# Semantics-Empowered Communications: A Tutorial-Cum-Survey

Zhilin Lu, Rongpeng Li, *Member, IEEE*, Kun Lu, Xianfu Chen, *Member, IEEE*,
Ekram Hossain, *Fellow, IEEE*, Zhifeng Zhao, *Member, IEEE*, and Honggang Zhang, *Fellow, IEEE*

*Abstract*—Along with the springing up of the semantics-empowered communication (SemCom) research, it is now witnessing an unprecedentedly growing interest towards a wide range of aspects (e.g., theories, applications, metrics and implementations) in both academia and industry. In this work, we primarily aim to provide a comprehensive survey on both the background and research taxonomy, as well as a detailed technical tutorial. Specifically, we start by reviewing the literature and answering the "what" and "why" questions in semantic transmissions. Afterwards, we present the ecosystems of SemCom, including history, theories, metrics, datasets and toolkits, on top of which the taxonomy for research directions is presented. Furthermore, we propose to categorize the critical enabling techniques by explicit and implicit reasoning-based methods, and elaborate on how they evolve and contribute to modern content & channel semantics-empowered communications. Besides reviewing and summarizing the latest efforts in SemCom, we discuss the relations with other communication levels (e.g., conventional communications) from a holistic and unified viewpoint. Subsequently, in order to facilitate future developments and industrial applications, we also highlight advanced practical techniques for boosting semantic accuracy, robustness, and large-scale scalability, just to mention a few. Finally, we discuss the technical challenges that shed light on future research opportunities.

*Index Terms*—Semantic communications, tutorial-cum survey, semantic information theory, semantic similarity metrics, semantic transmission and reasoning, enabling techniques, challenges and opportunities.

Zhilin Lu and Rongpeng Li are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310058, China (e-mail: lu_zhilin@zju.edu.cn; lirongpeng@zju.edu.cn).

Kun Lu was with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310058, China (e-mail: lukun199@zju.edu.cn).

Xianfu Chen is with the VTT Technical Research Centre of Finland, 90570 Oulu, Finland (e-mail: xianfu.chen@vtt.fi).

Ekram Hossain is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada (e-mail: ekram.hossain@umanitoba.ca).

Zhifeng Zhao is with Zhejiang Lab, Hangzhou 311121, China (e-mail: zhaozf@zhejianglab.com).

Honggang Zhang is with Zhejiang Lab, Hangzhou 311121, China, and also with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310058, China (e-mail: honggangzhang@ zhejianglab.com).

Digital Object Identifier 10.1109/COMST.2023.3333342

## I. INTRODUCTION

FROM 1G in the 1980s to recently commercialized 5G, modern mobile communications have experienced substantial technical revolutions in almost every decade. Meanwhile, tremendous new applications, such as massive machine-type connectivity with stringent latency and reliability constraints, spawn and exhibit significant differences from the classical throughput-oriented transmission. Correspondingly, it is highly demanded to re-design smarter and more efficient communications to satisfy the diversified quality of service (QoS) requirements whilst balancing the dilemma of finite communication resources and infrastructures. Basically, an intelligent communication system, which is a vital ingredient in both industry and academia endeavors, is envisaged to not merely memorize data flows with rigid rules, but also understand, process, and express the underlying semantics. In that regard, apart from post-processing in downstream information carriers, compression and understanding from the very beginning would be efficient [1], [2], which also emerges as a driving force to push the frontier of exploring "meanings" in a communication system. Building upon the classical information theory (CIT) as revealed by Shannon [3], modern transmission schemes provide cutting-edge protocols (e.g., CDMA, OFDM, MIMO, etc.) for reliable communications, securing bit-level transmission accuracy. In spite of its popularity, treating each bit with equal importance can be content- and intelligence-agnostic in real applications and inconsistent with human perception as well [4]. Essentially, bit-level accuracy is not sufficient for "intelligent" and semantic transmission - intuitively, different expressions may exactly convey a similar meaning, while an expression with a high word error rate (WER) [5] may still convey the meaning (or meaningful representation) of the message (i.e., the semantics [3]). Therefore, instead of fully transmitting every single bit, it will be desirable to develop a semantics-empowered paradigm that seeks to communicate the underlying meanings.

Transmission with semantics can date back to as early as 1949, when Weaver first introduced this idea [6]. Unfortunately, owing to the technical constraint and a more urgent need for Shannon's CIT, related researches entered a period of silence. As a matter of fact, it took the community several decades to find concrete clues to break out the "Shannon's trap" [7], since mastering semantics is widely beyond the scope of CIT. Benefiting from the latest advance in artificial intelligence (AI), the prerequisite techniques

are becoming gradually available and ready for building a prototype model. As a piece of evidence, research on semantics-empowered communications (SemCom for short) is now experiencing unprecedented growth. According to Google Scholar, the number of papers titled "semantic" and "communication" in 2022 exceeds the summation of those in the previous two years. Notably, this surging trend is expected to continue in the near future since SemCom is a promising enabler for 6G and beyond applications (e.g., Metaverse applications) [1].

Nevertheless, coming along with its popularity, researches on SemCom are relatively scattered and with varied research interests. The discussions on "*what*", "*why*" and "*how*" are at their infancy. We argue that a comprehensive tutorial-cum-survey on SemCom covering basic definitions, research taxonomy, and developing trends will be very beneficial for researchers and practitioners.

### A. Related Work and Scope

SemCom is currently undergoing extensive discussions and the understanding of SemCom is continually evolving. Looking back at the footprints, works on SemCom can be generally grouped into three main categories (i.e., conceptual exploration, detailed technical research, and surveys & tutorials) from the perspectives of contributions and research directions. The first category mainly concentrates on the "*what*" and "*why*" problems, by envisioning possible features and architectures, as well as explaining the necessity of this new communication scheme. The second category answers the "*how*" question and provides concrete implementation details and techniques for enabling a SemCom system. Belonging to the last category, this paper focuses on the summary and tutorials for facilitating further researches.

Prior to our work, publicly-available surveys & tutorials on SemCom have shed some light on summarizing the latest developments and future challenges. Strinati and Barbarossa [1] provide a holistic overview of the components along with the basic mathematical theory that supports a semantic level transmission, and highlight its potential in 6G scenarios. Lan et al. [8] incorporate the semantics and efficiency problem from an inspiring human-machine modality perspective and put an emphasis on the role of knowledge graph (KG) in SemCom. On the other hand, Qin et al. [9] concentrate more on the components, metrics, and multi-modal applications. Iyer et al. [10] and Yang et al. [11] focus on the potential of intelligent wireless networks and edge intelligence respectively. Later, Yang et al. [7] further envision possible applications and benefits along with technical challenges in future SemCom-powered 6G scenarios. Very recently, Wheeler and Natarajan [12] provide the latest survey, incorporating the history, taxonomy, techniques, and challenges of SemCom, while Chaccour et al. [13] present an end-to-end vision of SemCom from the perspective of AI-enabled semantic language and knowledge reasoning.

Similar to these works [1], [7], [8], [9], [10], [11], [12], [13], this tutorial-cum-survey is also designed to provide a clearer and more in-depth insight into SemCom. In a nutshell, we first provide a holistic overview of the historical developments, motivations, and research directions. Meanwhile, we clarify a few ambiguous concepts that are not specifically considered in previous works, and present the pioneering researches from varied directions. Outlined by the proposed taxonomy, we then present a comprehensive and thorough tutorial-cum-survey on the concrete techniques, frameworks, and most recent proposals that support the goal of semantic transmission. Moreover, from end-to-end semantic transmission to semantics-assisted networks, we identify the potential applications and future challenges, along with new possible directions to facilitate future researches. However, there exist some significant differences between this tutorial-cum-survey and the literature [1], [7], [8], [9], [10], [11], [12], [13]. In particular, we summarize and highlight the differences from five key aspects (i.e., basic history & definitions, research taxonomy, coverage of techniques, tutorial, and future prospect) in Table I. Meanwhile, the major contributions of this paper relative to the recent literature [1], [7], [8], [9], [10], [11], [12], [13] are as follows:

- (*History & Definitions*) Compared to other survey papers in the field, this tutorial-cum-survey provides clearer motivation & definitions by introducing the history of key SemCom ingredients (i.e., modern mobile communications and AI) and discussing the strong incentive to converge these ingredients. We also highlight "what is semantics", "how semantics is established, shared, and represented", as well as the conditions under which a communication system can be semantic.
- (*Taxonomy*) We provide a holistic research taxonomy of enabling techniques. Moreover, we explore the relationship between the semantic level and other levels (e.g., conventional communications). Taxonomy tables to summarize the key contributions, frameworks, metrics, datasets & toolkits, and limitations are also provided.
- (*Coverage*) Besides the basic structure and working pipeline, we summarize the corresponding ecosystems like theoretical advancement, learning metrics, wireless optimizations, etc.
- (*Tutorial*) We provide a tutorial-cum-survey spanning from history, theories to applications of SemCom, and present key lessons learned from the survey, which will help researchers and practitioners to quickly dive into this topic.
- (*Future Prospect*) We comprehensively survey potential applications for SemCom, and also identify future trends as well as technical and business challenges.

### B. Contents and Structures

This work is a tutorial-cum-survey covering the basic definitions, taxonomy, applications and challenges for SemCom. The organization of its contents and the relations is illustrated in Fig. 1. Our first goal is to provide a clear definition and explanation of semantic-empowered mechanism in typical transmission scenarios and illustrate how it differs from conventional communication schemes, which serves as the cornerstone for further in-depth analysis. Beforehand, in Section II, we will revisit the historical development of semantic transmission and its related driving forces for readers who are not familiar with the surveyed topic. Afterwards,

TABLE I
SUMMARY AND COMPARISON OF RELATED SURVEY AND TUTORIAL PAPERS ON SEMCOM

| Survey | History & Definitions | Taxonomy | Coverage | Tutorial | Future Prospect | Brief Description |
|---|---|---|---|---|---|---|
| Strinati *et al.* [1] | ● | ◑ | ◑ | ◑ | ◑ | Components along with the basic mathematical theory, and 6G applications |
| Lan *et al.* [8] | ◑ | ◑ | ● | ◑ | ● | New insights from human-machine communications |
| Qin *et al.* [9] | ● | ◑ | ◑ | ● | ◑ | A brief tutorial on the definition, components, metrics and multi-modal applications |
| Iyer *et al.* [10] | ◑ | ◑ | ◑ | ◑ | ◑ | Basic introduction and opportunities on wireless networks |
| Yang *et al.* [7], [11] | ◑ | ● | ● | ● | ● | Detailed technical taxonomy and challenges, along with future applications |
| Wheeler *et al.* [12] | ◑ | ● | ● | ◑ | ◑ | Summary of semantic theory and four enabling approaches with detailed analysis |
| Chaccour *et al.* [13] | ◑ | ◑ | ◑ | ● | ● | Design of a semantic communication network on the basis AI-enabled semantic language and knowledge reasoning |
| This paper | ● | ● | ● | ● | ● | Complete background, techniques and research taxonomy, balanced with adequate coverage and tutorial |

Notations: ● *indicates fully included;* ◑ *means partially included.*
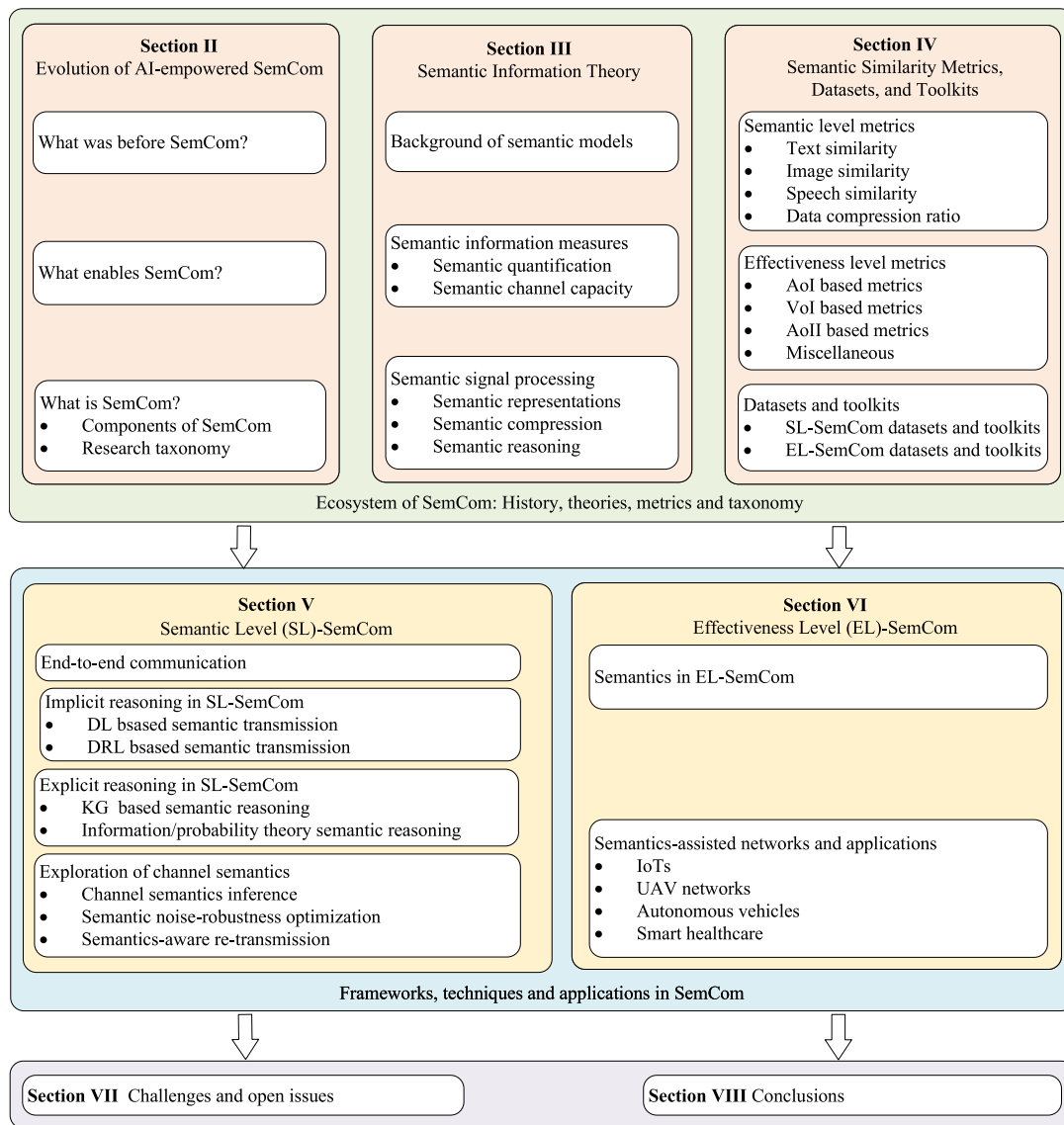


Fig. 1.   Main contents of this paper.

following the logical order, we will discuss the components of SemCom and provide the in-depth research taxonomy. In Sections III and IV, we would clarify the definitions and elaborate on the corresponding ecosystems like theoretical guidance, metrics, toolkits, etc. Next, we highlight each of the enabling techniques. In particular, in Section V, we

delve into the semantic-level SemCom (SL-SemCom) by presenting the exploration of both content semantics and channel semantics. Notably, content semantics focuses on extracting semantic information from the transmitted data to improve transmission efficiency and effectiveness, while channel semantics, which encompasses both channel state information (CSI) and surrounding environment information, puts more emphasis on leveraging channel and/or environment characteristics to facilitate the design of coding schemes, thereby better adapting to the channel and increasing transmission reliability. Furthermore, we propose to classify the utilization of content semantics in SL-SemCom to two major categories (i.e., implicit reasoning and explicit reasoning), wherein the former category typically relies on structured or unstructured parameterized models like deep neural networks (DNNs), while the latter involves explicitly defined rules or entities. Moreover, targeted at practical and large-scale commercial applications, recent progress on techniques to explore and exploit both content and channel semantics is also included. On top of these end-to-end SemCom approaches, we then transfer our viewpoint to effectiveness-level SemCom (EL-SemCom) and semantics-assisted networks in Section VI, where potential applications and opportunities in 6G (and beyond) are envisioned. Finally, some future challenges and open issues are discussed in Section VII while Section VIII concludes the paper.

## II. FROM PAST TO FUTURE OF AI-EMPOWERED SEMCOM

In this part, we begin with a brief history of conventional communications before SemCom, and talk about the potentially encountered obstacles and limits therein. Subsequently, inspired by the astonishing advance in AI, we provide the history of this driving force towards SemCom, and try to answer why it contributes to breaking out "Shannon's trap [7]". Afterwards, we formally discuss the differences between SemCom and conventional communications, and present our viewpoint of SemCom. Besides, the research taxonomy is summarized to provide a succinct overview of SemCom.

### A. What Was Before SemCom?

Before delving into the SemCom, it is meaningful to recall the evolution of mobile communications, which can be traced back to 1897 [14], [15]. Since then, we have witnessed tremendous progress in communications. In particular, 5G capably connects both humans and things, and provisions diversified services to vertical industries [16]. For example, 5G has begun to support three different service classes including enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (URLLC), and massive machine type communications (mMTC), and built up a service-based architecture (SBA) with network data analytics function (NWDAF) to make the implementation of AI feasible. Meanwhile, as summarized in Fig. 2(a), the drastic evolution of communications is also accompanied by cutting-edge techniques such as FDMA, TDMA, W-CDMA, and OFDMA [17].

In order to realize AI-based ultra-sensitive tactile and haptic communications in 6G [15], it is natural to explore more high-frequency bands and fully leverage powerful technologies like multi-layer precoding/beamforming and ultra-massive MIMO [18]. However, such an intuitive evolution often implies more intense resource demand in terms of both power and bandwidth, and gives rise to new challenging issues. For instance, though spreading signals at higher frequencies could partially resolve the over-exploitation issue of sub-6GHz spectra, it faces disadvantages such as disruptive attenuation and cost-ineffective deployment. Meanwhile, according to Shannon's CIT [3], the existing communication technologies have nearly approached the physical-layer capacity limit [19]. Therefore, we will need to develop new and more efficient technologies (e.g., SemCom) for future emerging applications.

### B. What Enables SemCom?

Considering many successful implementations of SemCom (e.g., semantic transmission & reasoning [7], [9], [13]), it can be observed that the development of SemCom is often accompanied by the flourishing of AI. Specifically, as shown in Fig. 2(c), albeit its tortuous progress at the early stages, AI, specifically deep learning (DL) based on DNNs, has made remarkable achievements recently. Inspired by this interesting observation, we go through the most commonly used AI technologies and try to explain the importance of breakthroughs in AI to SemCom.[1]

One major category of modern AI comes from discriminative models. For vision tasks, convolutional neural networks (CNNs) take advantage of the correlation of neighboring pixels, and play a key role in both high-level (e.g., detecting, tracking, classification) and low-level processing (e.g., inpainting, enhancing, denoising) [21], [22], [23], [24]. To model the sequence data, recurrent neural networks (RNNs) update their inner hidden states at each time step. In prominent research on semantics, sequential data (typically, text messages) is the most widely studied media. Benefiting from advanced AI models like long short-term memory (LSTM) [25] and attention-based transformer [26], both short-time and long-time semantic dependencies can be represented and extracted for downstream processing [27], [28], [29], [30]. Besides structured stream data, edge-node-based graph models handle the dynamics of data topology by aggregating information from their neighborhoods, which shows great success on a variety of graph media such as social networks and knowledge graphs [31], [32], [33]. Besides, deep transfer learning (DTL) [34] emerges as a useful technique for effectively using previously learned knowledge to solve novel tasks with minimum training. Meanwhile, theoretical researches such as information bottleneck (IB) theory [35], [36], [37] and probabilistic models [20], [38] contribute to easing concerns on the interpretability of DNN-based schemes.

Another category of AI models concentrates on the generative process, where the most representative ones are variational auto-encoder (VAE) [39], generative adversarial network (GAN) [40], and diffusion models [41]. These models are designed to mimic the source data distribution, which can be further used to augment finite data sets, support intelligent generation & editing, and even provide surrogate

---

[1]Notably, besides AI, there are alternative theoretical means (e.g., [20]) to implement SemCom.
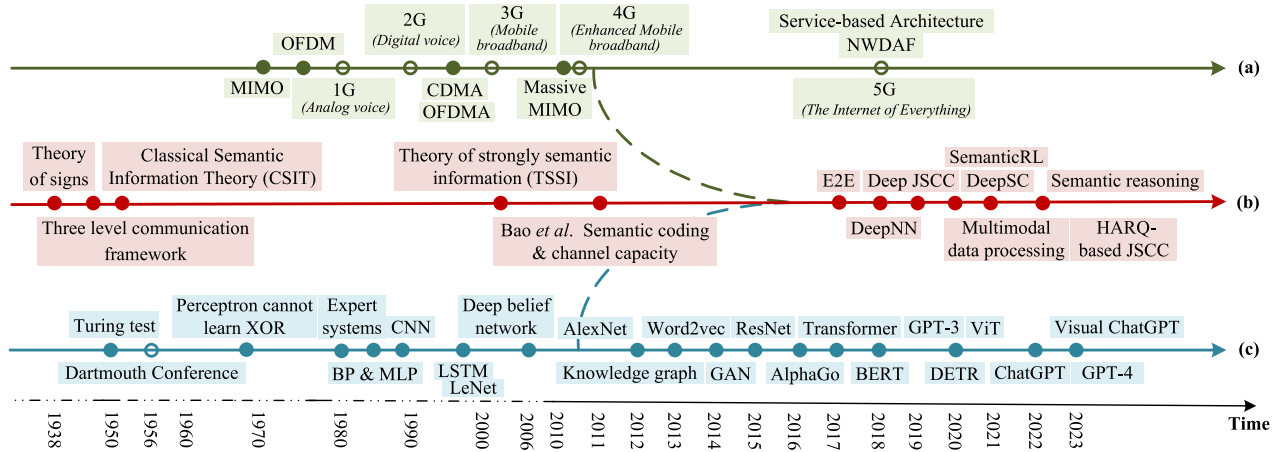
Fig. 2. Evolution of SemCom and its enabling ingredients: (a) Evolution of modern mobile communications (green dotted line), (b) Evolution of SemCom (red dotted line), and (c) Evolution of AI techniques (blue dotted line).

functions for non-differentiable processes like random communication channels [42], [43], [44]. Moreover, ChatGPT,[2] one representative large language model (LLM) fine-tuned from the generative pre-trained transformer (GPT)-3.5, has attracted intense interest due to its remarkable performance to handle contexts (e.g., the chain of thoughts). Last but not least, deep reinforcement learning (DRL) promises to learn by maximizing the total expected *reward* instead of relying on conventional loss functions. Learning by repetitive interactions with the environment, RL is believed to be an advanced form of AI, and shows its unprecedented superiority in decision-making tasks such as gaming, driving, and scheduling [45], [46], [47]. It is also commonly used to provide another alternative for non-differentiable functions since the reward can take any scalar value.

Notably, as data-driven DL remarkably boosts learning performance and raises increasing attention, conventional communications also benefit from this cutting-edge technique and evolve towards enhancing the awareness and cognition capability. In this new area as illustrated in Fig. 2(b), SemCom witnesses its new spring on the tacit intersection of both modern mobile communications and AI developments. Typically, AI/DL-based physical layer schemes [48], [49] and end-to-end communications [50], [51], [52] are now serving as important precursors of SemCom, demonstrating superior performance in a variety of tasks [53], [54], [55]. In other words, AI is becoming an indispensable ingredient for communications [56], [57]. When it emerges an urgent incentive to design a new communication paradigm (i.e., SemCom), motivated by the gradual maturity of current AI and DL technologies, it is natural for researchers to seek effective and promising AI-based strategies.

### C. What Is SemCom?

In a classical communication model, for a transmitted message $x$ encoded from the source signal $s$, the receiver tries to accurately decode a signal $\hat{s}$ from the received message $y$ in

terms of the bit-level accuracy. In particular, the received signal could be approximately derived from the channel information $H$ as $y = Hx + \varepsilon$, where $H$ is the channel gain, and $\varepsilon$ denotes the channel noise. In order to pursue the bit-level accuracy, various channel-oriented [44], [58], [59] and source-oriented methods [51], [52] have been proposed. The former category mainly focuses on signal estimation, channel detection, and resource allocation; while the latter puts emphasis on modulation and coding technologies, as well as constellation diagram design. Building on the accuracy principle, significant progress on reliable and efficient transmission has also been made, approaching the Cramer-Rao lower bound (CRLB) [60] and Shannon limit. However, this conventional design of communication systems faces a performance dilemma with the surging demand for data traffic and limited resources. It is expected and becomes imperative to revolutionize and explore alternative dimensions to solve the limitation and bottleneck of conventional communications [2].

