, S-NP layer, S-AI layer, and S-IF. Initial simulation results were presented to demonstrate the efficiency of semantic-based information delivery. Promising application scenarios and future directions were discussed to inspire further research efforts toward the vision of WePCNs.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (2019YFC1511302); in part by the National Natural Science Foundation of China (61871057); and in part by the Fundamental Research Funds for the Central Universities (2019XD-A13). In addition, the authors would like to acknowledge support from professors for technical support and expert advice. The authors also acknowledge Yimeng Zhang, Yuan Zheng, Sixian Wang, Zhongyi Wang, Jiangjing Hu, Haobing Gong, and Yuting Feng from Beijing University of Posts and Telecommunications and Yichi Zhang from National University of Defense Technology, for literature arrangement, writing assistance, and proof-reading.

Compliance with ethics guidelines

Ping Zhang, Wenjun Xu, Hui Gao, Kai Niu, Xiaodong Xu, Xiaoqi Qin, Caixia Yuan, Zhijin Qin, Haitao Zhao, Jibo Wei, and Fangwei Zhang declare that they have no conflict of interest or financial conflicts to disclose.

References

- Zhang P, Zhang J, Qi Q, Hu Z, Nie G, Niu K, et al. Ubiquitous-X: constructing the future 6G networks. *Sci Sin Inform* 2020;50(6):913–30. Chinese.
- You X, Wang CX, Huang J, Gao X, Zhang Z, Wang M, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. *Sci China Inf Sci* 2021;64(1):110301.
- Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;27(3):379–423.
- Jiang A, Li Y, Bruck J. Error correction through language processing. In: *Proceedings of 2015 IEEE Information Theory Workshop (ITW)*; 2015 Apr 26–May 1; Jerusalem, Israel; 2015.
- Xie H, Qin Z, Li GY, Juang BH. Deep learning enabled semantic communication systems. *IEEE Trans Signal Process* 2021;69:2663–75.
- Jankowski M, Gündüz D, Mikolajczyk K. Joint device-edge inference over wireless links with pruning. In: *Proceedings of 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*; 2020 May 26–29; Atlanta, GA, USA. New York: IEEE; 2020.
- Jankowski M, Gündüz D, Mikolajczyk K. Joint device-edge inference over wireless links with pruning. In: *Proceedings of 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*; 2020 May 26–29; Atlanta, GA, USA. New York: IEEE; 2020.
- Shi Y, Yao K, Chen H, Pan YC, Hwang MY, Peng B. Contextual spoken language understanding using recurrent neural networks. In: *Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2015 Apr 19–24; South Brisbane, QLD, Australia; New York: IEEE; 2015. p. 5271–5.
- Morris CW. *Foundations of the theory of signs*. Chicago: University of Chicago Press; 1938.
- Weaver W. Recent contributions to the mathematical theory of communication. *ETC Rev Gen Semant* 1953;10(4):261–81.
- Carnap R, Bar-Hillel Y. An outline of a theory of semantic information. *J Symb Log* 1954;19(3):230–2.
- Bar-Hillel Y, Carnap R. Semantic information. *Br J Philos Sci* 1953;4(14):147–57.
- Barwise J, Perry J. Situations and attitudes. *J Philos* 1981;78(11):668–91.
- Floridi L. Outline of a theory of strongly semantic information. *Minds Mach* 2004;14(2):197–221.
- D'Alfonso S. On quantifying semantic information. *Information* 2011;2(1):61–101.
- Zhong Y. A theory of semantic information. *China Commun* 2017;14(1):1–17.
- Kolchinsky A, Wolpert DH. Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface Focus* 2018;8(6):20180041.
- Kountouris M, Pappas N. Semantics-empowered communication for networked intelligent systems. 2020. arXiv:2007.11579.
- Rényi A. On measures of entropy and information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; 1960 Jun 20–Jul 30; Berkeley, CA, USA. Berkeley: University of California Press; 1960.