Different from securing bit-level accuracy, Weaver proposes an alternative communication model that is designed following a completely different philosophy [6]. In Weaver's vision, communications can take place on three levels. The first level is designed for *reliably* transmitting symbols (i.e., bits) of data from the sender to the receiver, and has been extensively investigated. The second level, which is called the *semantic* level, aims to extract the *semantic* information and transmits it via a channel with both semantic and physical noise; while the third level focuses on *effectively* performing tasks instantiated by the sender. Following Weaver's path, Carnap [61] envisions that efficient SemCom can be realized by quantifying semantic information and considering semantic compression errors. The proposal of semantic transmission is attractive, as it naturally "compresses" the original message and only transmits the core meaning beneath any complicated symbolic representation. Due to the remarkable reduction in information overhead, this philosophy allows SemCom to stand out from high-rate & reliability-based designs and better touches the essence of communications [1], [4]. In a word, SemCom focuses on accurately transmitting the semantic meanings rather than bits.

---

[2]Interesting readers could refer to https://openai.com/blog/chatgpt/ for further details.
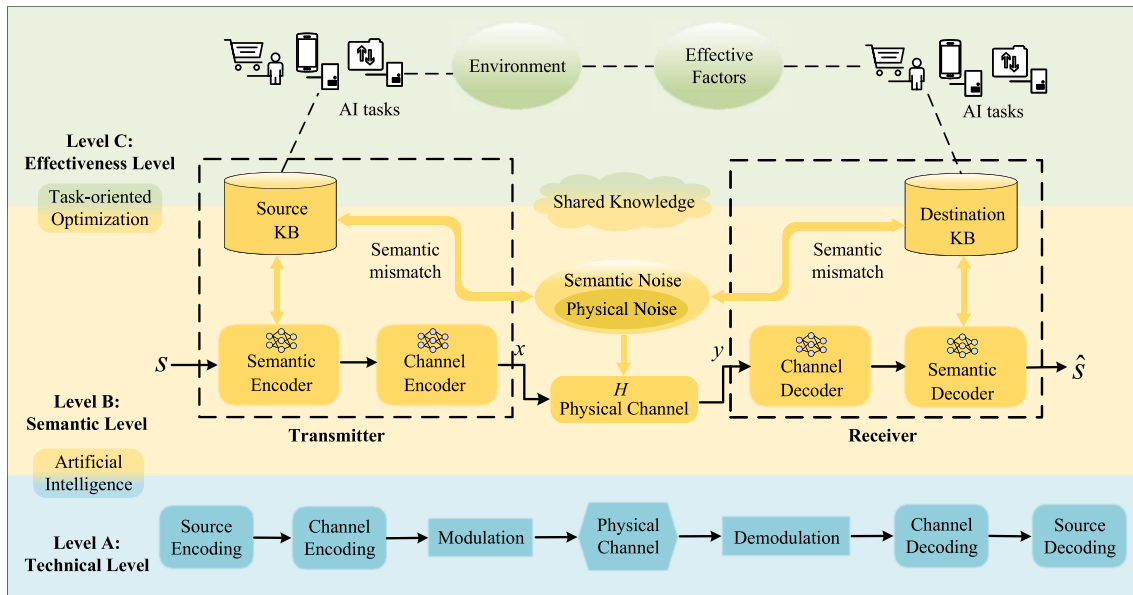
Fig. 3.   The three levels of communications.

Despite its promising features, SemCom did not gain sufficient attention for quite a long period. The reasons could be bi-folded. On one hand, Shannon believed that "the semantic aspects of communication are irrelevant to the engineering aspects" since the meaning of messages can be related to specific physical and conceptual entities [6]. As a result, semantic transmission inevitably affects the generality of mathematical model [11]. Moreover, owing to the much more urgent need for high-rate reliable communications at that time, researchers mostly follow the path of the CIT established by Shannon [62]. It has been pointed out in recent work that merely securing a symbolic recurrence, however, can be regarded as a means of preserving the syntactic information while lacking the capability of semantic awareness [4], [63], [64]. Analogous to the CIT, Carnap et al. [61] put forward the semantic information theory (SIT), which is henceforth referred to as classical semantic information theory (CSIT). Later, Floridi [65] extend CSIT to explore the measurements of information, i.e., *theory of strongly semantic information* (TSSI).[3] Recently, Bao et al. [66] propose a theoretical framework to measure the semantic information and channel capacity. These works clearly demonstrate the positive impact of semantics in communications, but are unable to provide theoretical analysis and feasible implementation methods for practical large-scale communication networks. Instead, benefiting from the impressive advance in data-driven AI or machine learning (ML) technologies, AI-powered SemCom is ushering into a new era and has achieved some preliminary results, on which will be elaborated later.

Fig. 2(b) highlights some early theoretical foundations including CSIT [61], TSSI [65] and semantic channel capacity [66], as well as representative achievements from the perspective of AI-powered SemCom. Generally, these SemCom works have significantly advanced Weaver's vision.

Beforehand, in order to summarize and keep our presentations succinct and easy to follow, we feel it essential to first outline the key components and prominent research directions of SemCom.

*1) Components of SemCom:* From top-layer design to bottom-layer implementation, we illustrate a mainstream structure in recently popular SemCom system in Fig. 3. Serving as a cornerstone for the second and third levels of Weaver's model, SemCom focuses on "how" to transmit and "what" to transmit in a more intelligent and semantically-compact way. Ideally in SemCom, the sender transmits the most useful information in terms of semantics but can ignore the redundant information. Meanwhile, the receiver in SemCom could understand and infer the expected semantic meanings correctly. This means that SemCom is only interested in semantic meanings and can tolerate the loss of certain semantically-irrelevant contents, wherein a pair of sender and receiver may not need to exchange all bits or pursue the bit-level accuracy. Equivalently, the bandwidth required for data transmission in SemCom can be saved significantly.

Meanwhile, to be consistent with Weaver's vision [6], SemCom can be further divided into SL-SemCom and EL-SemCom. In particular, the SL-SemCom systems mainly focus on semantic transmission for data reduction and put more emphasis on the delivery of content meaning rather than bit-level accuracy from a sender to a receiver, while the EL-SemCom systems consider more how to effectively utilize the semantic information at an appropriate time, thus facilitating successful task execution. Furthermore, Weaver believes that "the purpose of all communication is to influence the conduct of receiver", and "it is clear that communication either affects conduct or is without any discernible and probable effect at all". Therefore, we hold the view that all communications can fall into the scope of EL-SemCom from an engineering viewpoint. However, it is worth noting that EL-SemCom can achieve a goal on top of either extracted semantics

---

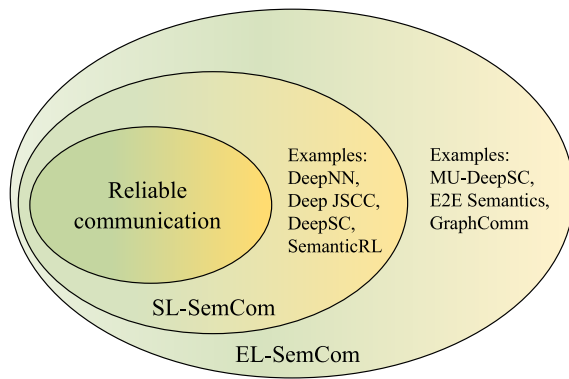[3]The details of CSIT and TSSI shall be given in Section III-B.

Fig. 4.   The relationship among the three levels of communications.

or simple bits. Typical works in the former case include MU-DeepSC [67], E2E Semantics [68], and GraphComm [37], while [69], [70] belong to the latter ones. In this way, when the goal of communications turns to accurate delivery of semantics only, it degenerates to SL-SemCom, such as *joint source-and-channel coding* (JSCC) on text or image transmission (e.g., DeepNN [5], deep JSCC [71]), DeepSC [53] and SemanticRL [54]. It is worth noting that the terminology "joint" in JSCC implies the training of the source and channel DNNs together, while the adopted DNN modules therein accomplish the functionalities like the source and channel codec in a separately designed yet end-to-end training manner. Hence, in Fig. 3, we still decompose it into a semantic encoder and channel encoder, so as to cover as many implementation means as possible. SemCom further falls back to the more fundamental and classical *reliable communications* when the data transmission focuses on the bit level accuracy. Intuitively, Fig. 4 depicts the relationship between the three-level communications.

Technically, SL-SemCom usually introduces an extra semantic encoder and decoder to process the information, while EL-SemCom adds some goal-oriented or task-oriented[4] factors to enhance the communication efficiency. As shown in technical and semantic levels in Fig. 3, SemCom performs various functions similar to conventional communications, and a SemCom architecture can encompass components such as *semantic encoder*, *channel encoder*, *channel decoder* and *semantic decoder*. Nevertheless, in contrast to independent modules in conventional communications to tackle statistical properties of symbols, SemCom is commonly contingent on *jointly* (i.e., end-to-end) trained DNNs on top of an additional, essential component (i.e., the KB). Specially, the semantic encoder extracts semantic features based on its KB from the input sequence *s*, during which the redundant information is filtered. Meanwhile, the channel encoder is responsible for generating the symbol (embedding representation) *x* to facilitate the subsequent transmission. Then, the extracted useful and relevant information will go through the physical channel. Similarly, the receiver is composed of a channel decoder for detecting the received symbol *y* from the received semantically noisy signals, and a semantic decoder for semantic-level or

---

[4]In this paper, we use the terminologies "task-oriented" and "goal-oriented" interchangeably.

effectiveness-level source estimation and recovery based on the KB [72]. Apparently, the most distinctive component of such a typical SemCom system attributes to a shared and continually updated KB, which can learn from the perceived environment and continuously evolve by repeated training or sharing via communications [19]. Another noticeable difference lies in the *semantic noise*, which extends the concept of physical noise in conventional communications, and can be defined as any noise that possibly leads to semantic infidelity [61], [66]. This extended definition implies that semantic noise could take into account the fact that even when a semantic concept is perfectly described and encounters no physical distortion during transmission, the receiver may not find a precise symbolic or semantic interpretation that matches the original meaning [61], [66] (typically when the KB of two communicating participants are not perfectly aligned). Therefore, besides physical noise and turbulence (e.g., Gaussian noise and multi-path fading), semantic noise also involves semantic mismatch, ambiguity and interpretation errors. Consequently, beyond simple *data compression*, SemCom strives to effectively combat the semantic noise and transmit the semantic meaning (e.g., maximizing the semantic similarity between the source data *s* and reconstructed data $\hat{s}$) with adequate encoding and decoding schemes. Moreover, the performance metrics such as semantic similarity and the compression ratio of the transmitted data constitute an important pillar of SemCom as well.

*2) Research Taxonomy:* In this paper, we take a coarse-to-fine routine where a macro research taxonomy is first provided for the community followed by detailed micro technical taxonomy based on the preliminary semantic information theory and the semantic similarity metrics in Sections III and IV, respectively.

As mentioned earlier, SemCom can be classified as SL-SemCom and EL-SemCom. As for the research of SL-SemCom, the exploration of semantics includes content semantics and channel semantics learned by DNNs shown in Fig. 5, wherein the former is realized over coding and modulation techniques, and the latter is contingent on transmission techniques. Specifically, the learning manner of content semantics can be categorized into implicit reasoning and explicit reasoning. In particular, explicit reasoning characterizes contextual probabilities with real entities or reasoning rules explicitly, wherein these real entities can be updated via pre-defined rules. We will mainly discuss mathematical probabilities and graph-modeled probabilities for this category [20], [38], [73]. On the contrary, the implicit reasoning methods, which are typically DL or DRL-based, learn how to express the semantics without a real mathematical probability distribution and infer with a set of learnable parameters supervised by human-crafted semantic similarity metrics as in Section IV-A, but in a discriminative manner. In particular, DL-based methods like JSCC-based frameworks [5], [53], [71], [74] assume the existence of differentiable objectives and adopt pre-trained DNNs as encoders and decoders. On the other hand, regardless of the differentiability, DRL-based methods [54], [75] model the SemCom as a Markov decision process (MDP) and dynamically learn the most appropriate
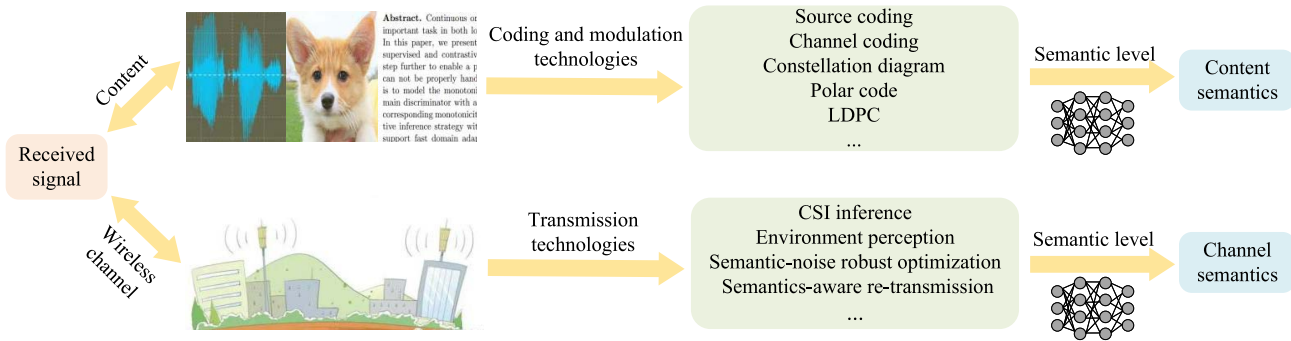
Fig. 5. A potential direction of communications evolving toward SemCom.

transmission methods by trial-and-error. Notably, without explicitly modeling the semantic expression and inference principles at the receiver side, these models embed the contextual probabilities in their parameters, but are more difficult to adapt to new and varying semantic environments (e.g., channels).

In addition to the aforementioned content-related semantics, some channel-related semantics are explored as well, so as to better adapt to the varying physical channel and further optimize the entire SemCom system. As a comparison, the content semantics of the transmitted information can be mined through semantic coding (SC), while by means of environmental perception, the channel semantics can be further inferred for semantics-aware transmission techniques and optimizing semantic noise-robustness.

In terms of EL-SemCom, we discuss the effectiveness of goal-oriented communications and some semantics-assisted networks. In this part, the SemCom pays more attention to the impact of semantic information on task execution, which can be evaluated in the corresponding effectiveness-level metrics in Section IV-B, such as energy efficiency [76], [77], *age of information* (AoI) [78], *value of information* (VoI) [79], [80] and the success or accuracy of task execution [37], [68], [81], rather than merely semantic transmission.

## III. SEMANTIC INFORMATION THEORY

### A. Background of Semantic Models

Before going through SIT, we would first recall the basic analytic structure of SemCom and how semantics are extracted and processed. Basically, in a SemCom system, KB, which consists of all semantically related "common knowledge" between the transmitter and the receiver, is deployed on both sides as a consensus. For example, the common knowledge for English speakers could be the lexical and grammatical English rules, while that for any two educated persons (even with no shared languages) comes from socialization. We assume that common knowledge is shared (but possibly not completely aligned) on both sides *before* any communications really takes place, i.e., an intrinsic property in SemCom. To start communicating, for any entity in the real world, a semantic concept would be first derived from the related observation $s$ and KB [82], after which this semantic concept is transformed into a symbolic representation or semantic message $x$. Next,

the message is encoded and transmitted through a physical channel, as illustrated in Fig. 3. Upon receiving the signal, a receiver could then decode and infer the corresponding semantic concept or symbol on the basis of the KB. Notably, a message $x$ can represent different types of media such as images, text (e.g., sentences) and audio, in which multiple semantic concepts can be described via simple or complex logical relations.

The SIT, hence, aims to quantitatively measure the amount of semantics, and explain how semantics flows in the communication process. Different from CIT which leverages the statistical probability of the message to measure the amount of information, in CSIT, Carnap et al. point out that the *information* carried by a sentence within a given language system is treated as synonymous with its content normalized in a certain way [61]. In this sense, the semantic information can be explicated by various measures of content based on logical probability, which is known as epistemic probability and refers to the degree of rational belief or confidence in a proposition within a given background knowledge and logical inference [83], [84]. Notably, in line with the consensus that the logical probability is defined within the context of a specific set [85], the logical probability discussed in CSIT and its extensions [61], [65], [66] is based on propositional logic within *specific* KB (i.e., sets), and thus significantly different from the independently defined statistical probability (and its induced Shannon theory), which is applicable to *any* given sets. Hence, in order to derive the proposition-oriented logical probability, the corresponding semantic universe and their induced messages (e.g., sentences) shall be firstly clarified. Specifically, taking the example of a language system $L_m^\pi$ with $m$ different individuals (e.g., things, events, or positions) and $\pi$ designating primitive properties for the individuals [61], [86]. Any message can be formed by atomic sub-messages with the help of five customary connectives (i.e., $\sim$ (Not), $\vee$ (Or), $\wedge$ (And), $\supset$ (If $\cdots$ then), and $\equiv$ (If and only if)). Analogously, complicated molecular predictors can be built by a combination of primitive properties and connectives. In other words, the "logical probability" of a message is measured by the likelihood of the atomic sub-messages inferred from the message using the logical connectives being true [61]. Mathematically, to stand in consistent with [61], [66], we will use $m(\cdot)$ to indicate the logical probability. For example, suppose a language system $L_2^1$ consisting of two independent

atomic sub-messages "A" and "B" (i.e., $m = 2$), which mean "true" or "false" with equal probability (i.e., 0.5). Hence, the logical probability for the message "A and B" is true is 0.25, while that for "A or B" is 0.75. Mathematically, $m(A \wedge B) = 0.25$ and $m(A \vee B) = 0.75$ [61], [66].

Furthermore, Bao et al. [66] extend this theory into semantic universe and introduce the interpretations from semantic source $s$ to message $x$. Specifically, a random *observation $S$* takes a value from a semantic world model $\mathcal{S}$ (i.e., a virtual world of all possible semantics), wherein the world model can be modeled as *interpretations* with a statistical probability distribution $\mu(\cdot)$, which can be regarded as a process of "semantic coding". Specifically, for a semantic message $x$,

$$m(x) = \frac{\mu(\mathcal{S}_x)}{\mu(\mathcal{S})} = \frac{\sum_{s \in \mathcal{S}, s \models x} \mu(s)}{\sum_{s \in \mathcal{S}} \mu(s)}, \qquad (1)$$

where $\mathcal{S}_x$ denotes a set that $s \in \mathcal{S}_x$ semantically entails $x$, i.e., $\mathcal{S}_x = \{s \in \mathcal{S}, s \models x\}$, $x$ is "true" in semantic world model $\mathcal{S}$ [66].

Based on the above basic background, some researchers have further studied and expanded the semantic information theory. Next, we will discuss in detail from the aspects of semantic information measures and semantic signal processing.

### B. Semantic Information Measures

*1) Semantic Quantification:* In CSIT, Carnap and Bar-Hillel [61] measure the information by the logical possibility defined above, rather than statistical distributions in CIT. In particular, for a message $x$ with the logical probability $m(x)$, the *content measure* of a message is defined negatively related to the logical probability as $\text{cont}(x) = 1 - m(x)$. Subsequently, the *information measure* $\inf(x)$ of one message $x$ is further defined as

$$\inf(x) = \log \frac{1}{1 - \text{cont}(x)}. \qquad (2)$$

(2) reveals the relationship between the amount of information and the semantic probability. Carnap's theory [61], [86] can be viewed as an inspiring and classical extension to Shannon's work.

However, CSIT measures the information for a statement based on the logical probability only, and ignores the influence of the truth on the amount of information, which will lead to a dilemma that the same logical probability of a false message and a true message imply the same amount of information. Meanwhile, it encounters the Bar-Hillel-Carnap (BHC) Paradox, that is, the ironical assignment of maximum information to the contradictions. For instance, for the message "A and not A", the corresponding logical probability equals 0, thus the information measure going to infinity. To address these issues, Floridi [65] propose TSSI to model the *informativeness* from both the polarity of a message $x$ (i.e., true or false) and the degree of discrepancy between $x$ and a given observation. Intuitively, for a received true message $x$ perfectly conforming to a given world model $\mathcal{S}$, the expected semantic

discrepancy is zero. Otherwise, depending on the truth value of $x$, a function $f(\cdot)$ is further introduced as

$$f(x) = \begin{cases} \frac{e}{l}, & \text{if } x \text{ is false;} \\ -\frac{n}{n_l}, & \text{if } x \text{ is true,} \end{cases} \qquad (3)$$

where $e$ is the number of false atomic sub-messages in the message $x$, and l is the total number of atomic sub-messages. $n$ denotes the cardinality of possible semantics consistent with message $x$, and $n_l$ is the number of all possible semantics that can be expressed in an $l$-length message. (3) can be interpreted as that for a true message, it is the most indiscriminate, and can be portrayed by *degree of vacuity*; while for a false message, $f(\cdot)$ depicts the *degree of inaccuracy*. Based on (3), [65] further defines the *degree of informativeness* as

$$\tau(x) = 1 - f^2(x). \qquad (4)$$

In addition, as the degree of vacuity and inaccuracy is largely dependent on the specific model of messages, it is not easy to quantify these two metrics for more complicated sentences rather than merely simple statements. As such, D'Alfonso [87] measures the semantic information by introducing the "truthlikeness", which can be further quantified by numerous distance metrics, such as Tichie-Oddie approach and Niiniluoto approach [87]. To some extent, these efforts [65], [87] partially solve the BHC paradox by effectively quantifying the semantic value of messages.

On the other hand, from the perspective of CIT [3], entropy, which is defined by the statistical probability of a random variable, could indicate the average level of information or "uncertainty" inherent to the random variable's possible outcomes. For example, when a random variable takes a deterministic value, we say it has no uncertainty (a low entropy). Consistent with CIT, Bao et al. [66] employ Shannon entropy to quantify information of semantic source (i.e., source semantic entropy), which can be formulated as

$$H(\mathcal{S}) = -\sum_{s \in \mathcal{S}} \mu(s) \log \mu(s), \qquad (5)$$

where $\mu(s)$ is the statistical probability of $s$ defined above. Analogized to the CIT, for each given semantic message $x$ encoded from $s$, Bao et al. [66] defines the semantic entropy of a message $x$ from the perspective of logical probability as

$$H_s(x) = -\log m(x). \qquad (6)$$

We use the subscript $s$ in (6) to distinguish a semantic entropy based on the logical probability in (1) from the statistical probablity-based classical entropy in CIT. Besides, it is worthwhile to note here that in terms of a logarithmic measure of a message, Bao's definition of the semantic entropy puts more emphasis on the informativeness of the message semantically encoded from the source, and is slightly abused from the standard definition of entropy in CIT, which takes the statistically average value.

In order to reduce semantic source entropy, D'Alfonso [87] exemplifies the positive contribution of the KB. Specifically, with the help of KB, even when the receiver is unable to directly decode a semantic message, the message could be correctly inferred given a set of logical relations, thus leading

to possibilities of lossless semantic compression with fewer encoded bits. In that regard, Choi et al. [20] further investigate the uncertainty and quantification of KB (denoted as $\mathcal{K}$) by using practical logic programming language, ProbLog, in which $\mathcal{K}$ consists of some logical clauses and each logical clause is annotated with a probability that indicates the degree of belief in the clause. Specifically, they define the *knowledge entropy* of a $\mathcal{K}$ as the uncertainty of answers it computes. As for a query $q$, a probability $p_q$ is computed with respect to $\mathcal{K}$ following the semantics of ProbLog (i.e., $\mathcal{K} \models q$). Therefore, the knowledge entropy is expressed as the average semantic entropy of query (message) $q$ computable from the KB $\mathcal{K}$ [20]. Mathematically,

$$H(\mathcal{K}) = \frac{1}{|\mathcal{H}_{\mathcal{K}}|} \sum_{q \in \mathcal{H}_{\mathcal{K}}} -H_{\mathcal{K}}(q), \qquad (7)$$

where $H_{\mathcal{K}}(q) = [p_q \log p_q + (1 - p_q) \log(1 - p_q)]$, and $\mathcal{H}_{\mathcal{K}}$ denotes the set of the terms which are the heads of all clauses in $\mathcal{K}$. For instance, suppose a knowledge base $\mathcal{K}$ with $\mathcal{H}_{\mathcal{K}} = \{a, b\}$. Accordingly, the possible queries that $\mathcal{K}$ can answer are $a$ and $b$. Besides, assume that $p_a = 0.2$, $p_b = 0.3$, and $p(b \to a) = 0.5$ (where $\to$ denotes "implies"). Therefore, the probability can be deduced as $p(\mathcal{K} \models a) = 1 - (1 - p_a)(1 - p_b \times p(b \to a)) = 0.32$ and $p(\mathcal{K} \models b) = 0.3$. Correspondingly, $H(\mathcal{K}) = \frac{1}{2}(H_{\mathcal{K}}(a) + H_{\mathcal{K}}(b)) \approx 0.839$. It can be observed that different from the logical probability that can be deduced from implicit reasoning rules, the probability in ProbLog is determined by the explicit rules defined in the knowledge base. Finally, (7) implies that in order to reduce the uncertainty of KB, a timely exchange and updating of the KB plays a critical role, and eventually contributes to less decoding uncertainty and more precise semantic transmission.

*2) Semantic Channel Capacity:* In this subsection, we present the related results on semantic channel capacity, and discuss the theoretical insights on semantic source coding and channel coding. As discussed in Section II-C1, the received message in SemCom may simultaneously suffer from physical noise and semantic mismatch & ambiguity, which leads to semantic infidelity. In other words, conventional reliable communications are concerned about the syntactic preservation of the message, whilst SemCom focuses on the semantic similarity between the input and output messages. In this sense, not all syntactic errors will lead to semantic errors due to the presence of semantic redundancy [66], while considering the possible incorrect inferring rules and unaligned KBs, a syntactic error-free process does not always guarantee perfect semantic decoding. To quantify the amount of information flow like in classical communication scheme, the semantic channel capacity $C$ for discrete memoryless channels can be developed in [66] as

$$C = \sup_{P(X|S)} \{I(X; Y) - H(S|X) + \overline{H_s(Y)}\}, \qquad (8)$$

where $I(X; Y)$ is the mutual information between the set of transmitted messages $X$ and received messages $Y$. $H(S|X)$ measures the semantic ambiguity as mentioned above; $\overline{H_s(Y)}$ denotes the average semantic entropy of received symbol $Y$, i.e., $\overline{H_s(Y)} = -\sum_y p(y) H_s(y)$, where $p(y)$ is the distribution of received $y$. Compared with Shannon theory, which defines the channel capacity as the supremum of mutual information between the channel input $X$ and channel output $Y$, i.e., $C = \sup \{I(X, Y)\}$, (8) states that there is a channel coding strategy through which the maximal probability of semantic error can be arbitrarily reduced within the semantic channel capacity. Notably, in SemCom, since both the encoding process and decoding process are dependent on inference rules and related local knowledge, the physical channel is no longer the only factor in shaping the semantic channel capacity. Instead, the relationship between "semantic channel capacity" and the standard "channel capacity" ($C = \sup I(X, Y)$) becomes intricate, as the exact value of $H(S|X)$ or $\overline{H_s(Y)}$ in (8) is governed by the practically implemented semantic inference rules and the shared knowledge base. It is also shown in [87] that a semantic encoder with low semantic ambiguity and a semantic decoder with strong inference capability, together with the help of a large shared KB, can jointly contribute to a better and precise SemCom system.

*Semantic source coding*, which is the process of mapping the observation $s$ into a symbolic representation (i.e., $x \in \mathcal{X}$), can be essentially formulated as a conditional distribution $P(x|s)$. Correspondingly, $P(x) = \sum_{s \in \mathcal{S}} \mu(s) P(x|s)$. Reference [66] points out that for a given semantic source, the source semantic entropy $H(S)$ and message entropy $H(X)$ satisfy

$$H(X) = H(S) + H(X|S) - H(S|X), \qquad (9)$$

where $H(X|S)$ and $H(S|X)$ measure the semantic coding redundancy and semantic ambiguity respectively. Equation (9) provides a general view of semantic source coding. It can be first concluded that depending on the coding redundancy and semantic ambiguity, message entropy can be larger or smaller than the source semantic entropy. Further, for a given semantic source $S$ with fixed coding redundancy $H(X|S)$, as we use more language symbols to describe, the message entropy tends to increase while the coding ambiguity decreases accordingly.

In order to quantify the beneficial side information (i.e., background knowledge and/or information obtained from previous communications), Xiao et al. [88] propose to extend (9) to characterize the positive impact of side information on decoding. Beforehand, the authors [88] give a showcase of side information by referring to a smart factory or smart city scenario, wherein a simple instruction sent by the controller may involve complex interactions between system components, and the side information corresponds to the underlying system structure and possibly related conditions to perform certain functions for individual components. Furthermore, as shown in Fig. 6, the authors [88] assume that the semantic information source $S$ includes some intrinsic states and features unable to be directly obtained by the encoder. Instead, the encoder can only leverage part of intrinsic observations, termed as the input signal $U$; while the decoder can have the privilege to access side information $E$ to assist the decoding from received signals. Therefore, in [88], (9) is reformulated as

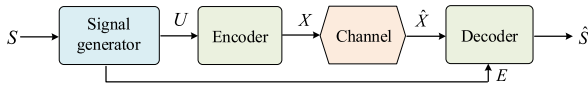$$I(S, U; Z|E) = I(S, U; Z, E) - I(S, U; E), \qquad (10)$$

Fig. 6. A SemCom model consists of a signal generator of semantic information source $S$, an encoder transforming $U$ to $X$, a channel $P(\hat{X}|X)$, a decoder to obtain $\hat{S}$ from the received $\hat{X}$ as well as the side information $E$.

where $Z$ is an auxiliary $E$-independent random variable introduced in the Wyner-Ziv [89] coding, denoting the output of a "test channel". Equation (10) implies the benefit of side information source under which the uncertainty of semantic source $S$ can be reduced at the semantic decoder side.

Taking into account the unique property of semantics that one semantic concept can correspond to multiple symbolic descriptions, it remains challenging for semantic source coding to determine an appropriate language system (typically those with low $H(X)$) to represent the semantic world. In that regard, [66] shows that due to the redundancy of semantic representation $X$, one can always build another language system $X'$ that is *equivalent* to the semantic representations but yields lower or equal message entropy. To experimentally approximate the ideal semantic source coding system, [81] provides a plausible solution by selecting the smallest subset of (learnable) semantic tokens via RL, which exhibits a remarkable encoding efficiency.

Besides semantic source coding, there also emerges significant research interest towards *semantic channel coding* to combat the semantic noise. A desired semantic channel coding strategy aims to eliminate the impact of semantic noise introduced by a physical channel and KB discrepancy. Practical semantic channel coding solutions often borrow ideas from classical channel coding schemes like introducing semantic redundancy, exploiting semantic noise-robust adversarial coding schemes, re-transmissions, semantic-correcting techniques, etc. [54], [74], [90], [91], [92]. It is noteworthy that due to the difficulty of theoretical semantic modeling for practical high-dimensional media, most of the existing semantic coding schemes follow a (typical) JSCC methodology. Meanwhile, it remains difficult to implement modular designs similar to the classical Shannon paradigm, which may be addressed as intensely new interest continues emerging.

### C. Semantic Signal Processing

Based on the aforementioned semantic information measures, which provide valuable insight into understanding why SemCom is feasible, we further discuss the design of semantic signal processing techniques in typical scenarios (e.g., semantic representations, compression, and reasoning).

*1) Semantic Representations:* Information theory-based semantic methods provide rigorous and convincing means of measuring the absolute amount of semantic information. In spite of its promising features, decomposing a general message (i.e., complex text, audio, images) into atomic semantic representations and modeling the complex semantic relations is rather challenging in reality. Till now, most semantic information theories are limited to addressing simple semantic concepts in a low-dimensional scenario. For real semantic representations, measuring the relative semantic information

(i.e., semantic similarity) provides another solution. It is also experimented in [54] that SemCom can be established from either semantic information rules or maximizing mutual semantic similarity.

Classical semantic similarity measurements are generally based on structured rules. One of the representative approaches is tree or graph-structured semantic representations like WordNet [93]. The distance of two semantic symbols can be measured by counting the path, depth or edit distance [94]. Besides building an explicit structured representation, statistical contextual metrics like *term frequency inverse document frequency* (TF-IDF) and the *information content* [95], [96] counts the similarity from semantic probabilities. In addition to the token's semantic similarity, more advanced approaches address phrase or sentence similarity by considering the contexts since similar semantic meaning often emerges from frequently used words. Some of the representative methods include *n*-gram contexts, customizing strategies for different syntactic and lexical structures (e.g., verb, noun, synonym), and machine learning-based decomposition like latent semantic analysis (LSA) and latent Dirichlet allocation (LDA) [97], [98], [99], [100].

While structured representations provide a plausible solution for simple semantic concepts, it is hard to generalize for complex relations and varied meanings. To address this issue, DL approaches build an automatic representation by exploiting the underlying semantic meanings from a unified semantic latent space [27], [28], [101]. Specifically, by comparing the distance of two semantic tokens or clauses in that latent space, DL-based representative learning methods, which can be interpreted as implicit probability modeling and automatic semantic atom discovery, can easily handle task-specific scenarios with flexible similarity measurements.

*2) Semantic Compression:* The aforementioned semantic information measures pave the way for lossless transmission. While in practical scenarios, semantic compression, which aims to transmit adequate information with certain tolerance of distortion at or below an acceptable level, is also of great importance. *Rate distortion function* belongs to one of the possible solutions to explore semantic compression based on information theory. In particular, for a simplified communication model under certain distortions, the optimization objective is to minimize the transmitted information volume between the semantic source $S$ and the semantic encoder output $X$ (denoted by mutual information $I(S; X)$), while reducing distortion function $d(s, x)$ as much as possible. To enable a reliable transmission under constrained expected distortion, the maximum bit rate can be solved by introducing a Lagrange multiplier $\beta$ and minimizing the following function:

$$\mathcal{L}[p(s, x)] = I(S; X)) + \beta d(s, x). \tag{11}$$

Equation (11) indicates that a smaller $\beta$ implies a higher compression ratio, and can be typically solved by Blahut–Arimoto algorithm [102].

However, as illustrated in (11), the rate distortion theory mainly focuses on the distance between source information $S$ and the encoded $X$, which ignores the information recovery process at the receiver side. In this case, IB theory [102]

extends rate distortion and provides theoretical guidance on balancing the distortion rate and compression rate. Typically in [102], IB aims to minimize the mutual information between $S$ and $X$ (which leads to a compression), whilst maximizing the mutual information between $X$ and $\hat{S}$ (which leads to fidelity). Mathematically,

$$\mathcal{L}[p(s, x)] = I(S; X) - \beta I\left(X; \hat{S}\right), \qquad (12)$$

whose optimal solution satisfies [102]

$$p(x|s) = \frac{p(x)}{Z(s, \beta)} \exp\left[-\beta \sum_{\hat{s}} \mathrm{KL}(p(\hat{s}|s)||p(\hat{s}|x)], \qquad (13)$$

where $Z(s, \beta) = \sum_x p(x) \exp(-\beta \, \mathrm{KL}\,[p(\hat{s}|s)||p(\hat{s}|x)])$ is the normalized function, and KL denotes the Kullback–Leibler divergence. Actually, (12) implies the means to combine semantic similarity metrics with the theoretical output. With the tunable parameter $\beta$, one can explore the tradeoff between the preserved semantic information and data compression at different resolutions. Hence, by defining proper distortion metrics, rate-distortion and IB-based loss can be mathematically converted to mean squared error (MSE) and cross-entropy (CE) loss and implemented with or without extra classifiers that model $p(\hat{s}|s)$ [53], [103].

*3) Semantic Reasoning:* Semantic reasoning is a concept that depicts the process of converting existing semantic cues to logical answers, which is mostly studied on the decoder side. A successful semantic decoder is expected to infer the true meaning from its local knowledge and received messages.

Seo et al. [38] investigate the progressive relationship among *entities*, *observations*, *concepts*, and *semantic symbols*, wherein they propose to model each of the semantic transmission processes (in the semantic flow) with contextual probabilities. Under their semantics-native framework [38], it is proved that the expected semantic coding length is bounded by the frequency of semantic concepts and decoding state; while semantic reasoning is equivalent to establishing a close communication context (termed as the *focal point* in their paper) for semantic agents on both encoder and decoder side. Iterative solutions and optimal transitions towards a semantics-native reasoning system are also involved.

Besides explicit mathematical probability-based representations, [55], [92], [104] model the reasoning process as completing a knowledge graph with entities and linked edges. Their experiments demonstrate that semantic contextual information can be successfully inferred with semantic concepts and logical rules, and verify that semantic reasoning is able to fill a few missing entities that can not be directly observed from the sent messages. In parallel, the authors in [20] additionally investigate the dynamic update of the reasoning KB, whose results demonstrate that benefiting from a self-updating KB, semantic inference contributes to the uncertainty reduction at the receiver side and even decreases the necessity for information delivery. These approaches can also be viewed as probability-based reasoning like [38], [88], wherein the task-specific probability distributions are now developed from large-scale training data.

## D. Discussions

As the aforementioned semantic information theories contribute to a preliminary yet meaningful exploration of SemCom, they are still evolving and not fully ready to become a complete theory. For example, most of the existing works succeed in modeling the semantic information with a pre-defined, proposition-oriented (and usually assumed to be easy-to-compute) logical probability, under which the basic semantic relationships can be theoretically measured but is hard to get generalized for more complex scenarios in real-life applications. Meanwhile, as a less-investigated yet critical component in the reasoning process, researches on the KB are scattered. A few key questions related to semantic KB, for instance, to what extent a shared KB influences the communication process; how to quantitatively model the semantic flow in partially-shared KBs, remain elusive and deserve more in-depth exploration. Meanwhile, theories towards a general semantic information theory, such as semantic security & robustness, semantic efficiency & generalization trade-off, and semantic computability are also expected to become important research topics.

Besides theoretical advances, experimental contributions serve as another strong and prominent driving force in SemCom. Among all techniques, there are encouraging signs that DNN-powered techniques are currently a promising direction and dominating methodology. These approaches take advantage of representative learning and provide a practical solution to measure semantic probability and semantic similarity. Opportunities can also be found in other techniques like JSCC, advanced AI pipelines (RL, adversarial learning, federated learning (FL), etc.), and game theory, and have demonstrated superior capability in handling a variety of challenging tasks including varying channels, KG based semantic reasoning, semantic compression [53], [54], [74], etc.

It is worth mentioning that advances in theoretical and experimental studies are complementary. Exploration of the experimental methodology provides plausible ways for evaluating and verifying the semantic theory. On the other hand, theoretical advances pave the way for discovering new opportunities for new SemCom systems. Unfortunately, both methodologies progress in a quasi-independent manner in the literature, where theoretical insight is generally a "later starter" due to the difficulties in semantic representation. In that regard, despite the initial success of SemCom, the semantic information theory [61], [65], [66], [86] is often limited to the measurement of semantic information or uncertainty, yet fails to directly gauge the semantic similarity of messages, with a few exceptions like the adoption of the IB theory [102] and its variants into the loss functions to optimize the DNN-enabled SemCom systems [53], [105]. Therefore, on one hand, it implies the necessity to continually explore the theoretical output of SemCom faced with semantic noise, given the apparent benefits of theoretical advances. For example, the integration of ProbLog [106] and practical SemCom systems bring a substantial performance boost in explicit reasoning models [20]. Meanwhile, the analysis of semantic coding and channel capacity also provides additional interpretability for

DNN-based coding schemes. On the other hand, in order to avoid the inconsistency between theoretical results and implementation means, it becomes inevitable to incorporate semantic similarity metrics with ease of computation and adoption, which shall be given later in Section IV.

To summarize, SIT tries to quantify how semantic information flows and how it is processed, along with some related features for the deployment purpose. These powerful theoretical tools will guide the design of future communications systems.

## IV. SEMANTIC SIMILARITY METRICS, DATASETS, AND TOOLKITS

During the establishment of either SL-SemCom or EL-SemCom, the semantic similarity metrics, datasets, and toolkits belong to one of the most important pillars. Basically, attributed to a part of SIT, semantic similarity metrics contribute to evaluating the differences of meaning rather than bits between the transmitted and recovered content, and are largely involved as the optimization objective in the design and measurement of SemCom systems. As for semantic similarity metrics, considering the differences among different types of media (i.e., text, image, and speech), metrics could be rather different, and could be classified dependent on the practical scenarios, such as error-based transmission metrics, data compression metrics (i.e., to measure the reduction ratio of data volume), and goal-oriented effectiveness metrics (e.g., age & value of information). Similarly, datasets and toolkits can be categorized as SL-SemCom-oriented (e.g., text, images, speech, and videos) and EL-SemCom-oriented. Table II summarizes the typical metrics in SemCom. Meanwhile, the datasets and toolkits frequently used in existing SemCom systems are provided in Table III as well.

### A. Semantic-Level Metrics

*1) Text Transmission Similarity:* Besides MSE or CE for training the DNN, some text similarity metrics can be used to quantify the difference between the transmitted $s$ and received $\hat{s}$. As a widely-accepted metric for text similarity, the *average semantic distortion* measures the distortion between the transmitted word and the recovered one [107]. Meanwhile, inheriting from measuring the speech-to-text accuracy of automatic speech recognition systems, *word error rate (WER)* [5] is another common metric in this field. Unfortunately, both of them lack the understanding capability of word ambiguity and pay more attention to the distance between individual words.

In order to avoid this word ambiguity issue and further measure the sentence similarity, the frequency of $n$-gram co-occurrence between the transmitted source $s$ and the recovered $\hat{s}$ is often leveraged. Specifically, the terminology $n$-gram refers to a group of $n$ consecutive words in a sentence. Essentially, a higher $n$-gram can reflect more contextual information, while as a special case of $n$-gram, 1-gram score totally ignores the contextual semantics and is equivalent to a word-level measurement. Typically, high-order $n$-gram similarities are often adopted in the literature [30], [54] as they reflect more on the semantic contexts.

On top of $n$-gram, several sentence similarity metrics like bilingual evaluation understudy (BLEU) and consensus-based image description evaluation (CIDEr) are proposed to measure the degree of consistency between the "candidate messages" and ground-truth references [53], [54], [105]. Benefiting from these similarity metrics, the authors in [54] further put forward a SemanticRL framework that directly optimizes the semantic distances rather than via word-level supervision.

In particular, BLEU score [108] is initially proposed for machine translation and has quickly become a dominating method against previous state-of-the-arts at that time. Without loss of generality, let $l_s$ and $l_{\hat{s}}$ denote the length of $s$ and $\hat{s}$. The BLEU score $\Theta_{\mathrm{BLEU}}$ in the logarithmic domain is defined as

$$\Theta_{\mathrm{BLEU}} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{4} u_n \ln p_n\right), \qquad (14)$$

where $u_n$ and $p_n$ denote the weight and the frequency of co-occurrence for $n$ consecutive words, respectively. Notably, it is commonly recommended that $n$ is up to 4 and $u_n$ takes the average value [108], i.e., $u_n = 1/4, \forall n \in \{1, \ldots, 4\}$. Besides, in order to avoid the misleading high matching degree under cases where a short candidate sentence is part of the long reference one, *brevity penalty* (BP) is introduced to punish short sentences and can be formulated as

$$\mathrm{BP} = \begin{cases} 1, & \text{if } l_{\hat{s}} > l_s; \\ e^{1-l_s/l_{\hat{s}}}, & \text{if } l_{\hat{s}} \le l_s, \end{cases} \qquad (15)$$

The BLEU score $\Theta_{\mathrm{BLEU}}$ ranges from 0 and 1, and a higher score indicates larger sentence similarity. Taking an example of the reference sentence "Going to play basketball in this afternoon?" and the received sentence "Going to play basketball this afternoon?", it is apparent that $l_s = 8$ and $l_{\hat{s}} = 7$, thus $\mathrm{BP} = e^{1-\frac{8}{7}} = e^{-1/7}$. As for the calculation of $n$-gram, for instance, the co-occurrence of 2-gram in these two sentences include "Going to", "to play", "play basketball", "this afternoon", thus $p_2 = 4/6$. Similarly, $p_1 = 6/7$, $p_3 = 2/5$, and $p_4 = 1/4$. Hence, $\mathrm{BLEU} = \mathrm{BP} \cdot \exp((\ln p_1 + \ln p_2 + \ln p_3 + \ln p_4)/4) \approx 0.4238$.

On the other hand, the CIDEr score [109], which has also proven effective for image captioning task [30], is adopted in [54] as a semantic metric for texts. Specifically, in order to calculate the accuracy and diversity of decoded $n$-gram phrases, CIDEr score $\Theta_{\mathrm{CIDEr}}$ measures the cosine similarity of the TF-IDF weights and can be formulated as

$$\Theta_{\mathrm{CIDEr}} = \sum_{n=1}^{4} u_n \frac{\|g_n(s) \circ g_n(\hat{s})\|}{\|g_n(s)\|\|g_n(s)\|}, \qquad (16)$$

where $\|\cdot\|$ denotes an $l_2$-norm and $(\circ)$ denotes the Hadamard product. Besides, the TF-IDF vector $g_n$, corresponding to all $n$-grams of length $n$, indicates the frequency of $n$-gram.

Nevertheless, both BLEU and CIDEr scores are still largely affected by polysemy. In that regard, [53] proposes a BERT (bidirectional encoder representation from transformers) [28]-based similarity metric (i.e., BERT-SIM [53]) as

$$\Theta_{\mathrm{BERT\text{-}SIM}} = \frac{\|B_\phi(s) \circ B_\phi(\hat{s})\|}{\|B_\phi(s)\|\|B_\phi(\hat{s})\|}, \qquad (17)$$

TABLE II
A SUMMARY OF SEMANTIC SIMILARITY METRICS FOR SEMCOM

| Metric type | Semantic Similarity Metrics | Advantages | Drawbacks | Ref |
|---|---|---|---|---|
| **Semantic-Level Metrics** | Average Semantic Distortion | It utilizes the semantic distances based on lexical taxonomies as a distortion measure. | It only computes the semantic similarity between individual words, and is difficult to calculate for large data sets. | [107] |
| | Word Error Rate (WER) | It can reflect the semantic similarity to a certain extent and is easy to calculate. | It does not capture the effects of synonyms or semantic similarity. | [5] |
| | Bilingual Evaluation Understudy (BLEU) | It considers the linguistic law that semantically similar sentences are consistent in the semantic space. | It can only compare the differences between words in two sentences rather than their semantic meaning in sentences. | [53] |
| | Consensus-based Image Description Evaluation (CIDEr) | Compared to BLEU, it does not evaluate semantic similarity on the basis of a reference sentence, but a group of sentences with the same meaning. | It concentrates more on the middle part of a sentence, and thus the middle part possesses more $n$-gram weight. | [54] |
| | BERT-based Similarity (BERT-SIM) | It can explain semantics at the sentence level due to BERT's sensitivity to polysemy. | It is not easy to generalize the pre-trained model on others. | [53] |
| | Peak Signal-to-Noise Ratio (PSNR) | It is easy to calculate and understand, and can roughly reflect the image similarity. | It is not always consistent with human perception. | [111] |
| | Structural Similarity Index Method (SSIM) | Compared with PSNR, SSIM is more consistent with human perception in image quality evaluation. | It reflects a higher evaluation than the actual. | [110] |
| | Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) | It exhibits distinctive robustness to noise. | Based on feature extraction, the spatial relationship between features cannot be fully utilized. | [118] |
| | Signal to Distortion Ration (SDR) | It is easy to calculate and can reflect the quality of voice to a certain extent. | Its evaluation results are sensitive to the volume of audios. | [120] |
| | Perceptual Evaluation of Speech Quality (PESQ) | Its evaluation is objective and close to human perception. | It has high computational complexity. | [120] |
| | Semantic Mutual Information (SMI) | It can reflect the semantic-level distortion. | SMI needs to be estimated by extra module. | [127] |
| | Reduction ratio of the number of bits | It can intuitively reflect the reduction of data volume. | High computational complexity | [126] |
| | Total number of bits | It can intuitively reflect the amount of data volume. | High computational complexity | [125] |
| **Effectiveness-Level Metrics** | Age of information (AoI) and peak age of information (PAoI) | It can reflect the freshness of information. | It may misjudge the value of information. | [78] |
| | VoI based approach | It can capture the value of information. | It is not easy to design the VoI function for some complicated systems. | [79], [80] |
| | Age of Incorrect Information (AoII) | It combines the age and value of information to reflect the significance of updates. | The optimal estimation of the penalty function needs to be further investigated. | [129] |

where $B_{\phi}$ is a pre-trained BERT model [28]. Apparently, $\Theta_{\text{BERT-SIM}} \in [0, 1]$, and an increase in the BERT-SIM implies a more accurate semantic recovery. However, the BERT model consumes significant resources for training and might be difficult to be extended to different tasks. As a remedy, the authors in [91] introduce a transformer-based DNN called *Sim32* to detect the meaning errors in the received sentences, which can tolerate some recovery error of several words as long as the meanings of the context remain unchanged. Sim32 could be effectively adopted in saving transmitting resources, since quite a large number of bit-lossy yet semantics-reserving sentences no longer need to be re-transmitted.

*2) Image Transmission Similarity:* In the context of image processing, *peak signal-to-noise ratio* (PSNR) [71] and *structural similarity index method* (SSIM) [110] are the most popular metrics for evaluating the image similarity. In particular, PSNR is the most widely adopted metric in early SemCom

systems [71]. However, due to the existence of estimation absolute errors, PSNR is error-sensitive and sometimes yields perception-inconsistent evaluation results [111]. As a result, it can not sufficiently incorporate the perception characteristics from the human vision perspective.