- Rao M, Farsad N, Goldsmith A. Variable length joint source–channel coding of text using deep neural networks. In: *Proceedings of 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*; 2018 Jun 25–28; Kalamata, Greece. New York: IEEE; 2018. p. 1–5.
- Guler B, Yener A, Swami A. The semantic communication game. *IEEE Trans Cogn Commun Netw* 2018;4(4):787–802.
- Bourtsoulatz E, Burth Kurka D, Gunduz D. Deep joint source-channel coding for wireless image transmission. *IEEE Trans Cogn Commun Netw* 2019;5(3):567–79.
- Weng Z, Qin Z. Semantic communication systems for speech transmission. *IEEE J Sel Areas Commun* 2021;39(8):2434–44.
- Alshbatat AI, Dong L. Cross layer design for mobile *ad-hoc* unmanned aerial vehicle communication networks. In: *Proceedings of 2010 International Conference on Networking, Sensing and Control (ICNSC)*; 2010 Apr 10–12; Chicago, IL, USA. New York: IEEE; p. 331–6.
- Awang A, Husain K, Kamel N, Aïssa S. Routing in vehicular *ad-hoc* networks: a survey on single and cross-layer design techniques, and perspectives. *IEEE Access* 2017;5:9497–517.
- Sukhbaatar S, Szlam AD, Fergus R. Learning multiagent communication with backpropagation. In: *Proceedings of the 30th Annual Conference on Neural Information Processing Systems 2016*; 2016 Dec 5–10; Barcelona, Spain. New York: Curran Associates Inc.; 2016. p. 2252–60.
- Wen L, Wang X, Dong Z, Chen H. Jointly modeling intent identification and slot filling with contextual and hierarchical information. In: *Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing*; 2020 Oct 14–18; Zhengzhou, China. Cham: Springer 2020.
- Chao W, Horiuchi S. Intent-based cloud service management. In: *Proceedings of 21st Conference on Innovation in Clouds, Internet and Networks and Workshops(ICIN)*; 2018 Feb 19–22; Paris, France. New York: IEEE; 2018.
- Wei Y, Peng M, Liu Y. Intent-based networks for 6G: insights and challenges. *Digit Commun Netw* 2020;6(3):270–80.
- Kučera H, Francis WN. *Computational analysis of present-day American English*. Providence: Brown University Press; 1967.
- Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. *J Mach Learn Res* 2003;3:1137–55.
- Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space 2013. arXiv:1301.3781v3.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv: 1412.6980.
- 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. New York: IEEE; 2018. p. 586–95.
- Xie H, Qin Z. A lite distributed semantic communication system for Internet of Things. *IEEE J Sel Areas Commun* 2021;39(1):142–53.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Proceedings of the 31th International Conference on Neural Information Processing Systems*; 2017 Dec 4–9; Long Beach, CA, USA. New York: Curran Associates Inc.; 2017. p. 5998–6008.
- Farsad N, Rao M, Goldsmith A. Deep learning for joint source–channel coding of text. In: *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2018 Apr 15–20; Calgary, AB, Canada. Stroudsburg: ACL; 2018. p. 2326–30.
- Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*; 2002 Jul 7–12; Stroudsburg, PA, USA; 2002. p. 311–8.

- [40] Li L, Zhao Y, Zhang Z, Niu T, Feng F, Wang X. Referring expression generation via visual dialogue. In: Proceedings of the 9th CCF International Conference on Natural Language Processing and Chinese Computing; 2020 Oct 16–18; Zhengzhou, China. Cham: Springer; 2020.
- [41] Xu W, Lin J, Feng Z, Xu W, Zhang P. Cognition flow in cognitive radio networks. *China Commun* 2013;10(10):74–90.
- [42] Jiang Z, Fu S, Zhou S, Niu Z, Zhang S, Xu S. AI-assisted low information latency wireless networking. *IEEE Wirel Commun* 2020;27(1):108–15.