On the other hand, belonging to a perception-based model, the SSIM is established on the idea that the pixels have strong inter-dependencies especially when they are spatially close. By leveraging these dependencies that carry information about the structure of the objects in the visual scene, the SSIM score $\Theta_{\text{SSIM}}(s, \hat{s})$ compares $s$ and $\hat{s}$ in terms of the *luminance*, *contrast* and *structure* [110] at the pixel level, and can be computed as

$$\Theta_{\text{SSIM}}(s, \hat{s}) = (\rho_{\text{l}}(s, \hat{s}))^{\lambda_1} \cdot (\rho_{\text{c}}(s, \hat{s}))^{\lambda_2} \cdot (\rho_{\text{s}}(s, \hat{s}))^{\lambda_3}, \quad (18)$$

where the functions $\rho_{\text{l}}$, $\rho_{\text{c}}$ and $\rho_{\text{s}}$ reflect the perceived changes in terms of luminance, contrast, and structural patterns

TABLE III
A SUMMARY OF DATASETS AND TOOLKITS FOR SEMCOM

| Applications | Datasets | Descriptions | Toolkits | Reference Samples |
|---|---|---|---|---|
| **Text** | Proc. European Parliament[a] | The English version has 2 million sentences, each with 4 to 30 words (or equivalently 53 million words). | TensorFlow PyTorch | [5] [53], [54], [74], [91] |
| | Tatoeba Project[b] | A large database of sentences and translations (e.g., from English to French). | \ | [105] |
| | WebNLG English dataset[c] | It comprises of sets of triplets describing facts with around 600 entities and 20 relations. | \ | [126] |
| | WordNet-WN18[d] | The subset of WordNet with 18 scraped relations for roughly $41,000$ synsets. | \ | [92] |
| | WMT 2018 Chinese-English news track[e] | It contains $202,221$ Chinese-English pairs for training while $50,556$ pairs for testing. | \ | [138] |
| **Speech** | LJSpeech[f] | A corpus of English speech at the sampling rate of $22,050$ Hz. | TensorFlow | [119] |
| | Librispeech[g] | A speech-to-text library based on public-domain audiobooks containing 960 hours of speech for training and $2,703$ utterances for testing. | \ | [139] |
| **Image** | CIFAR-10[j] | It consists of $60,000$ $32 \times 32$ color images in 10 classes. | Tensorflow \ TensorFlow, Keras | [71], [140] [90], [141], [142] [143] |
| | MSCOCO[k] | A large-scale object detection, segmentation, key-point detection, and captioning dataset $123,287$ images. | PyTorch | [144] |
| | ADE20K[l] | Scene parsing benchmarking dataset with $27,574$ images of 150 semantic labels. | PyTorch | [144] |
| | MNIST[m] | A handwritten digit dataset consists of $60,000$ $28 \times 28$ images. | \ | [142] |
| | PASCAL-VOC2012[n] | A object class recognition dataset consists of $2,913$ RGB images with sizes $513 \times 513$ in 20 categories. | \ | [142] |
| | CUHK03[o] | A pedestrian re-identification dataset including roughly $14,000$ images of over 1400 pedestrians. | \ | [145] |
| | Stanford Online Products[p] | It consists of $120,053$ online product images in $22,634$ categories. | \ | [138] |
| | CLEVR[q] | A diagnostic dataset with $70,000$ images and $699,989$ questions that capably tests a range of visual reasoning abilities. | \ | [138] |
| **Video** | UCF101[h] | It contains $13,000$ clips (27 hours) of realistic action videos from YouTube in 101 action categories. | PyTorch | [146] |
| | Vimeo-90k[i] | It consists of $89,800$ video clips with a large variety of scenes and actions. | \ | [147] |
| **UAVs** | \ | A self-defined 2D $3,750m \times 3,750m$ grid with 4 UAV, a server and several users. | Self-defined | [148] |
| | Drone detection dataset[r] | It consists of 7 different types of UAVs, and includes three flight modes (i.e., switched on, hovering, and flying). | \ | [149] |
| **IoTs** | \ | The CartPole consists of a cart that can be moved to the left or right and a pole positioned vertically above it; while Acrobot is a two-link pendulum that only activates at the second joint. | Gym (e.g., Cartpole and Acrobot)[s] | [150] |
| | \ | The IC3Net toolkits include traffic junction and predator prey Environments. | IC3Net[t] | [37] |

[a] https://www.statmt.org/europarl/  [b] http://www.manythings.org/anki/  [c] https://paperswithcode.com/dataset/webnlg

[d] https://wordnet.princeton.edu/  [e] https://statmt.org/wmt18/translation-task.html  [f] https://keithito.com/LJ-Speech-Dataset/

[g] https://librivox.org/  [h] https://www.crcv.ucf.edu/data/UCF101.php  [i] http://toflow.csail.mit.edu/  [j] http://www.cs.toronto.edu/~kriz/cifar.html

[k] https://cocodataset.org  [l] http://groups.csail.mit.edu/vision/datasets/ADE20K/  [m] http://yann.lecun.com/exdb/mnist/

[n] http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html  [o] http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html

[p] https://github.com/rksltnl/Deep-Metric-Learning-CVPR16  [q] http://cs.stanford.edu/people/jcjohns/clevr/

[r] https://ieee-dataport.org/open-access/dronedetect-dataset-radio-frequency-dataset-unmanned-aerial-system-uas-signals-machine

[s] https://github.com/openai/gym  [t] https://github.com/IC3Net/IC3Net

between two images. Besides, $\lambda_1$, $\lambda_2$ and $\lambda_3$ denote the exponential coefficients. However, the performance of SSIM becomes less effective when it is used to rate blurred and noisy images. Henceforth some variants of SSIM, such as 3-SSIM [112] and FSIM [113] are also proposed.

In order to evaluate the human perception consistency between generated images and natural images, some *distribution divergence*-based metrics like FID (Fréchet inception distance) [114] and KID (Kernel inception distance) [115] are proposed by computing the Wasserstein distance [116] instead of the MSE (and its variants), as the Wasserstein distance measures the distance between probability distributions on a given metric space, and could capture more information than the sole comparison of specific realizations between *s*
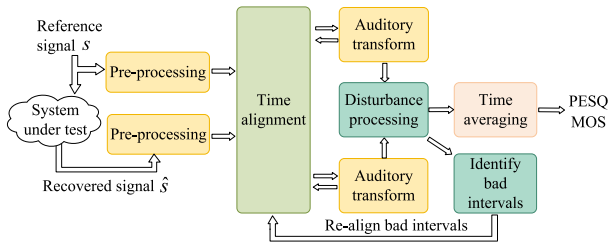
Fig. 7.    The procedures of PESQ [122].

and $\hat{s}$ in MSE [117]. Meanwhile, FID and KID compare the distribution of generated images with the distribution of a set of real images ("ground truth"). More specifically, lower FID proves to correlate well with higher-quality images, while KID measures the maximum mean discrepancy (MMD) in the feature space of a classifier, and a lower KID score implies better sampling quality [118]. Both of them are adopted to evaluate the efficiency of semantic exchange. However, based on feature extraction, FID and KID can not fully utilize the spatial relationship between features.

*3) Speech Transmission Similarity:* Given the similarity between the audio data and the text data, similar metrics have also been applied to measure the accuracy between speech signals. Quite similar to WER in text similarity evaluation, the *character error rate* (CER) metric operates on characters instead of words, so as to evaluate the accuracy of speech recognition [119]. On the other hand, in order to measure the semantic discrepancy, the authors in [120] adopt *signal to distortion ratio* (SDR) [121] as the semantic metric, which is defined on top of MSE and can be formulated as

$$\Theta_{\text{SDR}} = 10\log_{10}\left(\frac{\|s\|^2}{\|s - \hat{s}\|^2}\right). \tag{19}$$

Notably, as implied in (19), SDR and MSE are correlated and can be easily interpreted from one metric to the other. Specifically, a lower MSE often leads to higher SDR. Besides, as SDR produces clear performance differences, it can be conveniently used to optimize the DNNs.

Highlighted by International Telecommunication Union for Telecommunication Standardization Sector (ITU-T) recommendation P.862 [122], the *perceptual evaluation of speech quality* (PESQ) [123] takes into account the short memory in human perception and provides a popular tool for assessing speech quality [120]. As shown in Fig. 7, PESQ first preprocesses the reference source $s$ and the reconstructed one $\hat{s}$, after which the two signals are aligned in time to correct the time delays, and undergoes the auditory transform into a representation of perceived loudness in both time and frequency domain. Then the disturbance in the frequency and time domain is aggregated by the "disturbance processing" module, and the output is further mapped to compute the PESQ subjective mean opinion score (MOS) score [123]. In addition, if the "time alignment" module fails to correctly identify a delay change, the related bad intervals will be realigned and the PESQ MOS score will be recalculated. Clearly, in spite of its accuracy, PESQ retains high computational complexity, which on the other hand limits its light-weight applications.

Besides these carefully-designed empirical metrics, the distributions of two speech signals can also be quantified, providing another statistics-related interpolation. For instance, Fréchet deep speech distance (FDSD) and kernel deep speech distance (KDSD) [124] are adopted in the SemCom system [119] to assess the speech similarity and evaluate the quality of recovered speech signals. Specifically, a lower FDSD or KDSD score indicates higher similarity between real and synthesized speech sample sequences.

*4) Ratio of Data Compression:* Considering that SemCom aims to maximize the semantic similarity with as minimal amount of data as possible, the above error-based semantic similarity metrics focus more on the measurement of the semantic similarity, but neglect to measure the data volume before and after semantic transmission. As implied in (11), a higher compression ratio means lower distortion, but requires corresponding more bandwidth. Therefore, it entails a tradeoff to minimize distortions with the smallest possible bit rate. Hence, besides error-based semantic similarity metrics such as BLEU, PSNR and SDR, subject to a particular semantic loss, it is still essential to design some semantic similarity metrics to measure the reduction of data volume compared to conventional communications.

In the context of text compression, Wang et al. [125] propose a metric named *reduction ratio of the number of bits* to compute the ratio of data size between the semantic information and the classical encoded text. Specifically, assuming that each English letter is encoded by 8 bits and each token includes 10 letters. Therefore, a token can conventionally be encoded with 80 bits by classical 16-dimensional encoding vectors, while the SemCom method can represent the semantics of the original token by a DNN, wherein a decimal is encoded into 32 bits. Compared with the classical methods, the simulation results demonstrate that their proposed SemCom schemes can actually reduce the data size by up to 41.3%. Moreover, Zhou et al. [126] adopt a similar metric *total number of bits* to evaluate the compression ratio. Jiang et al. [91] leverage the metric of *average bits* consumed per sentence to measure the efficiency of the SemCom. On the other hand, as for the image compression, Sun et al. [127] adopt the *semantic mutual information* (SMI) and *accuracy* (ACC) simultaneously to evaluate the semantic-level distortion, wherein the SMI is estimated by utilizing CLUB [128] due to its excellent accuracy.

### B. Effectiveness-Level Metrics

On top of the semantic-level metrics, effectiveness-level metrics are primarily leveraged to evaluate the contribution of information to the task accomplishment.

*1) Age of Information-Based Metrics:* Different from the traditional metrics which ignore the temporal impact of one message, AoI is put forward by describing the time lag between the current time and timestamp of packets [78]. In this end, the receiver is more interested in the freshness of the received information, so as to avoid repeated transmission of useless information and bandwidth wastage [129]. AoI has been widely used in cyber-physical systems, such as
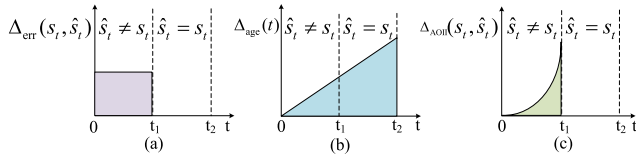
Fig. 8. Example of the different penalty functions: (a) Error-based penalty function, (b) AoI-based penalty function, (c) AoII-based penalty function [129].

unmanned aerial vehicle (UAV) [130], [131], [132] and sensor networks [133], [134]. Notably, the definition of AoI can be slightly different across tasks. For example, [78] introduces AoI as the average age of packets during a period, while the *peak age of information* (PAoI) [135] represents the peak age of packets during a period.

Although AoI can indicate the freshness of data in various applications, it also encounters some critical flaws. Taking the example of sensor networks involving temperature monitoring and control [129], communications could act as an anchor to ensure the controller responds swiftly to any abnormal temperatures. As shown in Fig. 8, intuitively, the error-based metrics are penalized only when $s \neq \hat{s}$. It can also observed from Fig. 8(b) that although $s_t = \hat{s}_t$ in the time interval $t \in [t_1, t_2]$ and the monitor perfectly estimates the actual process, the system still penalizes the AoI metric since messages become less valuable as time goes by [136]. In other words, the AoI-based metrics focus more on the freshness of data and are more suitable for applications with stringent time requirements.

*2) Value of Information-Based Metrics:* Similar to AoI, the exact computation method of *Value of information* (VoI) is task-dependent as well [80]. However, different from the AoI which is concerned about the freshness but ignores the content itself, VoI measures the contribution of content towards achieving a specific goal, and becomes especially popular in the networked control systems. For example, in [79], VoI is defined to minimize the quadratic norm of estimation errors, and can be regarded as a measure of the uncertainty reduction from the information at the receiver after a successful transmission. Moreover, VoI can be utilized as a part of a prioritizing scheduler [80] in which the essential information with a larger VoI will be given higher priority.

*3) Age of Incorrect Information-Based Metrics:* In addition to the aforementioned metrics, different SemCom systems also design specific measurements according to different tasks. The *age of incorrect information* (AoII) [137] deals with the shortcomings of the AoI and conventional error-based metrics, as it neatly extends the notions of information updates to "informative" updates [137] and focuses on the usefulness to the communication's goal. Afterwards, [129] takes a step forward by further considering the objective of communications towards enabling the SemCom. In this sense, as semantics can reflect the significance and usefulness of information, in order to reflect the effectiveness to this task, AoII is introduced to incorporate the semantics of data more meaningfully. More specifically, AoII measures the captured information by two aspects (i.e., information-penalty and time-penalty function), which can be formulated as $\Theta_{\text{AoII}}(t) = \kappa(t) \times \iota(s_t, \hat{s}_t)$, where

$\kappa(t)$ is a monotonically increasing time penalty function, while $\iota(s_t, \hat{s}_t)$ denotes the information penalty function to reflect the difference between the actual data $s_t$ and estimate data $\hat{s}_t$ at time $t$. Meanwhile, in different scenarios, the functions $\kappa(\cdot)$ and $\iota(\cdot)$ need to fit the interest of applications. As illustrated in Fig. 8(c), in the time interval $[t_1, t_2]$, compared to the error and AoI-based metrics, AoII-based metrics allow to capture a more meaningful purpose of the transmitted data. The authors also show that the metrics adopted in many goal-oriented communications can be regarded as variants of the AoII after tweaking specific parameters [129].

*4) Miscellaneous:* In addition, some goal-oriented metrics are also proposed to achieve the prescribed purpose. For instance, in order to evaluate the sample updates, the authors in [68] adopt the metrics like real-time reconstruction error and the cost of actuation error. In particular, the reconstruction error reflects the discrepancy in real-time data exchange, while the cost of actuation error captures the significance of the error at the actuation point considering that there may have a mismatch between the original source $s$ and the reconstructed one $\hat{s}$. For example, assume that there exist three states for the original source. If $s$ belongs to State 1, while $\hat{s}$ is wrongly projected to other states (e.g., State 2 or State 3), an actuation error occurs. Besides, compared to State 2, State 3 deviates further from State 1. Therefore, the cost of actuation error for $\hat{s} = \text{State 3}$ will be larger than that for $\hat{s} = \text{State 2}$. When the $s$ and $\hat{s}$ belong to the same state, there is no actuation error [68].

### C. Datasets and Toolkits

On the basis of semantic similarity metrics above, datasets and toolkits, which are typically originated from authoritative institutions and organizations like Google, OpenAI and Facebook, manifest their significance in the development of SemCom as well.

*1) SL-SemCom Datasets and Toolkits:* As the primary case to demonstrate the potential of SemCom, text-oriented SemCom has rather abundant datasets, which can be categorized into varying-length sentence datasets and KG-based entity-relationship-entity triplets. In particular, the former category includes *Proceedings of the European Parliament* and *WMT 2018 Chinese-English news track* [53], [54], [74], [91], [138], while the latter encompasses *WebNLG English dataset* and *WordNet-WN18* [92], [126]. Furthermore, since text has a pivotal role in providing training and evaluation labels for multimedia transmission, some cross-modal datasets are applied in SemCom as well. For instance, LJSpeech belongs to a popular text transcription dataset, and is particularly suitable for testifying voice speech recognition and synthesis results [119]. Meanwhile, Librispeech, which is derived from read audiobooks and provides a corpus of approximately 1,000 hours of read English speech, has been leveraged to examine the efficacy of speech transmission. On the other hand, the availability of image datasets like CIFAR-10 and MNIST greatly promotes the development of AI and facilitates the verification of semantic image transmission [71], [90], [140], [141], [142], [143]. Furthermore, some researchers, who apply

the semantics from 2D discrete images to enhance video transmission accuracy, leverage UCF101 and Vimeo-90k video datasets to investigate the performance [146], [147]. Finally, SL-SemCom commonly experiments on Python-based toolkits like PyTorch or TensorFlow.

*2) EL-SemCom Datasets and Toolkits:* EL-SemCom apparently enjoys the applicability of aforementioned text and image datasets and toolkits [138], [151], [152]. For instance, on top of the CLEVR dataset [138], [152] designs a visual question answering (VQA) task to investigate visual reasoning abilities in response to a text-based question. In addition, EL-SemCom leverages some well-designed UAV or Internet of things (IoT) experimental toolkits as well, and typical examples include the OpenAI Gym toolkit and the IC3Net toolkit. Specifically, the former toolkit, which provides a standard API to communicate between learning algorithms and environments, as well as a standard set of environments (e.g., CartPole and Acrobot) compliant with that API, is becoming the de-facto scenario for evaluating RL algorithms [150]. Meanwhile, the latter toolkit includes traffic-junction and predator-prey environments [37], and benefits the study on learning when to communicate at scale in multi-agent cooperative and competitive tasks. Besides, some propriety datasets on UAVs and IoTs start to emerge, and the drone detection dataset [149] can be categorized into this scope. In particular, it consists of 7 different types of UAVs with three flight modes (i.e., switched on, hovering, and flying) captured.

To sum up, there is no doubt that datasets play an essential role in boosting the development of AI and DL, and act as an important pillar to support the validation of techniques in SemCom. Additionally, toolkits significantly facilitate the fostering and reproduction of interesting ideas. Together with semantic similarity metrics, the datasets and toolkits lay the foundation of SemCom.

## V. TOWARD SEMANTIC LEVEL SEMCOM

The design of models in SL-SemCom is mainly explored from two aspects (i.e., content semantics and channel semantics). For the sake of clarity, we will begin with some preliminary works, i.e., DL-based end-to-end communications, which treat the entire communication system as an end-to-end reconstruction task [153]. Furthermore, most of the existing SL-SemCom frameworks explore and extract the semantic information in an implicit reasoning process by a pre-trained model within a given KB, while some schemes explicitly reason part of hidden semantics that cannot be directly observed in the data by introducing the reasoning mechanisms, such as KG-based inference rules and probability theory-based semantic reasoning rules. As such, we will then introduce these frameworks from the perspectives of implicit reasoning and explicit reasoning. Finally, we provide an overview of channel semantics inference and exploitation, while optimization techniques around the channel semantics are also discussed.

### A. End-to-End Communications

As stated in Section II-B, conventional communication systems are divided into multiple independent blocks, each of
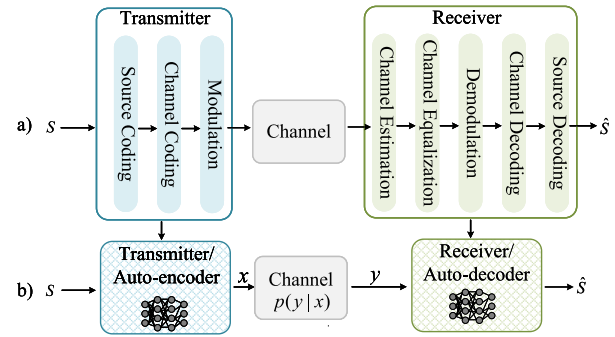


Fig. 9.    a) The conventional communication system; b) End-to-end communication system [50].

which is independently optimized. Despite its simplicity of implementation in engineering, such a design is known to be sub-optimal [161]. Benefiting from the rapid development of DL, data-driven methods pave the way to completely interpret the communication system as an end-to-end one, and promise performance improvement in complex communications scenarios that are conventionally difficult to describe with tractable mathematical models.

End-to-end communication systems with a given CSI usually assume the AWGN channel model. On top of such an assumption, O'Shea and Hoydis [50] first propose to use the autoencoder to implement an end-to-end communication system by replacing the transmitter and receiver with DNNs. Fig. 9 depicts the comparison of conventional and end-to-end communication systems [50], wherein the transmitter and receiver can be implemented by DNNs (i.e., the auto-encoder and auto-decoder respectively). Specifically, the DNNs can be pre-trained from both data and expert knowledge (e.g., the KB) by optimizing the end-to-end loss function (e.g., CE and MSE) in a supervised manner. Afterwards, the auto-encoder-based transmitter removes the redundancy of the source information and converts them into encoded symbols, which are further transmitted through the channel. Based on the received symbols with channel noise, the auto-decoder-based receiver attempts to accurately recover the information. The result [50] shows that in terms of the block error rate (BLER), i.e., $P(s \neq \hat{s})$, the end-to-end system can achieve competitive performance in comparison with the conventional communication system employing binary phase-shift keying (BPSK) modulation and a Hamming (7, 4) code, and also outperforms the combination of a Hamming (7, 4) code with either binary hard-decision decoding or maximum likelihood decoding.

Based on the pioneering work of [50], some variants [51], [52] are proposed to extend and optimize the end-to-end communication system. For example, the authors in [51] focus on communications under Rayleigh fading channels and extend the use of autoencoders from SISO systems to MIMO systems by designing multi-antenna-oriented DNNs. Erpek et al. [52] study the physical layer scheme for MIMO systems in the presence of annoying interference, and build two autoencoder-based communication systems, which share the same channel and can be jointly optimized to minimize their symbol error rate (SER). These works also verify that based on auto-encoders, the end-to-end systems can obtain

TABLE IV
A SUMMARY OF TECHNIQUES AND APPLICATIONS FOR SEMCOM

| Communication Level | Techniques | Applications | Contributions | Objective Function | Semantic Similarity Metrics | References |
|---|---|---|---|---|---|---|
| SL-SemCom | DL, BLSTM, JSCC | Text transmission | First implementation of the JSCC for text; Lower WER; Preservation of semantic information. | \ | WER | [5]□ |
| | DL, Transformer, DTL | Text transmission | Optimization of the JSCC at the semantic level; Robustness to the varying channel. | CE and MI | BLEU, BERT-SIM | [53]□ |
| | DL, Transformer, NN compression | IoTs | A lite DeepSC; Robustness especially under low SNR. | CE | MSE, BLEU | [154]□ |
| | DL, Universal Transformer, JSCC | Text transmission | Enhanced flexibility for semantic coding; Better adaptation to varying channels. | CE | SER, BLEU | [72]□ |
| | DL, Transformer, JSCC, HARQ | Text transmission | Robustness with adaptive bit rate control; Lower communication cost. | CE | Average bits for each word, BLEU | [74]□ |
| | DL, Transformer, JSCC | Text transmission | Effective SemCom | CE, KL | BLEU | [105]□ |
| | DRL, JSCC | Text and image transmission | Suitability for non-differentiable metrics and channels. | CE, rewards | WER, BLEU, CIDEr | [54]□ |
| | DRL, JSNC | Text transmission | Semantic confidence based semantic distillation mechanism. | MSE, CE and rewards | WER, BLEU | [4]□ |
| | DRL, GAN, VGG | Semantic segmentation and object detection | A remarkable reduction of data volume with high semantic similarity; Naturalness. | Rate-semantic-perceptual | FID, KID, PSNR, SSIM, mIoU, Semantic loss | [155]□ |
| | DL, JSCC, CNN | Image transmission | Initial introduction of JSCC to image transmission. | MSE | PSNR | [71]□ |
| | DL, JSCC, CNN, GDN/IGDN [156] | Image transmission | Practical implementation of JSCC to fully exploit channel feedback; Bandwidth saving. | MSE | PSNR, Average bandwidth ratio | [140]□ |
| | DL, CNN, ResNet-152, LSTM, JSCC | Image transmission | Multi-level semantic extraction. | MSE | PSNR, SSIM | [144]□ |
| | DL, CNN, JSCC | Image transmission, Digit recognition, Image classification | Adaptive to the observable datasets with high performance. | MSE, CE, KL divergence | PSNR, Accuracy | [142]□ |
| | DL, CNN, ViT, Residual block | Image transmission | Real-time SemCom. | MSE | SSIM | [141]□ |
| | DL, CNN, Attention mechanism, ResNet | Telephone systems and multimedia transmission systems | Higher accuracy of signal recovery with more weights given to essential speech information. | MSE | MSE, SDR, PESQ | [120]□ |
| | FL, CNN | Audio | Application of FL to the edge devices and server to improve audio semantic accuracy. | Normalized root mean squared error (NRMSE) | MSE | [157]□ |
| | DL, CNN, GRUs | Speech recognition and synthesis | Reduction of the volume of transmitted data without performance degradation. | The connectionist temporal classification (CTC) loss | CER, WER, FDSD, KDSD | [119]□ |
| | DL, LSTM, VGG, Transformer, GAN | Speech-to-text and speech-to-speech transmission | Speech signal recovery by combining text-related features and additional speech-related features with a pre-trained GAN. | CTC loss, CE, MSE | Bert-SIM, WER | [139]□ |

*(Continued)*

significantly superior performance than conventional modulation and coding schemes, which lay the foundation for SemCom. Notably, despite the performance improvement, this end-to-end system is still dedicated to accurately recovering data at the bit level, and there is no significant improvement in data compression and semantic transmission.

TABLE IV
(*Continued*) A SUMMARY OF TECHNIQUES AND APPLICATIONS FOR SEMCOM

| Communication Level | Techniques | Applications | Contributions | Objective Function | Semantic Similarity Metrics | References |
|---|---|---|---|---|---|---|
| SL-SemCom | DL, Transformer, HARQ | Text transmission | Significant reduction in the amount of transmitted data; Decreased sentence error rate. | MSE, CE | BLEU, WER, average bits | [91]□ |
| | DL, CNN, Attention mechanism, IB, DIB | IoT, Image classification, Multi-view object recognition | A Flexible control of the communication overhead. | \ | Accuracy, Relevance | [158]□ |
| | DL, JSCC, GDN, IGDN, CNN | Image transmission | A channel-wise soft attention mechanism to scaling features according to SNR | MSE | PSNR, Storage | [143]□ |
| | DL, ViT, Codebook | Image classification | Higher robustness to semantic noise. | CE | Classification accuracy | [90]□ |
| | DRL, KG | Text transmission | A graphical representation of semantic meaning. | Semantic distance-based rewards | Loss values, Accuracy | [73]■ |
| | DL, KG | Text transmission | Utilization of the life-long learning to automatically update the reasoning rules. | Semantic distance-based rewards | Error rate, Loss values, Accuracy | [92]■ |
| | KG, DL | Text transmission | A simple and general solution with higher compression ratio and reliability of communications. | \ | Number of bits, Bert-SIM, BLEU | [126]■ |
| | KB, ProbLog | Message choice problem | A probabilistic logic approach to improve the KB. | Entropy | \ | [20]■ |
| EL-SemCom | DL, JSCC, ResNet-50 | Image retrieval, Edge devices | Robust JSCC retrieval-oriented image compression. | CE | Accuracy, CE loss | [145]□ |
| | DL, Transformer | VQA, Image retrieval, Machine translation | Multi-user SemCom systems for transmitting both single-modal data and multi-modal data. | CE, MSE | BLEU, answer accuracy, recall@1 | [67]□ |
| | DL, Transformer, Codebook | VQA, sentiment analysis, Image classification | A unified multi-task SemCom for multimodal data. | MSE | BLEU, Accuracy, PSNR | [159]□ |
| | DRL, Curriculum learning | IoTs, Control tasks or automotive production in factories | Lower task execution time; Reduced transmission cost. | Transmission cost and execution time | Transmission cost and task execution time | [81]□ |
| | DRL, KG, VD based DQN | IoTs, traffic control | A significant reduction of transmission delay and energy consumption. | Rewards | Energy, Similarities | [76]■ |
| | DRL, KG | IoTs, heterogeneous agents | Semantics-aware collaborative DRL for wireless networks. | Rewards | Average return, Maximum average return | [150]□ |
| | DRL, IB, GIB, GCN | IoTs, Predator-prey game | Semantics-empowered multi-agent cooperation to relive the IB. | Rewards and IB | Success rate, entropy, reward | [37]□ |
| | DL | IoTs, edge devices | A trading mechanism with higher utilities for semantic models. | Rewards | Average utility of winning sellers, BLEU, Bert-SIM | [151]□ |
| | DRL, QMIX [160], RNN, Attention mechanism | UAVs | Lower latency; Inter-UAV collision avoidance. | Temporal difference(TD) error | Latency, Error rate | [148]□ |
| | DL, ResNet-18, SVM, Few-shot learning | UAVs | A higher recognition accuracy than other few-shot learning schemes. | CE and Euclidean distance | Accuracy | [149]□ |

*Notations: The superscript ■ indicates the learning process is "explicit"; while the superscript □ denotes an "implicit" learning process.*

## B. Implicit Reasoning in SL-SemCom

*1) DL-Based Semantic Transmission:* In some sense, the aforementioned DL-based end-to-end communications lay the basic foundation for the development of SemCom, as DL proves itself as an effective tool to improve the reliability of data transmission. Therefore, by shifting from the bit-level accuracy to a semantic-level one, the DL-based semantic transmission also promises a new communication paradigm. In other words, DL can be leveraged in SemCom to extract and reproduce the semantics of information, so as to implicitly understand the semantics rather than decode the received bits. Therefore, basically, the joint optimization of the DL-based

encoder and decoder could achieve the end-to-end semantic delivery and reconstruction of the source information. This section will present the latest work in this aspect.

*a) Text processing:* The rapid advance of natural language processing (NLP) which lays the very foundation for understanding the semantics behind the text, inspires researchers to redesign the transceiver to achieve SemCom. With the aid of a KB and some preliminary information from the CSI, SemCom primarily leverages a semantic encoder to understand and extract feature information from the source information, while the semantic decoder attempts to reconstruct the received semantic symbols into the semantic information as consistent with that from the transmitter as possible. For instance, the words "three" and "3" share the same meaning at the semantic level, but they are completely different in terms of the encoded bit sequences. Meanwhile, channel coding is also complemented to combat channel impairment. Since the semantic encoder and decoder are commonly implemented in DNNs, an essential part of content-related SemCom becomes to design an appropriate loss function to minimize the semantic error and distortion caused by semantic noise, so as to achieve end-to-end optimization.

LSTM enabled JSCC for text transmission [5], i.e., DeepNN, belongs to one of the pioneering works, which is capable of preserving semantic information by embedding semantically similar sentences located in a closer semantic space, and recovering the transmitted sentence in terms of the WER rather than the bit error rate (BER). Compared with the traditional end-to-end optimization, [5] shows the great potential of JSCC-based SemCom, especially in scenarios with sentences encoded by a small number of bits. Although the concept of SemCom is not mentioned in [5], the idea of semantic extraction and preservation has been put forward and greatly inspires later researchers.

In order to measure the semantic similarity at the sentence level, Xie et al. [53] propose a DL-based SemCom system (i.e., DeepSC), the framework of which is similar to the semantic level of Fig. 3. Similar to end-to-end communications, the physical-layer blocks in the conventional communication system are merged together. The DNN structure of DeepSC consists of a transformer [26]-based source encoder to extract the embedding information and dense layers to generate semantic symbols, thus facilitating subsequent transmission. As for the channel environment, the AWGN channel is interpreted as one layer of the pre-trained model. Likewise, the receiver is composed of a channel decoder for symbol detection and a transformer decoder for text estimation. It should be noted that a loss function $L_{\text{total}} = \text{CE}(s, \hat{s}) - \gamma I(x, y)$ is first developed in [53] to train the whole process, where the first term takes the CE loss between $s$ and $\hat{s}$ for minimizing semantic differences. Besides, $\gamma$ is a parameter to balance the relative importance of CE and MI. As depicted in Fig. 10, the training framework for DeepSC by stochastic gradient descent (SGD) consists of two phases due to the introduction of the two-term loss function for jointly training the whole DNN. Particularly, Phase I trains the estimation model for MI, while Phase II trains the whole DNN in terms of CE and MI. Moreover, DTL can also be adopted to better adapt to different
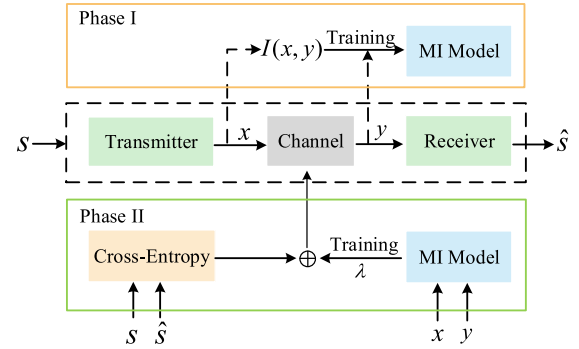


Fig. 10. The training framework of the DeepSC [53].

channel environments. By carefully calibrating the layers of semantic encoders and channels, DeepSC can effectively extract semantic information and ensure reliable transmission. Finally, the simulation results [53] show that DeepSC outperforms other conventional communication systems, especially for low signal-to-noise ratio (SNR) scenarios. The BLEU score of DeepSC also converges to about 90%, which implies that DeepSC can capture most of the relationship of words and features of the syntax to facilitate the understanding of the texts.

Furthermore, on the basis of DeepSC, some variants are developed. For example, in [154], a lite version "L-DeepSC" is studied, in which the authors consider an affordable structure for context transmission in IoT devices by pruning the model redundancy and decreasing the resolution of DNN weights. In order to capture the effects of end-to-end semantic distortion and minimize the errors during reconstructing the sequences, Sana and Strinati [105] propose a novel loss function, which incorporates the MI and Kullback-Leibler divergence and is defined as $L_{\text{total}} = I(x, s) - (1 + \alpha)I(x, y) + \beta \text{KL}(y, \hat{s})$. Besides, $\alpha$ and $\beta$ are the parameters to reflect the relative importance. The mutual information $I(x, s)$ and $I(x, y)$ is leveraged to ensure compactness and informativeness, respectively. Meanwhile, $\text{KL}(s, \hat{s})$ minimizes the semantic distortion between the intended $s$ and the decoded $\hat{s}$. Another contribution in [105] lies in a semantics-adaptive mechanism, which can dynamically adjust the number of symbols per word to better balance the tradeoff between accuracy and complexity. By experiments, for the same accuracy, the adaptive method in [105] could dynamically use fewer symbols for each word, thus outperforming the fixed one.

Many existing methods [53], [154] utilize the transformer-based encoder to extract semantic features from the source. However, these methods utilize a fixed transformer and ignore the impact of the difference in length and semantic complexity for each sentence. Therefore, it becomes essential to design a flexible architecture for the semantic encoder to better adapt to different inputs and channel variations. Zhou et al. [72] adopt a universal transformer (UT) [162]-based encoder to dynamically process sentences, wherein UT introduces a circulation mechanism to adjust the DNN structure of the transformer and an adaptive computation time (ACT) [163] to adaptively allocate the number of computational resources to process each symbol. In this way, the semantic system

can be more flexible to deal with different input and channel conditions.

Similarly, considering complicated content and channel conditions, the encoder side in [74] adopts an adaptive bit rate control mechanism for the transmission of semantic information. Specifically, this bit rate control mechanism consists of two parts, that is, an encoder capable of transmitting sentences at different rates and a policy network to choose the suitable rate for the encoder. In [91], the transformer-based semantic coding adjusts the code length according to the sentence length, so as to enable semantic coding to outperform other coding methods with fixed length and reduce SER under different channel conditions.

Most of the aforementioned SemCom systems directly use semantic error as the performance metrics but ignore the semantic loss in the process of signal detection. Hence, in order to compensate the semantic loss, Guo and Wang [164] propose a signal shaping method for SemCom system, which measures the semantic loss during detecting signals by the BERT model [28]. Afterwards, an efficient projected gradient descent method is adopted to solve the signal optimization problem subject to the power constraint.

*b) Image processing:* Nowadays, the image and video data accounts for 75% of IP traffic and is more informative than the text [165]. Hence, it becomes imperative to build an image SemCom (ISC) system that can significantly reduce the amount of data without sacrificing the fidelity of the image. In this part, we focus on the DL-based image compression and transmission for ISC.

*DL-based image compression*: The conventional image compression methods such as JPEG [166] and JPEG2000 [167] rely on its quantization and entropy coder to compress images [168]. However, these methods might be inflexible to encode all types of image content and image formats. By contrast, the DL-based end-to-end compression methods [168], [169], [170] exhibit an excellent learning ability wherein their deep autoencoder encodes the images into low dimensional vectors to realize efficient compression. Nevertheless, the legacy image compression approaches mostly focus on pixel-level consistency but neglect the signal distortion caused by the channel transmission on downstream tasks. The latest image compression technology transforms traditional pixel-level reconstruction into a DL-based semantic-level one, which becomes capable to maintain semantic consistency and stably guarantee the performance for downstream tasks.

Many DL solutions have been proposed for image compression. Dependent on the type of utilized DNNs, those solutions can be mainly classified as CNN-based [169], [171], RNN-based [172], [173], and GAN-based [174], [175], [176].

- *CNN-based methods*: CNNs have prime importance in feature extraction and image compression. In [169], the authors present a CNN-based method, in which semantically-salient regions are highlighted by an intentional map and encoded at a higher quality than those background regions. The authors in [171] propose a CNN-based end-to-end multi-spectral image compression method for feature partitioned extraction, which consists

of two parallel parts (i.e., one for spectral features and the other for spatial features). Then, a rate-distortion optimizer is adopted to make the representation data more compact, and yields superior performance in terms of PSNR than JPEG2000 [171].

- *RNN-based methods*: Commonly, the RNN-based methods [172], [173] also involve some convolutional layers. For instance, Toderici et al. [172] present a set of full-resolution lossy image compression methods consisting of an RNN-based encoder and decoder, a binarizer, and LSTM modules for entropy coding. The RNN encoder encodes the residuals between the previous reconstruction and the uncompressed layer, and combines the extracted key features with the output of the previously hidden layer at each recurrent step. The performance of typical RNNs (e.g., LSTM, associative LSTM, a new hybrid DNN of gate recurrent unit (GRU) and residual network (ResNet)) has been validated in [172] and indicates fine image compression quality.

- *GAN-based methods*: GAN encompasses one generator and one discriminator, the game of which contributes to generating new, synthetic data samples that seem authentic and exhibit many realistic characteristics [177]. Rippel and Bourdev [174] introduce GAN to realize a real-time adaptive compression. By adopting the pyramidal decomposition and regularization for feature extraction and processing with adaptive codelength, the codec in [174] efficiently produces files 2.5 times smaller than JPEG and JPEG 2000. Agustsson et al. [175] also present a full-resolution image compression framework based on GAN operating under low bit rates, and successfully build an extremely-effective image compression system by constraining the application domain to street images. In particular, on the basis of available semantic labels, this system can fully synthesize unimportant regions in decoded images with preserved ingredients such as streets and trees.

In addition, researchers also resort to signal processing methods to boost the semantic preserving performance of DNNs [178], [179]. In summary, all these efforts increase the interpretability of image semantics, by learning the natural distribution of images instead of merely local information construction.

*SemCom for image transmission*: Classical methods such as JSCC for image transmission aim to directly map pixel-level values into complex-valued symbols, and jointly learn the encoder and decoder with feedback from the channel. Fig. 11 depicts a general framework for semantic image transmission [110], [111], wherein the DL-based semantic encoder first extracts original low-dimensional information as a form of semantic representations (e.g., semantic labels, graph, and embedding) by a joint source-channel optimization. Based on the received semantic signals, the semantic decoder recovers the images according to some semantic similarity metrics such as PSNR and SSIM. Assuming the consistency between the data distribution at the transmitter and the shared KB, Bourtsoulatze et al. [71] propose a JSCC technique for wireless image transmission (i.e., deep JSCC)
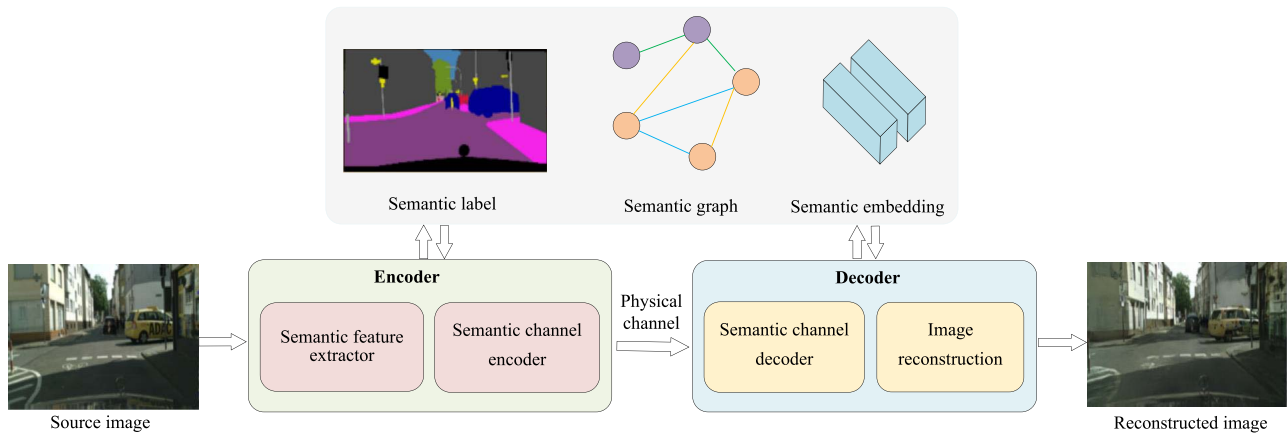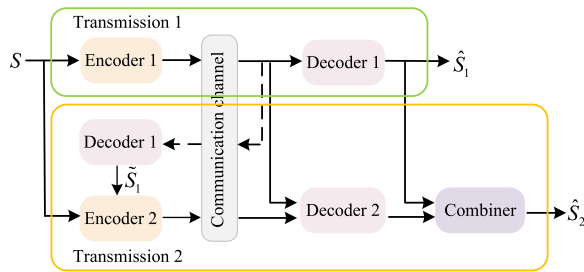
Fig. 11.   A general SemCom framework for image transmission.



Fig. 12.   Architecture of DeepJSCC-f for the $L = 2$ case [140].

by modeling the encoder and decoder as two CNNs and incorporating the channel in the DNNs as a non-trainable layer, which can significantly outperform the separation-based transmission scheme under both AWGN and Rayleigh fading channels. Kurka and Gündüz [140] introduce an autoencoder-based JSCC with flexibly varying-length coding scheme by utilizing channel feedback, called *DeepJSCC-f*, and adopt CNN-based layered autoencoders encompassing an encoder, a decoder and a combiner. The transmission of each image $s$ is divided into $L$ layers, in which each layer tries to improve the quality of recovered images by additional information of the residual error from the previous layer. Especially, as depicted in Fig. 12, for a 2-layer DeepJSCC-f, the estimation $\tilde{s}_1$ in the previous Layer 1 works as feedback to Layer 2 for obtaining the final recovered image $\hat{s}_2$, so as to reduce construction errors. Furthermore, Kurka and Gündüz re-design a DeepJSCC-l framework [180] for adaptive transmission, in which images are transmitted over multiple channels, allowing a flexible and bandwidth-adaptive transmission. By simulations, DeepJSCC-l leads to superior performance over state-of-the-art progressive transmission schemes in challenging scenarios with low SNRs and limited bandwidth.

Furthermore, Zhang et al. [144] propose a multi-level semantic feature extractor for image transmission, named MLSC-image, which consists of three feature extraction modules (i.e., a semantic feature module composed of ResNet-152 and LSTM [23] for text-form semantic features, a pre-trained SPNet-based module [181] for high-level segmentation feature, as well as a "Concat" module for low-level features

such as local details of images). These extracted features are further aggregated and encoded by the joint source and channel encoder into semantic symbols. In addition, the adaptive rate control can also be used for ISC. In that regard, Yang and Kim [182] propose an SNR-adaptive module and a policy DNN to dynamically adjust the number of active intermediate features over various SNRs. Meanwhile, the authors demonstrate that this scheme can maintain a comparable visual quality compared to a specially trained model for a particular rate.

In order to address the issue of distribution discrepancy between the actually observed data by the transmitter and the shared KB, the authors in [142] develop a semantic coding DNN to learn how to extract and transmit the semantic information using a receiver-leading training process. In particular, on the basis of a data adaptation (DA) DNN, the semantic coding DNN can directly leverage the converted quasi-empirical distribution from the observed data rather than complete retraining.

Instead of sticking to pixel-based metrics (e.g., PSNR and SSIM), Huang et al. present a GAN-based coarse-to-fine ISC for multimedia [118] by adopting a base layer and an enhancement layer. Specifically, the base layer of an image is first generated to reconstruct the semantic information while the residuals coded by better portable graphics (BPG) [183] further refine details as the enhancement layer. The proposed model can output images visually more similar to the source and significantly outperforms the baselines (i.e., BPG, JPEG2000) in terms of perception metrics such as FID and KID.

There emerge exciting prototype results for ISC as well. For instance, in a field-programmable gate array (FPGA) prototype, Yoo et al. [141] demonstrate the feasibility of ISC in a real-time wireless channel with promising results compared to the traditional 256-QAM in actual wireless environments, especially at low SNRs. Meanwhile, Zhang et al. [184] propose a model division multiple access (MDMA) scheme to capably excavate the shared and personalized information for different users. On this basis, shared semantic information is transmitted within the same time-frequency resources, while personalized semantic information is delivered separately. Experimental

results show a 5-dB improvement over the conventional non-orthogonal multiple access (NOMA) scheme [185] under low SNRs.

*c) Speech processing:* Unlike text that only contains characters, the speech signal is more complex and difficult to be semantically understood [19], since the speech content not only involves signal values and frequency, but also its loudness and tone. For example, the text "come on" can express a kind of encouragement for someone. Meanwhile, it can also be a modest and negative emotion. In this case, the speech possibly produces more confusing meanings. Besides, dialect also becomes an obstacle to speech recognize & transmission. Therefore, semantic speech processing puts more emphasis on technical enhancement to semantic speech transmission, and also sheds light on cross-modal transmission involving both text and speech.

In order to enhance the recovery accuracy at the semantic level, Weng and Qin [120] propose a DL-enabled SemCom system for speech, named DeepSC-S. Both the speech encoder and speech decoder in DeepSC-S are based on an attention mechanism employing squeeze-and-excitation (SE) modules, named SE-ResNet, in which one or multiple SE-ResNet modules are sequentially connected to learn attention weights of features by capturing inter-dependencies. Reference [120] shows that in terms of SDR and PESQ, DeepSC-S outperforms the conventional communication systems, especially at the low SNRs. Moreover, FL can also be utilized in designing encoder/decoder [186], so as to mitigate the problem of privacy issues to share training data among devices. For example, Tong et al. [157] develop a FL trained model for audio SemCom to reduce overhead between edge devices and the server and improve the accuracy of semantic information extraction. Specifically, in order to collaboratively train the autoencoder over multiple devices and the server, a wave-to-vector framework is devised to extract the semantic information of audio signal and update CNN models locally. Afterwards, the local parameter updates are transmitted to the server at each time step. Finally, the server aggregates the collected local updates into a global model and broadcasts it to all participated devices. With the help of FL, this SemCom framework can not only improve the accuracy of semantic transmission, but also further alleviate the privacy issues associated with the network. Shi et al. [2] design a first-understand-then-optimization methodology for semantic speech transmission. Besides the typical modules for SemCom, an additional symbol recognition module is added to further reduce errors, so as to largely alleviate the data traffic burden and improve semantic fidelity.

DeepSC-ST [119] jointly optimizes the tasks of speech recognition and speech synthesis, on top of text transcription and speech synthesis modules. As shown in Table V, in terms of both speech recognition metrics (e.g., CER and WER) and speech synthesis metrics (e.g., FDSD and KDSD), DeepSC-ST [119] yields superior performance than conventional speech transceivers and text transceivers in a real-time human speech input prototype, and also significantly reduces the amount of transmitted data, since it can comprise useful characters in the long speech samples.

TABLE V
RECOGNIZED SENTENCES IN DIFFERENT SYSTEMS OVER
RAYLEIGH CHANNEL (FOR SNR = 4 dB) [119]

| | |
|---|---|
| Original Sentence | "he concluded that school had nothing to offer him" |
| DeepSC-ST [119] | "he concluded that school had noghing to offer him" |
| Speech Transceiver | "it cood sili ite bebou a pims t lup ar of ig mote terigytit w" |
| Text Transceiver | "h ea aahesourhhtntchchen ehaoeitdcofo offer him" |

Similarly, Han et al. [139] propose a semantics-oriented communication system for speech-to-text transmission and speech-to-speech transmission, which exploits an attention-based soft module and a redundancy removal module to extract semantics-related features. In particular, a language model-based semantic corrector computes the most matching text transcription while a connectionist temporal classification (CTC)-based speech information extractor obtains the additional semantics-irrelevant but speech-related information like duration, power and pitch.

*2) DRL-Based Semantic Transmission:* Despite the aforementioned works can successfully extract essential features of data at the semantic level, there also exist some problems that need to be addressed. First and foremost, most of the popular works commonly focus on the DNN design of autoencoder for semantic extraction while using bit-level as objective functions, such as semantics-blind CE or MSE, which introduces an extra "semantic gap". On the other hand, most works have now studied the optimization of differentiable objectives, there is little light towards a universal framework that allows the optimization of both differentiable and non-differentiable objectives, which is instead commonly seen in real wireless scenarios. As DRL is an efficient and stable learning for any user-defined rewards and capable to address the non-differentiability issue, it is attracting increasing interest in SemCom.

To the best of our knowledge, the authors in [54] first introduce the DRL to the SemCom system to preserve the semantic-level accuracy instead of bit-level optimization, called SemanticRL. First, a semantic similarity-oriented JSCC solution (i.e., SemanticRL-JSCC) is designed by calibrating the semantic-level metrics like BLEU and CIDEr, so as to close the "semantic gap". Afterwards, the learning process is formulated as an MDP and a self-critic algorithm is leveraged to solve the non-differentiability issue, wherein the *state* is defined as the recurrent state of decoder combined with historical actions taken so far, and the *action* is the decoding of words from the dictionary. Besides, the *policy* is the probabilistic preference to take an action based on the current state, and the *reward* is the similarity metrics between input sequence $s$ and output sequence $\hat{s}$. In this way, SemanticRL models the output likelihood as a probabilistic multinomial distribution and continually optimizes the semantic similarity by maximizing the total rewards. Moreover, in order to cope with non-differentiable channels, SemanticRL further regards the encoder and decoder as two independent agents

and introduces the DRL into the receiver side. Therefore, in the SemanticRL-based JSCC, the whole learning system turns into a collaborative semantic transceiver. Compared with the CE baseline, SemanticRL yields competitive results in terms of BLEU, BERT-SIM, and CIDEr scores. Meanwhile, as SemanticRL performs better in BLEU (2-gram), BLEU (3-gram) and BLEU (4-gram), but acts slightly poorly in BLEU (1-gram), it strongly proves that it pays more attention to the completeness of context and has the ability to preserve semantic information. As for noisy channels, SemanticRL also exhibits strong robustness under varying AWGN and fading channels with SNR randomly fluctuating from 0 to 20 dB, especially in low SNRs from 2.5 dB to 7.5 dB. The authors [54] also extend this framework to other SemCom tasks like image transmission, which also performs well on the MNIST dataset.

In order to further tackle the impact of time-varying channels on semantic transmission, [4] also develops a joint DRL-based semantics-noise coding (JSNC) mechanism to adapt to the varying channel and preserve semantics. Specifically, [4] adopts a confidence-based distillation mechanism at both the encoder and the decoder, which automatically adjusts the depth of semantic representations according to the underlying channel state and sentence structure. As such, only when the semantic confidence reaches a pre-defined threshold, the encoder and decoder could leverage the extracted semantic information for further processing. The simulation results [4] prove that JSNC outperforms the baselines without a distillation mechanism, especially under low SNRs, and the distillation time also increases along with the sentence's length.

Moreover, Huang et al. [155] propose an RL-based adaptive semantic coding (RL-ASC) approach that encodes images beyond the pixel level, by optimizing a triple factor consisting of semantic loss (i.e., mean intersection over union, mIoU), perceptual loss and transmitted bit rate. In particular, the RL-ASC encompasses a semantic encoder, an RL-based semantic bit allocation for adaptive quantization, and a GAN-based semantic decoder. Specifically, the RL-based semantic bit allocation module assigns different quantization levels for different semantic concepts with varying degrees of importance in downstream tasks. In other words, the crucial semantic concepts should be encoded with higher precision so as to incur as little loss as possible, while the task-irrelevant concepts can be relaxed. Finally, the GAN-based semantic decoder can fuse both local and global semantic features via an attention-based module to reconstruct images. In this way, the data volume can be significantly reduced while maintaining a high semantic similarity.

Furthermore, since wireless resources are limited, the base station (BS) may not be able to disseminate the entire semantic information to all user equipment (UE) under stringent latency constraints. Therefore, Wang et al. [125] propose a semantics-driven wireless network in which the BS extracts the semantic information from the textual data by a KG [187] and transmits it to attached UEs, while the users recover the original text using a graph-to-text generation model [188]. In order to optimize the *metric of semantic similarity*, which

considers the semantic accuracy and semantic completeness, an attention proximal policy optimization (APPO) algorithm is leveraged by evaluating the importance of each triple in the semantic information. On top of that, RL tries to find effective policies for resource allocation and semantics transmission. The experiments show that this approach can reduce the required volume of data transmission by 41.3% meanwhile significantly boosting higher semantic similarity than conventional communications.

In summary, benefiting from some advanced NLP and computer vision (CV) models to extract feature embeddings and model the context, the DRL-based works successfully improve the effectiveness of SemCom. Besides, some DRL-based works start to address semantic and channel non-differentiable issues.

### C. Explicit Reasoning in SL-SemCom

The aforementioned SemCom frameworks implicitly learn from the context based on pre-trained models, and may lack inference and error correction functionalities. However, it is unwise to preclude the context, which usually includes rich information that is hard to be represented and recovered at first glance. As such, explicit reasoning can contribute to the semantic transmission by inferring incomplete information using inference rules at the receiver, and thus effectively guide the task execution. In terms of the underlying methodology, explicit reasoning can be classified as KG-based, information theory-based, and probability theory-based.

*1) Knowledge Graph-Based Semantic Reasoning:* In order to find an appropriate representation of semantic meaning, a graph-based structure emerges as an intuitive way. Consistent with common settings in a graph, the KG consists of some nodes and undirected edges as well, wherein the node and edge can be formally termed as the entity and relationship, respectively. Moreover, the relationship connects distinguishable and independent entities together, and the basic unit of KG is an entity-relationship-entity triplet. Interesting readers could refer to [187], [189], [190] for more aspects (e.g., knowledge graph completion) of the KG as well.

Typically, the KBs derived from real-world facts and grammar can be built in the form of a KG and represented as well-design DNNs. For example, WordNet [93], which is one of the most popular lexical KBs, consists of 118,000 English words and the corresponding relationships. The entities and relationships in a specific KB facilitate the semantic reasoning of implicit meanings. Taking the example of a sentence "*I want to buy an apple*", the entity "*apple*" may link to some hidden entities such as "a smartphone brand" or "a kind of fruit". If the context has explained that "*I want to buy a phone*", the meaning of "apple" can be referred to as the former case. In SemCom, the semantic KBs, which possibly contain some private knowledge for specific individuals contingent on the particular environment or backgrounds, can be shared and updated among transmitters and receivers. Therefore, the receiver in SemCom can better infer the meaning of received messages based on the semantic KBs and other learned inference rules.

Recently, several works [73], [92], [126] leverage KG to extract the essential information and realize the semantic reasoning for SemCom. For example, [73] proposes an implicit SemCom to optimize the semantic representation and delivery of messages. To this end, the source signals are first mapped into a semantic KG by an entity detector, and a commonly used graph embedding solution (i.e., TransE) [191] is adopted to obtain the semantic embeddings of entities and relations. Afterwards, a generative imitation learning-based reasoning mechanism is employed at the decoder side to learn and minimize the semantic discrepancy between the possible reasoning path and the expert path. Therefore, the decoder is capable of generating a reasoning path as similar as possible to the expert one. Liang et al. [92] also utilize a KG-based approach to infer the implicit information in incomplete entities that cannot be directly observed in the received message. In particular, a life-long learning-based updating process is taken into account to automatically update the reasoning rules, so as to improve the interpretation accuracy. Due to the over-simplicity of TransE, more thoughtful graph embedding models, such as TransG [192] and TranSparse [193], are leveraged as well.

Different from mapping triples to embeddings [73], [92], Zhou et al. [126] propose a cognitive SemCom system, in which the "cognitive" characteristic is enabled by the KG. Specifically, the triples are regarded as semantic symbols while the entities or relationships are uniquely mapped as integers. Reference [126] demonstrates that this cognitive SemCom framework is capable of capturing and recovering semantic information with reduced errors.

*2) Information Theory and Probability Theory-Based Semantic Reasoning:* Information theory and probability theory are also adopted to implement semantic reasoning in communication systems, in which the entropy and probability are leveraged to express the uncertainty of logical clauses (e.g., rules or facts) in the KB. Choi et al. [20] provide a unified and technical approach for SemCom through using a practical ProbLog [106]. In ProLog, each logical clause is annotated with a probability (by a programmer) that indicates the degree of (the programmer's) belief in the clause. Specifically, the ProLog can learn from the facts to determine the relationship between objects, and then automatically match the target that needs to be queried. Besides, a SemCom layer is additionally introduced on the basis of the reliable communication layer to exchange logically meaningful clauses in the KBs, wherein both layers interact with each other to improve the efficiency of communications.

Furthermore, Seo et al. [38] introduce the probability distribution-based contextual reasoning to semantics-native communications, which can significantly reduce the bit-length of semantic representation with high reliability. In order to improve the interpretability of the data, Thomas and Saad [194] propose a neuro-symbolic AI for learning the causal structure behind the observed data. In particular, a symbolic component characterizes the high-level semantic representations while a neural network (NN) component is responsible for logical deduction and learning. Meanwhile, [194] shows that it yields significant gain in bandwidth saving for reliable data transmission compared to conventional communications.

In addition, different from most existing works that the encoder and decoder are jointly optimized to maximize the semantic information, Xiao et al. [88] consider a case where the transmitter and the receiver can have different distortion measures, and make rational decisions about their encoding and decoding strategies independently. On basis of this assumption, [88] studies the impact of strategic decision-making on semantic communications and focuses particularly on the case of a transmitter committing to an encoding strategy based on rate-distortion theory. In this regard, recalling the formulation in (10), a transmitter can have a personal understanding of semantic messages based on its indirect source observations $U$, and the receiver can recover the full semantic information based on its KB and the assistance of side information $E$. To figure out the impact of strategic communications by utilizing the side information $E$, they study three types of equilibrium solutions including optimal Stackelberg equilibrium (OSE) [195], robust Stackelberg equilibrium (RSE) [196] and Nash equilibrium (NE) [197]. Finally, though different decoding strategies may have different distortion rates, Xiao et al. [88] verify a sufficient condition, committing to which an encoding strategy can always improve distortion performance.

### D. Exploitation of Channel Semantics

Given the wide adoption of JSCC-oriented DNNs in SemCom, it is challenging to independently model the channel layer. Nevertheless, channel semantics, which refers to CSI and critical surrounding environment information, still makes significant senses in SemCom. In this part, we first discuss the inference and exploitation of channel semantics, so as to facilitate the design of end-to-end DNNs in SemCom. Afterwards, given the possible existence of semantic mismatch and ambiguity, which adds to the difficulty to circumvent the semantic noise, we provide some interesting optimization techniques for noise-robustness. Finally, we show how the shift from bit-level accuracy to semantic-level accuracy could motivate the development of semantics-aware transmission/retransmission techniques as one of the effective means to ensure transmission reliability.

*1) Channel Semantics Inference:* In a real communication system, it is often difficult to obtain the distribution of the real channel as the channel impairments such as channel noise and annoying channel-varying properties are usually hard to be modeled mathematically and expressed analytically. However, apart from the widely assumed wireless channels (e.g., Rayleigh fading channel, AWGN), DeepSC [53] and MLSC-image [144] still impose an assumption that the CSI must be accurately given before the training phase to minimize the reconstruction error, though the assumptions might not always hold and possibly lead to biased DNN weights. Hence, researchers shed some attention on learning channel semantics from varying channels, so as to make the SemCom capable of agnostically adapting to varying channel conditions, and the rapid progress in AI greatly accelerates such a process [198], [199].

*a) Inference from pilot-assisted information:* In order to learn diverse channel effects, Ye et al. propose to approximate
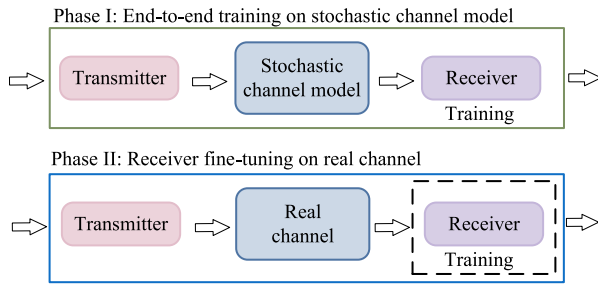
Fig. 13. Two-phase training strategy [204].

the distribution of channels (i.e., $p(Y|X)$) by using a conditional GAN (cGAN) from historical data [200]. In particular, [200] first adds some pilot information as conditional information for the agnostic channel, and tries to learn a surrogate channel from corresponding outputs [44], [201]. Afterwards, [200] takes advantage of an end-to-end learning method to jointly optimize the loss function. Regardless of the inaccuracy of expert knowledge about the channel, the simulation results [200] confirm the effectiveness of this method, and show competitive performance compared to those with known channel models under AWGN and Rayleigh channel configurations. On the other hand, inspired by the concept of model-agnostic meta-learning [202], [203] proposes to first send multiple pilot packets over varying channel conditions to learn a set of channel models, and then carry out online meta-training on any SGD-based SemCom frameworks with minimal modifications. Similarly, as shown in Fig. 13, [204] introduces a two-stage method [205], within which DNN-based transmitter and receiver are first trained on available datasets collected from a stochastic channel that resembles the behavior of the expected channel as closely as possible. Afterwards, these DNNs get partly fine-tuned under the real channel, so as to accelerate the training process [204]. Meanwhile, the authors extend this idea towards continuous data transmission entailing the synchronization issue by introducing a DNN-based frame synchronization module. However, due to the existence of differences between the practical channel and the stochastic channel for training, substantial efforts are still required before conducting practical over-the-air transmissions [204].

Additionally, some researchers [206], [207] propose to estimate the noisy channel by RL instead of supervised learning, since these RL-based systems cannot depend on any prior knowledge. Taking [206] as an example, the transmitter is considered as an agent while both the channel and the receiver can be regarded as the environment. Furthermore, the key idea turns to approximate the gradient of the loss function with respect to the DNN parameters of the transmitter by regarding the channel input as a random variable. Specifically, in order to circumvent the issue due to the missing channel gradient, the transformer adopts RL to gradually learn how to take actions to minimize the loss function provided by the environment. Besides, the training of the receiver is formulated as a supervised learning task, under the assumption that the receiver has the privilege to access the send messages (e.g., preambles). Hence, by using the policy gradient algorithm, the transmitter

can learn without requiring the gradients of the channel. However, some prior information about the channel is still required to achieve a competitive performance [200].

*b) Inference through environment perception:* Strong interests are emerging towards exploiting surrounding environment information for channel semantics inference, and preliminary results validate their effectiveness. Specifically, critical scatterers, which could significantly affect signal propagation in the wireless channel, can be located after filtering environment information redundancies, so as to save inference time and communication cost [198], [199], [208]. For example, Wen et al. [208] propose a vision-aided detection technique to build keypoint heatmaps. Accordingly, the optimal mmWave beam pair can be conveniently learned by another DNN, thus significantly saving beam selection overhead while boosting the accuracy. Analogously, [199] verifies that, the environment information can contribute to predicting beam and blockage situations in an extremely efficient and timely manner even without pilot training or costly beam scanning.

*2) Semantic Noise-Robustness Optimization:* Semantic noise can be mainly categorized into semantic mismatch & ambiguity and physical noise (e.g., interference-induced symbol or bit errors). In order to cope with the former category of semantic noise, it mainly requires a timely update of shared KBs to reduce the mismatch or periodic retraining of the DNNs with data containing semantic noise. For example, Hu et al. [90] propose an adversarial training method with weight perturbation and train the DNNs by samples with semantic noise [209]. For the latter, Lu et al. [4] establish a confidence-based distillation mechanism to continually refine the embedding in the encoder and decoder, which endows the transceiver with a proper semantic extraction means to adapt to the changing sentences and channel condition. Xu et al. [143] propose an attention-based JSCC method for ISC under different SNRs, in which a channel-wise soft attention network is adopted to dynamically adjust the bit allocation ratio between the source coding and the channel coding. More specifically, when the channel turns worse, more bits will be allocated for the channel coding to compensate for channel changes. Otherwise, extra bits will be allocated for the source coding to improve image quality. The results [143] also confirm that the semantic-level features are more robust against noise than pixel-level features.

*3) Semantics-Aware Re-Transmission:* Semantics-Aware re-transmission (e.g., Hybrid automatic repeat request [HARQ]), which plays a vital role in guaranteeing reliability in conventional communications, is also re-designed in SemCom to reduce semantic transmission errors. For example, considering the inflexibility of directly adopting fixed-length codeword-based SemCom frameworks to transmit varying-length sentences, Jiang et al. [91] propose an SC-RS-HARQ framework by combining semantic coding with *Reed Solomon* (RS) channel coding and HARQ, and demonstrate the superiority of semantic coding and reliability of conventional methods. In particular, as shown in Fig. 14(a), they introduce an end-to-end architecture called SC-HARQ with multiple isomorphic semantic coding-based encoders and decoders.
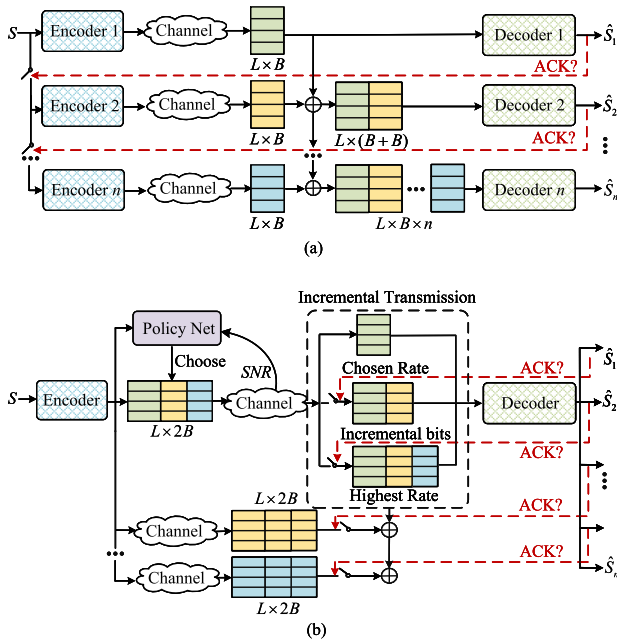
Fig. 14. The illustration of HARQ enhancements for SemCom: (a) DNN structure of SC-HARQ with the maximum re-transmission number *n* and a fixed bit rate *B* [91], (b) DNN structure of IK-HARQ [74] with the peak bit rate 2*B*.

the progress of DL in many fields (e.g., NLP, CV), implicit reasoning has greatly facilitated the semantic transmission of text, images, voice and videos, and yielded fruitful results. Notably, most of the popular semantic transmission designs adopt pre-trained models, and among them, attention-based transformer combined with carefully calibrated loss functions accounts for an overwhelming percentage. Meanwhile, for non-differentiable cases where DL can not be directly applied, DRL sounds like a promising alternative. On the other hand, in order to leverage rich information that is hard to be represented and recovered at first glance, KG-based or probability theory-based explicit reasoning techniques complement the aforementioned implicit reasoning ones, by inferring the incomplete information at the receiver. For channel semantics, existing efforts primarily focus on unveiling CSI and critical surrounding environment information. Meanwhile, the existence of semantic noise and the shift towards semantic-level accuracy requires a re-thinking on noise-robustness optimization and re-transmission techniques, so as to better deal with the channel dynamics. These aforementioned results of SL-SemCom encourage to make bold endeavors towards the effectiveness level of Weaver's vision in Fig. 3.

When one decoder fails to interpret the correct source message, the same content will be re-transmitted at a fixed bit rate but the previously received bits will be concatenated with the latest received bits and re-used as incremental information for decoding at another decoder. In the SC-HARQ, the process repeats until the receiver successfully recovers the sentence or reaches the maximum number of re-transmissions. Notably, Sim32 is adopted to check the semantic accuracy in the recovered sentences. In this way, the whole system confirms the potential of introducing HARQ to SemCom, and can save bandwidth resources by reducing the number of re-transmission requests for some lossy channels. Different from [91], which employs multiple encoders and decoders for retransmission at a fixed bit rate, as shown in Fig. 14(b), Zhou et al. [74] exploit an incremental knowledge HARQ-based JSCC scheme, which can work even with one unified single decoder, thus further simplifying the DNN. In addition, Shao et al. [158] present a selective retransmission-based communication scheme that leverages Tishby's IB principle [102] for task-relevant feature extraction and capably identify and skip redundant features to avoid compensating for the potential performance loss due to the variational approximations for IB, thus contributing to further reducing the communication load.

### E. Summary of Lessons Learned

Recalling the technical progress of SL-SemCom, end-to-end communications play a rather contributing role. In some sense, though semantics has not been taken into account, the successful replacement of conventional communication modules by end-to-end autoencoders significantly promotes the arrival of SL-SemCom. On the other hand, SL-SemCom manifests itself in many aspects. For the content semantics, benefiting from

## VI. TOWARDS EFFECTIVENESS-LEVEL SEMCOM AND NETWORKED SYSTEMS

As the myriad of autonomous smart devices, such as robots and UAVs, emerge and are empowered with advanced sensing, computing, and learning capabilities, the transmission of a huge amount of data (on the order of zettabytes) possibly congests the networks. For example, a swarm of mobile robots may involve the transmission of 1 GB aggregated data per second for target tracking or collaborative sensing [68]. These communication bottlenecks, if left unresolved, will severely limit the growth and utilization of networked systems [210]. In this regard, adopting SemCom as a bridge between multiple agents to communicate and exchange desired information can effectively contribute to successful task execution. Notably, earlier works [211], [212] on goal-oriented communications address the issue of potential "misunderstanding" among parties, which arises from lack of initial agreement, protocols and languages used in communications due to the lack of consistent semantics [1]. Hence, SemCom can also be regarded as a special kind of goal-oriented communications.

Furthermore, for safe and successful task execution, it is crucial to take into account the timeliness and value of information. Therefore, some SemCom works focus on the semantics of specific tasks, and only the semantic information closely relevant to task execution is transmitted for decision-making at the receiver. In other words, in quest of exploring the semantic information at the effectiveness level, EL-SemCom is different from the aforementioned SL-SemCom, and is envisioned at a higher level. In particular, instead of transmitting all the semantics, EL-SemCom attempts to only transmit the essential goal-oriented semantics in a timely manner.

In this section, we first explore the latest advances towards the EL-SemCom. More specifically, we introduce point-to-point and point-to-multi-point task processing, such as

MU-DeepSC [67], U-DeepSC [159]. Meanwhile, we discuss the exploration of the age and value of semantics, as the semantics might be dynamic and some networked intelligent systems could incorporate time as part of semantics. Afterwards, we discuss some representative semantic networks and applications.

### A. Semantics in EL-SemCom

In this part, we begin with the works of EL-SemCom frameworks on single-modal and multi-modal data transmission. Then, we generalize EL-SemCom by extending the application of semantics to effectively facilitate task execution.

For the single-modal EL-SemCom frameworks, Jankowski et al. [145], [213] study the image re-identification task for edge devices, by remotely retrieving similar images from an edge server, such as persons or cars captured by other cameras. Correspondingly, a JSCC-based solution significantly increases the end-to-end accuracy and speeds up the encoding process. Yang et al. [214] present a semantic communication paradigm with AI tasks (SC-AIT), in which the functionalities corresponding to the three levels of communications are clearly identified. Specifically, the effectiveness level mainly concerns the source semantics, the desired conduct of the AI tasks and the KB, while the semantic level mainly involves semantic extraction and transmission. Experiments in image classification and surface defect detection have verified the superiority over conventional schemes in terms of classification accuracy and delay. In order to be tightly coupled with downstream tasks, Kang et al. [215] aim at building an aerial image transmission paradigm under limited computation resources for remote sensing, in which the high-resolution images captured by onboard cameras should be classified immediately. Furthermore, Kang et al. [215] propose a DRL-based algorithm to exploit the semantic blocks most contributing to the back-end classifier under various channel conditions, thus better balancing the tradeoff between transmission latency and classification accuracy.

On the other hand, the multimodal data transmission has also been investigated under various configurations. Xie et al. [67] propose a multi-user SemCom system (MU-DeepSC) for the VQA task to improve the answer accuracy, where one user transmits text-based questions about images, while the answered images are transmitted from another user. In particular, the MU-DeepSC transmitter adopts a memory, attention, and composition-based DNN to extract the essential semantic information of initial data from several transmitters, and then the receiver tries to predict the answers by directly merging different semantics, where LSTM and CNN are adopted for text and image transmitters respectively. On top of MU-DeepSC, a transformer-based framework [138], which leverages the same DNN structure for the text delivery at the transmitter side but different DNN structures for image retrieval at the receiver side, is presented to cope with various tasks (e.g., machine translation and VQA tasks). Nevertheless, though the DNN structure of the transmitter in [138] can be shared across different tasks, it remains a non-unified receiver incapably adapting to various tasks. In order to cope with this drawback, Zhang et al. [159] present a unified DL-enabled
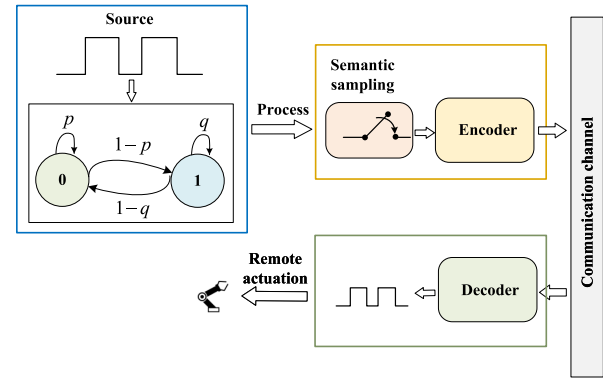


Fig. 15. An illustration of the goal-oriented end-to-end information sampling and transmission in EL-SemCom [68].

SemCom system (U-DeepSC) to serve various transmission tasks by unifying both the transmitter and the receiver, which can simultaneously handle five tasks and significantly outperform those task-oriented models designed for a specific task. In addition, to further explore the useful semantic information for executing a certain task, Farshbafan et al. [81] introduce the concept of beliefs and define a common language between the speaker and the listener, by which the speaker could describe the environmental observations to the listener and thus the listener becomes competent to take corresponding actions. By using a top-down curriculum learning [75] framework based on RL, [81] simultaneously minimizes the required time and transmission cost of task execution.

Moreover, for edge and IoT devices with limited computation resources, some schemes also incorporate the communications between edge devices in the framework of SemCom, so as to realize an end-to-end semantics transmission and enhance communication efficiency. Kountouris and Pappas [68] propose an end-to-end EL-SemCom framework (named E2E Semantics), in which the samples are triggered, generated and transmitted by informative data to control the remote actuator, and the smart devices have the ability to steer their traffic via semantics-aware active sampling. For example, in an end-to-end time-slotted system shown in Fig. 15, the device monitors a two-state Markovian source with different transition probability $p$ and $q$, and transmits sampled updates of the source's status to a remote actuator. Then the real-time source reconstruction is performed at the receiver side based on the received updates, to achieve a real-time actuation goal. It is noted that different from the change-triggered sampling policy, this end-to-end semantic policy can simultaneously measure the changes at the source and track the differences between the two ends, thus avoiding some redundant transmissions. As such, the most useful semantics for goal execution is transmitted to control the remote actuator. As for performance evaluation, in the slow varying sources ($p = 0.95$, $q = 0.9$), the end-to-end semantic policy significantly outperforms the change-triggered policy as it can eliminate the discrepancy quickly. For the rapidly varying sources ($p = 0.8$, $q = 0.3$), the end-to-end semantics provides a lower actuation error without wasting resources [68].

In addition, some RL-based works [76], [150] have been developed to facilitate the accomplishment of specific goals. The goal-oriented communications can be generalized to

multi-agent scenarios, where the agents communicate with each other in an abstract manner over noisy channels while trying to maximize prescribed reward functions. In these problems, besides taking actions and interacting with the environment, all agents transmit and receive semantics. Here, semantics serves as a measure for the usefulness of messages with respect to the goal of data exchange, thus enabling coordination and cooperation among the agents to maximize their accumulated reward. For example, in [76], SemCom is applied to the energy harvesting networks to reduce the transmission delay and energy consumption, where users jointly determine the transmission of partial semantic information and the number of resource blocks used. The authors in [216] formulate the multi-agent collaboration with noisy communications problem as a multi-agent MDP, where the objective of communications aims to speed up the accomplishment of a common goal, and adopt a value-decomposition-based DRL. Moreover, the authors in [150] develop a semantics-aware collaborative DRL method to enable a group of untrained heterogeneous agents with semantically-linked tasks to collaborate efficiently across a resource-constrained wireless cellular network.

To further explore the relevance and importance between semantic features and corresponding tasks, Liu et al. [217] develop a framework for task-oriented multi-user SemCom that enables users to extract, compress, and transmit the semantics of the raw data effectively to the edge server. Specifically, [217] considers a scenario composed of an edge server and a set of users, where the user aims at gathering data locally and sending the extracted semantics to the edge server in a scheduled manner. Besides, the users are prioritized according to the specific service level with distinctive objectives. Correspondingly, [217] proposes an adaptable semantic compression approach to compress extracted semantics. Furthermore, [217] introduces a compression ratio and resource allocation (CRRA) algorithm to obtain the convergent solution and a CRRA-based dynamic user selection algorithm to handle various service levels. The simulation results [217] also demonstrate the superiority in reducing the size of transmitted data and successful task execution over the baselines, especially under limited communication resources.

Similarly, as there exists a tradeoff between the compactness and informativeness (measured in the IB) of messages, the aforementioned work in [37] also applies an attention-based neighbor sampling mechanism to the graph convolutional network (GCN)-based communication (GraphComm) for multi-agent cooperation to relieve the IB. On one hand, the observations and intentions are coded compactly by variational IB (VIB) [35] to extract the semantics as much as possible. Meanwhile, a graph IB (GIB) [218] is utilized to avoid sending similar information multiple times and thus saves the bandwidth consumption.

### B. Semantics-Assisted Networked Intelligent Systems

The effectiveness of end-to-end SemCom frameworks promises potential performance improvement for intelligent networks such as industrial IoTs (IIoTs), UAVs systems, autonomous vehicles, and smart healthcare. Next, we will introduce some typical cases of these semantics-assisted networks.

*1) Industrial IoTs (IIoTs):* IIoTs such as intelligent transportation [219], smart grids [220], and industrial automation [221], [222] belong to a generic framework that exploits the abundance of available data generated by a wirelessly connected set of sensors, actuators, and a remote controller to improve the efficiency, reliability & accuracy of an industrial manufacturing process. Essentially, the networked sensors observe the environment and report the states of interest. Based on the received report, the remote controller on the cloud generates the control signals, and then wireless actuators follow the received signals to take actions.

Since the energy consumption of sensor nodes is mainly due to data transmission, some works [223], [224], [225] concentrate on improving the energy efficiency of wireless transmission. Li et al. [223] consider the freshness, relevance, and value of information of transmitted messages and demonstrate their superiority. Meanwhile, as for semantics-aware data processing, the optimal sampling theory for communication and control systems shall be reconsidered from the perspective of data significance. Mohanty et al. [224] present an RNN-LSTM-based distributed data mining model consisting of RNN and LSTM, which divides the layers of DNNs into separate parts, and deploy them to different sensors. In this sense, information can be exchanged over the network with reduced overhead at the fusion center along with a decrease in data transmission.

On the other hand, as DNN models are usually pre-trained by specific datasets under some particular channels, it becomes inevitable to fine-tune the DNN models to adapt to the changes of KB and channels. However, such a mechanism, which has been extensively discussed in Section V-D1, becomes cumbersome for resource-limited IoT devices. Therefore, inspired by the model trading framework in collaborative edge learning [226], Liew et al. [151] adopt a DL-based hierarchical trading system to support the joint trading of semantic model and semantic information, where model providers trade the well-trained models to support devices with limited transmission resources. In particular, a semantic model provider has more resources to train high-quality semantic models with the relevant KBs and channel models, and the edge devices can directly exploit the semantic model to extract semantics. Thus, the semantic information can be collected and traded between devices from the interested information buyers, while ensuring individual rationality and incentive compatibility. In this regard, the proposed method [151] contributes to the utility enhancement of devices by semantic model trading.

*2) UAV Communications Networks:* UAV systems play an important role in not only working as a stand-alone system in dedicated areas [227] but also serving as complementary parts of the cellular networks [228], [229], [230], [231].

Stand-alone UAV systems aim to collaboratively accomplish tasks (e.g., providing services for distributed IoT devices) under constraints like low latency, energy consumption, and high reliability. Reference [148] introduces a graph attention exchange neural network to satisfy the stringent reliability and latency requirements for real-time air-to-ground communication networks. In particular, each UAV agent swaps attention weights with its neighbors so as to reduce the attention mismatch. Moreover, considering the public safety problems

caused by illegal UAVs, by using a specifically designed tri-residual Siamese neural network (TRSN) based on the radio frequency (RF) fingerprints, [149] proposes two few-shot learning UAV recognition methods (i.e., F-TRSN and SF-TRSN). Specifically, contingent on limited trading samples, F-TRSN directly extracts features of the RF signal, while SF-TRSN utilizes the semantic features of the obtained RF signal by combining the TRSN with a support vector machine (SVM). Notably, outperforming the F-TRSN in terms of UAV recognition accuracy, SF-TRSN further demonstrates that the exploitation of the semantic features provides additional benefit than that of data features only.

On the other hand, the flexibly deployed UAVs can communicate cooperatively to accomplish tasks (e.g., joint rescuing or sensing), and provide emergent ultra-reliable communications to ground users [229] and boost the network performance [230], [231]. Especially, in the latter aspect, it becomes imperative to design appropriate path-planning methods for UAVs. Reference [227] develops a DRL-based path-planning method to collect data generated by sensors in a manner of centralized training and decentralized execution (CTDE). In other words, a flying UAV system can be regarded as an implementation of DRL-based EL-SemCom.

*3) Autonomous Vehicles:* For autonomous vehicles, a self-driving car with no or little human intervention is becoming possible [232]. The most crucial module of autonomous vehicles lies in perception, which targets to sense the surrounding environments and extract the useful information for navigation [233]. Though each vehicle has its individual perception sensors and is fully trained based on large-scale training data, it is still insufficient to meet the high-demand capacity to complement the limited field of view. Hence, the vehicles need to share the collective perception message (CPM) with each other in the same area from a holistic perspective [233], which can be regarded as a multi-agent system.

Benefiting from the advance in V2X (Vehicle-to-Everything) communications, vehicles could reliably exchange their messages with their neighbors, thus enhancing the robustness and safety of transportation systems [234]. However, sharing the raw sensory data requires huge overhead and might congest the communication network, which is impractical in most cases. Therefore, SemCom can be applied to aggregate the most useful information in an efficient manner. Coincidentally, it is natural to apply SemCom to efficiently aggregate the essential information from neighboring vehicles, which has been widely discussed in some multi-agent collaboration works [235], [236], [237]. Besides, as SemCom possesses the capability of reasoning, it is also feasible to supplement the missing part of the received information and enhance its robustness. Furthermore, since autonomous vehicles require the real-time information delivery of events, locations, and time, it is significant to introduce the concept of the VoI and AoI to achieve accurate temporal and spatial alignment, such as DiscoNet [179] and V2VNet [238].

*4) Smart Healthcare:* Smart healthcare promises a new technological shift toward efficient, convenient, and faster medical services based on AI and Internet of medical things (IoMT), and could significantly benefit the living quality of human beings [239]. For example, smart healthy monitoring (SHM) [240] could pose many significant advantages, such as preventing some unnecessary visits to hospitals and checking the patient's condition in a real-time manner. Meanwhile, the data generated by SHM and IoMT devices can be analyzed to prevent chronic diseases and fatality of patients [241], [242].

Considering the battery limitations of wearable sensors and healthcare devices, it belongs to one of the paramount important issues to accomplish efficient services under constrained energy. AI and edge computing techniques can be used for real-time responses and energy consumption minimization. As SemCom supports data filtering by feature extraction, it can help devices only exchange truly meaningful data between devices, thus improving energy efficiency and shortening communication latency.

### C. Summary of Lessons Learned

The effectiveness of executing tasks on the basis of EL-SemCom sounds promising, as the comprehensive results from both single-modal and multi-modal data transmission provide encouraging evidence to facilitate subsequent down-stream tasks. Meanwhile, together with the wide adoption of AI (e.g., RL and GCN), incorporating the EL-SemCom at the edge also brings preliminary yet appealing outcomes. On the other hand, as the process of networked intelligent systems entails a timely and effective methodology towards satisfying the urgency and value of task execution, semantics-assisted networked intelligent systems could enhance the capability to effectively accomplish tasks with minimal cost. Nevertheless, despite the research progress, fundamental performance enhancements of distributed edge learning [243] are still at their infancy and the application of semantics-assisted networks is still worth further exploration.

## VII. CHALLENGES AND OPEN ISSUES

There is a plethora of challenges that need to be tackled before SemCom can be applied to practical communication scenarios. In this section, we list several key challenges to be further explored in future investigations.

1) *Semantic Theory*: In the past few decades, researchers have mostly followed the classic framework of CIT and shed little light on the SIT by extending the logic probability to the scope of semantic entropy. The rationality of quantifying SemCom through semantic entropy and semantic channel capacity remains a fundamental question. In addition, the transmitted symbols in conventional communication systems are assumed to be fixed. However, in SemCom, the flexibility of semantics and the complexity of the languages causes the symbol set to change dynamically and possibly exhibit polysemy. Therefore, how to process and model these dynamic sets is still a problem worth studying. Moreover, though the source-channel separation theorem plays a vital role in reliable communications, the widely adopted JSCC in SemCom makes it interesting to study the theoretical validity of such a separation theorem.

2) *Semantic Similarity Metrics*: Several works have developed the semantic similarity metrics for different types of sources (e.g., text, image, or speech), but considering its important role in implementing the SemCom, a general semantic similarity metric suitable for various tasks needs further exploration. For instance, in order to determine the loss function and pre-train the parameters of DNNs for diversified tasks, an appropriate semantic similarity metric is essential. Meanwhile, the adopted DNN also needs to be carefully designed to avoid gradient vanishing.

3) *Real-time Requirement*: In general, the transceiver of SemCom is relatively more complicated, and it is natural to ask whether SemCom can meet the ultra-low latency requirements for future communications. In addition, although semantic reasoning can correct errors in transmission, an extra delay will be introduced in the semantic reasoning process. Meanwhile, the applicability of SemCom to low-cost IoT devices is worthy of further study. Hence, it is of significant importance by developing lightweight algorithms and improving the hardware design.

4) *Scalability*: SemCom can provide an effective and sustainable service in bandwidth saving and task processing. Besides, the physical layer and application layers are mutually independent in the widely-adopted open-system interconnection (OSI) model, and thus the upper application layer is primarily responsible for the semantic understanding of contents. Nevertheless, though some works for multimodal transmission have proven the effectiveness to process the text and image simultaneously, it is still a huge challenge to deal with more complex data types under the legacy OSI models. In other words, a general semantic-level framework for different types of sources has not been available yet.

On the other hand, since it is a pre-requisite to share semantic KBs between the transmitter and receiver, the continual update and maintenance of a semantic KB surely involve extra storage costs and algorithm design, thus requiring intensive computational and storage costs. Consequently, how to ensure the scalability of SemCom remains to be addressed.

5) *Privacy*: As most of the implementation of SemCom depends on AI technologies, especially DL techniques, it poses significant privacy concerns to collect the possibly sensitive user data for the training data or extract the hidden user information of the trained models. In this sense, beyond the capability of ML, we need to figure out the trustworthiness [244]. In particular, privacy protection corresponds to two different levels (i.e., data level and model level). As for the former level, simply deleting sensitive features and entities might violate the data integrity and consistency, since the "missingness" may reveal some data properties. Hence, data obfuscation, sanitization, and synthesis are often applied to mask, scramble, or overwrite the sensitive information with a realistic fake [245], [246]. As for model level, the defense techniques (e.g., model augmenting, differential privacy, FL, and data encryption) can be leveraged to protect existing trained models from leaking sensitive information [247], [248]. Furthermore, considering that the correct information recovery in SemCom requires the matching of KBs at both the sender and receiver, from the perspective of privacy protection in SemCom, encrypting the KB provides a viable means to prevent semantic leakage. For example, even though a user successfully eavesdrops on bits sent to others, the user without the required KB can not accurately interpret these bits, thus enhancing privacy and security. From this regard, it is necessary to address the security of KBs and investigate the means to encrypt the semantic messages during transmission.

6) *Semantic networks*: 6G is anticipated to transform into a sizable and decentralized system, so as to better support intelligent networked systems at different communication levels [249]. We boldly argue that SemCom can be applied to distributed intelligent networks and even become the dominant communication architecture of 6G with less data transmission and more knowledge exploitation. The embedded generative capability in SemCom pays the way for the orchestration of LLMs in 6G, so as to fully unleash converged communications and computing abilities [250]. However, until to now, there is no specific definition for the blueprint of the semantic network. Besides the aforementioned challenges for end-to-end SemCom, a semantic network also faces the difficulty to disseminate the semantic KB to distributed devices, so as adapt to the network heterogeneity.

## VIII. CONCLUSION

The purpose of this tutorial-cum-survey is to provide a comprehensive understanding of the state-of-the-art works for SemCom and its applications. First, we have reviewed the evolution of SemCom and provided a clear definition and explanation of the semantic-empowered mechanism in typical communication scenarios. Next, we have elaborated on the corresponding ecosystem consisting of history, theoretical guidance, metrics, toolkits, etc. Afterwards, we have carefully reviewed two major categories of SemCom techniques, i.e., implicit reasoning and explicit reasoning, wherein the former typically relies on structured or unstructured parameterized models (e.g., DNNs), while the latter involves explicitly defined rules or entities. In particular, we have presented an overview of the recent progress on techniques to explore and exploit the content and channel semantics. On top of these end-to-end SemCom approaches, we have discussed typical semantics-assisted networked systems, where possible applications and opportunities are envisioned as well.

Through exchanging the most informative, timely, and effective information, SemCom is capable of improving resource utilization, communication efficiency and effectiveness towards accomplishing a task, and can provide a paradigm and technical basis for the next generation communication systems. Therefore, semantics-aware communication will play

an important role in future intelligent systems. The objective of this article has been to provide a primer and a unified view of this semantics-aware communications.

## REFERENCES

[1] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107930.

[2] G. Shi et al. "A new communication paradigm: From bit accuracy to semantic fidelity." Jan. 2021. [Online]. Available: https://arxiv.org/abs/2101.12649

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[4] K. Lu et al., "Rethinking modern communication from semantic coding to semantic communication," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 158–164, Feb. 2023.

[5] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 2326–2330.

[6] W. Weaver, "Recent contributions to the mathematical theory of communication," *Math. Theory Commun.*, vol. 10, no. 4, pp. 261–281, Sep. 1949.

[7] W. Yang et al., "Semantic communications for future Internet: Fundamentals, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, 1st Quart., 2023.

[8] Q. Lan et al., "What is semantic communication? A view on conveying meaning in the era of machine intelligence," *J. Commun. Inf. Netw.*, vol. 6, no. 4, pp. 336–371, Dec. 2021.

[9] Z. Qin, X. Tao, J. Lu, and G. Y. Li. "Semantic communications: Principles and challenges." Jun. 2022. [Online]. Available: https://arxiv.org/abs/2201.01389

[10] S. Iyer et al., "A survey on semantic communications for intelligent wireless networks," *Wireless Pers. Commun.*, vol. 129, no. 1, pp. 569–611, Nov. 2022.

[11] W. Yang et al., "Semantic communication meets edge intelligence," *IEEE Wireless Commun.*, vol. 29, no. 5, pp. 28–35, Oct. 2022.

[12] D. Wheeler and B. Natarajan, "Engineering semantic communication: A survey," *IEEE Access*, vol. 1, pp. 13965–13995, 2023, doi: 10.1109/ACCESS.2023.3243065.

[13] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor. "Less data, more knowledge: Building next generation semantic communication networks." Nov. 2022. [Online]. Available: https://arxiv.org/abs/2211.14343

[14] W. J. Baker, *A History of the Marconi Company 1874–1965*. London, U.K.: Routledge, 2013.

[15] L. Qiao, Y. Li, D. Chen, S. Serikawa, M. Guizani, and Z. Lv, "A survey on 5G/6G, AI, and robotics," *Comput. Electr. Eng.*, vol. 95, Oct. 2021, Art. no. 107372.

[16] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[17] M. N. Patwary, S. J. Nawaz, M. A. Rahman, S. K. Sharma, M. M. Rashid, and S. J. Barnes, "The potential short-and long-term disruptions and transformative impacts of 5G and beyond wireless networks: Lessons learnt from the development of a 5G testbed environment," *IEEE Access*, vol. 8, pp. 11352–11379, 2020.

[18] A. Mchangama, J. Ayadi, V. P. G. Jiménez, and A. Consoli, "MmWave massive MIMO small cells for 5G and beyond mobile networks: An overview," in *Proc. Int. Symp. Commun. Syst. Netw. Digit. Signal Process. (CSNDSP)*, Porto, Portugal, Nov. 2020, pp. 1–6.

[19] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.

[20] J. Choi, S. W. Loke, and J. Park, "A unified view on semantic information and communication: A probabilistic logic approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Seoul, South Korea, May 2022, pp. 705–710.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[22] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Feb. 2015, pp. 1440–1448.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jul. 2016, pp. 770–778.

[24] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Glasgow, U.K., Nov. 2020, pp. 213–229.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[26] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*. Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[27] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. Int. Conf. Learn. Represent. Workshop (ICLR Workshop)*, Scottsdale, AZ, USA, May 2013, pp. 1–12.

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapt. Assoc. Comput. Linguist. Human Lang. Technol. (NAACL HLT)*, Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186.

[29] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. "Improving language understanding by generative pre-training." 2018. [Online]. Available: https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

[30] S. Liu, Z. Zhu, N. Ye, S. Guadarrama, and K. Murphy, "Improved image captioning via policy gradient optimization of spider," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 873–881.

[31] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 1025–1035.

[32] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Apr. 2017, pp. 1–14.

[33] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, Apr. 2018, pp. 1–12.

[34] R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, "TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2000–2011, Apr. 2014.

[35] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep variational information bottleneck," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Apr. 2017, pp. 1–19.

[36] M. Tucker, J. Shah, R. Levy, and N. Zaslavsky. "Towards human-agent communication via the information bottleneck principle." Jun. 2022. [Online]. Available: https://arxiv.org/abs/2207.00088

[37] Q. Yuan, X. Fu, Z. Li, G. Luo, J. Li, and F. Yang, "GraphComm: Efficient graph convolutional communication for multiagent cooperation," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16359–16369, Nov. 2021.

[38] H. Seo, J. Park, M. Bennis, and M. Debbah. "Semantics-native communication with contextual reasoning." Aug. 2021. [Online]. Available: https://arxiv.org/abs/2108.05681

[39] D. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Banff, AB, Canada, Apr. 2014, pp. 1–14.

[40] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, Dec. 2014, pp. 2672–2680.

[41] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2020, pp. 1–12.

[42] K. Lu, R. Li, and H. Zhang, "Contrastive monotonic pixel-level modulation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2022, pp. 493–510.

[43] Y. Hua, R. Li, Z. Zhao, X. Chen, and H. Zhang, "GAN-powered deep distributional reinforcement learning for resource management in network slicing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 2, pp. 334–349, Feb. 2020.

[44] H. Ye, F. Gao, J. Qian, H. Wang, and G. Y. Li, "Deep learning-based denoise network for CSI feedback in FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1742–1746, Apr. 2020.

[45] D. Silver et al., "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[46] R. Li et al., "Deep reinforcement learning for resource management in network slicing," *IEEE Access*, vol. 6, pp. 74429–74441, 2018.

[47] B. R. Kiran et al., "Deep reinforcement learning for autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 4909–4926, Jun. 2022.

[48] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "ComNet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, Oct. 2018.

[49] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[50] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[51] T. J. O'Shea, T. Erpek, and T. C. Clancy. "Deep learning based MIMO communications." Jul. 2017. [Online]. Available: https://arxiv.org/abs/1707.07980

[52] T. Erpek, T. J. O'Shea, and T. C. Clancy, "Learning a physical layer scheme for the MIMO interference channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–5.

[53] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, Apr. 2021.

[54] K. Lu, R. Li, X. Chen, Z. Zhao, and H. Zhang. "Reinforcement learning-powered semantic communication via semantic similarity." Aug. 2021. [Online]. Available: https://arxiv.org/abs/2108.12121

[55] Y. Wang, M. Chen, W. Saad, T. Luo, S. Cui, and H. V. Poor, "Performance optimization for semantic communications: An attention-based learning approach," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[56] R. Li, Z. Zhao, X. Xu, F. Ni, and H. Zhang, "The collective advantage for advancing communications and intelligence," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 96–102, Aug. 2020.

[57] R. Li et al., "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[58] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal. Process.*, vol. 66, no. 11, pp. 2905–2917, Jun. 2018.

[59] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, Feb. 2020.

[60] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Hoboken, NJ, USA: Prentice-Hall, 1993.

[61] R. Carnap and Y. Bar-Hillel. "An outline of a theory of semantic information." Oct. 1952. [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/4821/RLE-TR-247-03150899.pdf?sequence=1

[62] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE. Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001.

[63] A. Jiang, Y. Li, and J. Bruck, "Error correction through language processing," in *Proc. IEEE Inf. Theory Workshop. (ITW)*, Jerusalem, Israel, Apr. 2015, pp. 1–5.

[64] Y. Zhong, "A theory of semantic information," *China Commun.*, vol. 14, no. 1, p. 129, Jun. 2017.

[65] L. Floridi, "Outline of a theory of strongly semantic information," *Mind Mach.*, vol. 14, no. 2, pp. 197–221, May 2004.

[66] J. Bao et al., "Towards a theory of semantic communication," in *Proc. IEEE Net. Sci. Workshop (NSW)*, West Point, NY, USA, Jun. 2011, pp. 110–117.

[67] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented multi-user semantic communications for VQA," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 553–557, Mar. 2022.

[68] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 96–102, Jun. 2021.

[69] N. Pappas and M. Kountouris, "Goal-oriented communication for real-time tracking in autonomous systems," in *Proc. IEEE Int. Conf. Auton. Syst. (ICAS)*, Montreal, QC, Canada, Aug. 2021, pp. 1–5.

[70] C. Zhang, H. Zou, S. Lasaulce, W. Saad, M. Kountouris, and M. Bennis, "Goal-oriented communications for the IoT and application to data compression," *IEEE Internet Things Mag.*, vol. 5, no. 4, pp. 58–63, Dec. 2022.

[71] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[72] Q. Zhou, R. Li, Z. Zhao, C. Peng, and H. Zhang, "Semantic communication with adaptive universal transformer," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 453–457, Mar. 2022.

[73] Y. Xiao, Y. Li, G. Shi, and H. V. Poor, "Reasoning on the air: An implicit semantic communication architecture," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Seoul, South Korea, May 2022, pp. 289–294.

[74] Q. Zhou, R. Li, Z. Zhao, Y. Xiao, and H. Zhang, "Adaptive bit rate control in semantic communication with incremental knowledge-based HARQ," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1076–1089, 2022.

[75] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum learning for reinforcement learning domains: A framework and survey," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 1–50, Jul. 2020.

[76] M. Chen, Y. Wang, and H. V. Poor, "Performance optimization for wireless semantic communications over energy harvesting networks," in *Proc. IEEE Int Conf. Acoust. Speech Signal Process. (ICASSP)*, Singapore, May 2022, pp. 8647–8651.

[77] J. Chen, N. Skatchkovsky, and O. Simeone, "Neuromorphic wireless cognition: Event-driven semantic communications for remote inference," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 2, pp. 252–265, Apr. 2023.

[78] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, Mar. 2021.

[79] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *Proc. ACM/IEEE Int. Conf. Cyber Phys. Syst. (ICCPS)*, Montreal, QC, Canada, Apr. 2019, pp. 109–117.

[80] A. Molin, H. Esen, and K. H. Johansson, "Scheduling networked state estimators based on value of information," *Automatica*, vol. 110, Dec. 2019, Art. no. 108578.

[81] M. K. Farshbafan, W. Saad, and M. Debbah, "Common language for goal-oriented semantic communications: A curriculum learning framework," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 1710–1715.

[82] C. K. Ogden and I. A. Richards, *The Meaning of Meaning: A Study of the Influence of Thought and of the Science of Symbolism*. New York, NY, USA: Harcourt, Brace & World, 1923.

[83] Z. Ognjanović, M. Rašković, and Z. Marković, *Probability Logics*. Cham, Switzerland: Springer,, 2009.

[84] D. Ellerman, "An introduction to logical entropy and its relation to Shannon entropy," *Int. J. Semant. Comput.*, vol. 7, no. 2, pp. 121–145, 2013.

[85] D. Ellerman, *New Foundations for Information Theory: Logical Entropy and Shannon Entropy*. Cham, Switzerland: Springer, 2021.

[86] Y. Bar-Hillel and R. Carnap, "Semantic information," *Brit J. Philos. Sci.*, vol. 4, no. 14, pp. 147–157, Sep. 1953.

[87] S. D'Alfonso, "On quantifying semantic information," *Information*, vol. 2, no. 1, pp. 61–101, 2011.

[88] Y. Xiao, X. Zhang, Y. Li, G. Shi, and T. Başar, "Rate-distortion theory for strategic semantic communication," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Mumbai, India, Nov. 2022, pp. 279–284.

[89] N. Merhav and S. Shamai, "On joint source-channel coding for the Wyner-Ziv source and the Gel'Fand-Pinsker channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2844–2855, Nov. 2003.

[90] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications against semantic noise," in *Proc. Veh. Technol. Conf. (VTC)*, London, U.K., Jan. 2023, pp. 1–6.

[91] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep source-channel coding for sentence semantic transmission with HARQ," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5225–5240, Aug. 2022.

[92] J. Liang, Y. Xiao, Y. Li, G. Shi, and M. Bennis, "Life-long learning for reasoning-based semantic communication," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Seoul, South Korea, May 2022, pp. 271–276.

[93] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.

[94] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and application of a metric on semantic nets," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 19, no. 1, pp. 17–30, Jan./Feb. 1989.

[95] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *J. Document.*, vol. 28, no. 1, pp. 11–21, Jan. 1972.

[96] P. Resnik. "Using information content to evaluate semantic similarity in a taxonomy." Nov. 1995. [Online]. Available: https://arxiv.org/abs/cmp-lg/9511007

[97] S. Amir, A. Tanasescu, and D. A. Zighed, "Sentence similarity based on semantic kernels for intelligent text retrieval," *J. Intell. Inf. Syst.*, vol. 48, no. 3, pp. 675–689, Jun. 2017.

[98] Y. Le, Z.-J. Wang, Z. Quan, J. He, and B. Yao, "ACV-tree: A new method for sentence similarity modeling," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Stockholm, Sweden, Jul. 2018, pp. 4137–4143.

[99] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discour. Process.*, vol. 25, nos. 2–3, pp. 259–284, 1998.

[100] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.

[101] J. Sarzynska-Wawer et al., "Detecting formal thought disorder by deep contextualized word representations," *Psychiat. Res.*, vol. 304, Oct. 2021, Art. no. 114135.

[102] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2000, pp. 368–377.

[103] F. Liu, W. Tong, Z. Sun, and C. Guo. "Task-oriented semantic communication systems based on extended rate-distortion theory." Apr. 2022. [Online]. Available: https://arxiv.org/abs/2201.10929

[104] B. Wang, R. Li, J. Zhu, Z. Zhao, and H. Zhang, "Knowledge enhanced semantic communication receiver," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1794–1798, Jul. 2023.

[105] M. Sana and E. C. Strinati, "Learning semantics: An opportunity for effective 6G communications," in *Proc. IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 2022, pp. 631–636.

[106] P. M. Nugues, *An Introduction to ProLog*. Berlin, Germany: Springer, 2006.

[107] B. Guler and A. Yener, "Semantic index assignment," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM Workshops)*, Budapest, Hungary, Mar. 2014, pp. 431–436.

[108] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. Annu. Meet. Assoc. Comput. Linguist. (ACL)*, Jul. 2002, pp. 311–318.

[109] R. Vedantam, C. L. Zitnick, and D. Parikh, "CIDEr: Consensus-based image description evaluation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 4566–4575.

[110] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study," *J. Comput. Commun.*, vol. 7, no. 3, pp. 8–18, Mar. 2019.

[111] X. Zhang and W. Zhang, "Semantic image compression based on data hiding," *IET Image Process.*, vol. 9, no. 1, pp. 54–61, Jan. 2015.

[112] C. Li and A. C. Bovik, "Three-component weighted structural similarity index," in *Proc. SPIE. Int. Soc. Opt. Eng*, San Jose, CA, USA, Jan. 2009, pp. 252–260.

[113] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[114] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 6627–6638.

[115] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, May 2018, pp. 1–36.

[116] L. N. Vaserstein, "Markov processes over denumerable products of spaces, describing large systems of automata," *Problemy Peredachi Informatsii*, vol. 5, no. 3, pp. 64–72, 1969.

[117] M. Soloveitchik, T. Diskin, E. Morin, and A. Wiesel. "Conditional Frechet inception distance." Mar. 2021. [Online]. Available: https://arxiv.org/abs/2103.11521

[118] D. Huang, X. Tao, F. Gao, and J. Lu, "Deep learning-based image semantic coding for semantic communications," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[119] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, "Deep learning enabled semantic communications with speech recognition and synthesis," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 6227–6240, Sep. 2023, doi: 10.1109/TWC.2023.3240969.

[120] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.

[121] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1462–1469, Jun. 2006.

[122] P. C. Loizou, "Speech quality assessment," in *Multimedia Analysis, Processing and Communications*. Berlin, Germany: Springer, 2011, pp. 623–654.

[123] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal. Process. (ICASSP)*, Salt Lake, UT, USA, May 2001, pp. 749–752.

[124] M. Bińkowski et al., "High fidelity speech synthesis with adversarial networks," in *Proc. Int. Conf. Learn. Repsent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–17.

[125] Y. Wang et al., "Performance optimization for semantic communications: An attention-based reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2598–2613, Sep. 2022.

[126] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, and R. Q. Hu, "Cognitive semantic communication systems driven by knowledge graph," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seoul, South Korea, May 2022, pp. 4860–4865.

[127] Q. Sun, C. Guo, Y. Yang, J. Chen, and X. Xue. "Semantic-assisted image compression." Jan. 2022. [Online]. Available: https://arxiv.org/abs/2201.12599

[128] P. Cheng, W. Hao, S. Dai, J. Liu, Z. Gan, and L. Carin, "CLUB: A contrastive log-ratio upper bound of mutual information," in *Proc. Int. Conf. Machin. Learn. (ICML)*, Jul. 2020, pp. 1779–1788.

[129] A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: An enabler of semantics-empowered communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2621–2635, Apr. 2023.

[130] S. F. Abedin, M. S. Munir, N. H. Tran, Z. Han, and C. S. Hong, "Data freshness and energy-efficient UAV navigation optimization: A deep reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 9, pp. 5994–6006, Sep. 2021.

[131] M. Yi, X. Wang, J. Liu, Y. Zhang, and B. Bai, "Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Toronto, ON, Canada, Jul. 2020, pp. 716–721.

[132] M. A. Abd-Elmagid, A. Ferdowsi, H. S. Dhillon, and W. Saad, "Deep reinforcement learning for minimizing age-of-information in UAV-assisted networks," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[133] H. Zheng, K. Xiong, P. Fan, Z. Zhong, and K. B. Letaief, "Age-based utility maximization for wireless powered networks: A Stackelberg game approach," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[134] Z. Ning et al., "Mobile edge computing enabled 5G health monitoring for Internet of Medical Things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 463–478, Feb. 2021.

[135] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE Int. Symp. Inf. Theor. (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 1583–1587.

[136] P. Popovski et al. "Internet of Things (IoT) connectivity in 6G: An interplay of time, space, intelligence, and value." Nov. 2021. [Online]. Available: https://arxiv.org/abs/2111.05811

[137] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, Oct. 2020.

[138] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2584–2597, Sep. 2022.

[139] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, "Semantic-preserved communication system for highly efficient speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 245–259, Jan. 2023.

[140] D. B. Kurka and D. Gündüz, "Deep joint source-channel coding of images with feedback," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 5235–5239.

[141] H. Yoo, T. Jung, L. Dai, S. Kim, and C.-B. Chae, "Demo: Real-time semantic communications with a vision transformer," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2022, pp. 1–2.

[142] H. Zhang, S. Shao, M. Tao, X. Bi, and K. B. Letaief, "Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 170–185, Jan. 2022.

[143] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2315–2328, Apr. 2022.

[144] Z. Zhang, Q. Yang, S. He, M. Sun, and J. Chen, "Wireless transmission of images with the assistance of multi-level semantic information," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Hangzhou, China, Feb. 2022, pp. 1–6.

[145] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "Deep joint source-channel coding for wireless image retrieval," in *Proc. IEEE Int. Conf. Acoust. Speech. Signal. Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 5070–5074.

[146] T.-Y. Tung and D. Gündüz, "DeepWiVe: Deep-learning-aided wireless video transmission," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2570–2583, Sep. 2022.

[147] S. Wang et al., "Wireless deep video semantic transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 214–229, Jan. 2023.

[148] W. J. Yun et al., "Attention-based reinforcement learning for real-time UAV semantic communication," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Berlin, Germany, Sep. 2021, pp. 1–6.

[149] H. Liang, R. Wang, M. Xu, F. Zhou, Q. Wu, and O. A. Dobre, "Few-shot learning UAV recognition methods based on the TRI-residual semantic network," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2072–2076, Jun. 2022.

[150] F. Lotfi, O. Semiari, and W. Saad, "Semantic-aware collaborative deep reinforcement learning over wireless cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 5256–5261.

[151] Z. Q. Liew et al., "Economics of semantic communication system: An auction approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13559–13574, Oct. 2023. doi: 10.1109/TVT.2023.3278467.

[152] J. Johnson, B. Hariharan, L. Van Der Maaten, L. Fei-Fei, C. Lawrence Zitnick, and R. Girshick, "CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1988–1997.

[153] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.

[154] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, Jan. 2021.

[155] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 55–71, Jan. 2023.

[156] J. Ballé, V. Laparra, and E. P. Simoncelli, "Density modeling of images using a generalized normalization transformation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2016.

[157] H. Tong et al., "Federated learning for audio semantic communication," *Front. Commun. Netw.*, vol. 2, Sep. 2021, Art. no. 734402.

[158] J. Shao, Y. Mao, and J. Zhang, "Task-oriented communication for multidevice cooperative edge inference," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 73–87, Jan. 2023, doi: 10.1109/TWC.2022.3191118.

[159] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, and X. Tao. "A unified multi-task semantic communication system for multimodal data." Sep. 2022. [Online]. Available: https://arxiv.org/abs/2209.07689.

[160] T. Rashid, M. Samvelyan, C. Schroeder, G. Farquhar, J. Foerster, and S. Whiteson, "QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, Jul. 2018, pp. 4295–4304.

[161] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *Proc. Conf. Rec. Asilomar Conf. Signals Syst. Comput. (ACSSC)*, Pacific Grove, CA, USA, Oct. 2018, pp. 298–303.

[162] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, and Ł. Kaiser, "Universal transformers," in *Proc. Int. Conf. Learn. Repsent. (ICLR)*, New Orleans, LA, USA, May 2019, pp. 1–23.

[163] A. Graves. "Adaptive computation time for recurrent neural networks." Apr. 2016. [Online]. Available: https://arxiv.org/abs/1603.08983

[164] S. Guo and Y. Wang, "Signal shaping for semantic communication systems with a few message candidates," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, London, U.K., Sep. 2022, pp. 1–5.

[165] G. Forecast, "Cisco visual networking index: Global mobile data traffic forecast update, pp. 2017–2022," San Jose, CA, USA, 2019.

[166] G. K. Wallace, "The JPEG still picture compression standard," *ACM Commun.*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[167] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal Process. Image. Commun.*, vol. 17, no. 1, pp. 3–48, Jan. 2002.

[168] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional autoencoder-based lossy image compression," in *Proc. Picture Coding Symp. (PCS)*, San Francisco, CA, USA, Jun. 2018, pp. 253–257.

[169] A. Prakash, N. Moran, S. Garber, A. DiLillo, and J. Storer, "Semantic perceptual image compression using deep convolution networks," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Apr. 2017, pp. 250–259.

[170] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Apr. 2017, pp. 1–19.

[171] F. Kong, K. Hu, Y. Li, D. Li, and S. Zhao, "Spectral–spatial feature partitioned extraction based on CNN for multispectral image compression," *Remote Sens.-Basel.*, vol. 13, no. 1, p. 9, Dec. 2020.

[172] G. Toderici et al., "Full resolution image compression with recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5306–5314.

[173] A. Punnappurath and M. S. Brown, "Learning raw image reconstruction-aware deep image compressors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 1013–1019, Apr. 2020.

[174] O. Rippel and L. Bourdev, "Real-time adaptive image compression," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Sydney, NSW, Australia, Aug. 2017, pp. 2922–2930.

[175] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative adversarial networks for extreme learned image compression," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2019, pp. 221–231.

[176] S. Kudo, S. Orihashi, R. Tanida, S. Takamura, and H. Kimata, "GAN-based image compression using mutual information for optimizing subjective image similarity," *IEICE Trans. Inf. Syst.*, vol. 104, no. 3, pp. 450–460, Mar. 2021.

[177] S. Jamil, M. J. Piran, M. Rahman, and O.-J. Kwon, "Learning-driven lossy image compression: A comprehensive survey," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106361.

[178] S. Zhou, Y. He, Y. Liu, C. Li, and J. Zhang, "Multi-channel deep networks for block-based image compressive sensing," *IEEE Trans. Multimedia*, vol. 23, pp. 2627–2640, Aug. 2020.

[179] M. Li, K. Zhang, J. Li, W. Zuo, R. Timofte, and D. Zhang, "Learning context-based nonlocal entropy modeling for image compression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 3, pp. 1132–1145, Mar. 2023.

[180] D. B. Kurka and D. Gündüz, "Bandwidth-agile image transmission with deep joint source-channel coding," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8081–8095, Dec. 2021.

[181] Q. Hou, L. Zhang, M.-M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4003–4012.

[182] M. Yang and H.-S. Kim, "Deep joint source-channel coding for wireless image transmission with adaptive rate control," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process Proc. (ICASSP)*, Singapore, May 2022, pp. 5193–5197.

[183] F. Bellard. "BPG image format." 2017. [Online]. Available: http://bellard.org/bpg/

[184] P. Zhang et al., "Model division multiple access for semantic communications," *Front. Inform. Technol. Electron. Eng.*, vol. 24, no. 6, pp. 801–812, Jun. 2023.

[185] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[186] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6G: Applications, challenges, and opportunities," *Engineering*, vol. 8, pp. 33–41, Jan. 2022.

[187] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 494–514, Feb. 2022.

[188] R. Koncel-Kedziorski, D. Bekal, Y. Luan, M. Lapata, and H. Hajishirzi, "Text generation from knowledge graphs with graph transformer," in *Proc. Conf. North Amer. Chapt. Assoc. Comput. Linguist. Human Lang. Technol. (NAACL HLT)*, Minneapolis, MN, USA, Jun. 2019, pp. 2284–2293.

[189] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Exp. Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112948.

[190] W. Yu et al., "A survey of knowledge-enhanced text generation," *ACM Comput. Surveys*, vol. 54, no. 11s, pp. 1–38, Jan. 2022.

[191] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, Dec. 2013, pp. 2787–2795.

[192] H. Xiao, M. Huang, Y. Hao, and X. Zhu, "TransG: A generative mixture model for knowledge graph embedding," in *Proc. Annu. Meet. Assoc. Comput. Linguist. (ACL)*, Berlin, Germany, Aug. 2016, pp. 2316–2325.

[193] G. Ji, K. Liu, S. He, and J. Zhao, "Knowledge graph completion with adaptive sparse transfer matrix," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, Phoenix, AZ, USA, Feb. 2016, pp. 985–991.

[194] C. K. Thomas and W. Saad, "Neuro-symbolic artificial intelligence (AI) for intent based semantic communication," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Rio de Janeiro, Brazil, Dec. 2022, pp. 2698–2703.

[195] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. New York, NY, USA: SIAM, 1998.

[196] C. Chen and J. Cruz, "Stackelberg solution for two-person games with biased information patterns," *IEEE Trans. Autom. Control*, vol. 17, no. 6, pp. 791–798, Dec. 1972.

[197] V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica J. Econometr. Soc.*, vol. 50, no. 6, pp. 1431–1451, Nov. 1982.

[198] Z. Qin, F. Gao, B. Lin, X. Tao, G. Liu, and C. Pan. "A generalized semantic communication system: From sources to channels." Mar. 2023. [Online]. Available: https://arxiv.org/abs/2301.04552

[199] Y. Yang, F. Gao, X. Tao, G. Liu, and C. Pan, "Environment semantics aided wireless communications: A case study of mmWave beam prediction and blockage prediction," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 7, pp. 2025–2040, Jul. 2023.

[200] H. Ye, G. Y. Li, B.-H. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," in *Proc. IEEE Globecom Workshops (GCWkshps)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–5.

[201] T. J. O'Shea, T. Roy, and N. West, "Approximating the void: Learning stochastic channel models from observation with variational generative adversarial networks," in *Proc. Int. Conf. Comput. Netw. Commun. (ICNC)*, Honolulu, HI, USA, Feb. 2019, pp. 681–686.

[202] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Sydney, NSW, Australia, Aug. 2017, pp. 1126–1135.

[203] S. Park, O. Simeone, and J. Kang, "End-to-end fast training of communication links without a channel model via online meta-learning," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Atlanta, GA, USA, May 2020, pp. 1–5.

[204] S. Dörner, S. Cammerer, J. Hoydis, and S. Ten Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.

[205] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[206] F. A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2503–2516, Nov. 2019.

[207] V. Raj and S. Kalyani, "Backpropagating through the air: Deep learning at physical layer without channel models," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2278–2281, Nov. 2018.

[208] F. Wen, W. Xu, F. Gao, C. Pan, and G. Liu, "Vision aided environment semantics extraction and its application in mmWave beam selection," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1894–1898, Jul. 2023.

[209] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–11.

[210] E. Uysal et al., "Semantic communications in networked systems: A data significance perspective," *IEEE Netw.*, vol. 36, no. 4, pp. 233–240, Jul./Aug. 2022.

[211] B. Juba, "Universal semantic communication," in *Proc. 40th ACM Symp. Theory Comput. (SCOT)*, Victoria, BC, Canada, May 2008, pp. 123–132.

[212] O. Goldreich, B. Juba, and M. Sudan, "A theory of goal-oriented communication," *J. ACM*, vol. 59, no. 2, pp. 1–65, Apr. 2012.

[213] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "Wireless image retrieval at the edge," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 89–100, Jan. 2021.

[214] Y. Yang et al. "Semantic communications with AI tasks." Sep. 2021. [Online]. Available: https://arxiv.org/abs/2109.14170

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

[216] T.-Y. Tung, J. R. Pujol, S. Kobus, and D. Gunduz, "Effective communications: A joint learning and communication framework for multi-agent reinforcement learning over noisy channels," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2590–2603, Aug. 2021.

[217] C. Liu, C. Guo, Y. Yang, and N. Jiang. "Adaptable semantic compression and resource allocation for task-oriented communications." Apr. 2022. [Online]. Available: https://arxiv.org/abs/2204.08910

[218] T. Wu, H. Ren, P. Li, and J. Leskovec, "Graph information bottleneck," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2020, pp. 20437–20448.

[219] H. Grover, T. Alladi, V. Chamola, D. Singh, and K.-K. R. Choo, "Edge computing and deep learning enabled secure multitier network for Internet of vehicles," *IEEE Internet Things J.*, vol. 8, no. 19, pp. 14787–14796, Oct. 2021.

[220] J. Wang, X. Wang, C. Ma, and L. Kou, "A survey on the development status and application prospects of knowledge graph in smart grids," *IET Gener. Transm. Distrib.*, vol. 15, no. 3, pp. 383–407, Feb. 2021.

[221] W. Liu, P. Popovski, Y. Li, and B. Vucetic, "Wireless networked control systems with coding-free data transmission for industrial IoT," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1788–1801, Mar. 2020.

[222] W. Liu, P. Popovski, Y. Li, and B. Vucetic, "Real-time wireless networked control systems with coding-free data transmission," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[223] G. Li, B. He, Z. Wang, X. Cheng, and J. Chen, "Blockchain-enhanced spatiotemporal data aggregation for UAV-assisted wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4520–4530, Jul. 2022.

[224] S. N. Mohanty, E. L. Lydia, M. Elhoseny, M. M. G. Al Otaibi, and K. Shankar, "Deep learning with LSTM based distributed data mining model for energy efficient wireless sensor networks," *Phys. Commun.*, vol. 40, Jun. 2020, Art. no. 101097.

[225] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, "Internet of Things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios," *IEEE Access*, vol. 8, pp. 23022–23040, 2020.

[226] W. Y. Bryan Lim et al., "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, 3rd Quart., 2020.

[227] S. Wan, J. Lu, P. Fan, and K. B. Letaief, "Toward big data processing in IoT: Path planning and resource management of UAV base stations in mobile-edge computing system," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5995–6009, Jul. 2020.

[228] A. Mostaani, T. X. Vu, S. K. Sharma, V.-D. Nguyen, Q. Liao, and S. Chatzinotas, "Task-oriented communication design in cyber-physical systems: A survey on theory and applications," *IEEE Access*, vol. 10, pp. 133842–133868, 2022, doi: 10.1109/ACCESS.2022.3231039.

[229] M. M. Azari, F. Rosas, K.-C. Chen, and S. Pollin, "Ultra reliable UAV communication using altitude and cooperation diversity," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 330–344, Jan. 2018.

[230] A. M. Seid, G. O. Boateng, B. Mareri, G. Sun, and W. Jiang, "Multi-agent DRL for task offloading and resource allocation in multi-UAV enabled IoT edge network," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 4, pp. 4531–4547, Dec. 2021.

[231] Y. Gu, Y. Huang, H. Hu, W. Gao, and Y. Pan, "Energy efficiency optimization of cognitive UAV-assisted edge communication for semantic Internet of Things," *Wireless Commun. Mobile Comput.*, vol. 2021, Feb. 2021, Art. no. 6682340.

[232] S. Hakak et al., "Autonomous vehicles in 5G and beyond: A survey," *Veh. Commun.*, vol. 39, Feb. 2023, Art. no. 100551.

[233] S. Ren, S. Chen, and W. Zhang, "Collaborative perception for autonomous driving: Current status and future trend," in *Proc. 5th Chin. Conf. Swarm Intell. Cooperat. Control. (CCSICC)*, Shenzhen, China, Jan. 2022, pp. 682–692.

[234] S. Gyawali, S. Xu, Y. Qian, and R. Q. Hu, "Challenges and solutions for cellular based V2X communications," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 222–255, 4th Quart., 2021.

[235] Y.-C. Liu, J. Tian, C.-Y. Ma, N. Glaser, C.-W. Kuo, and Z. Kira, "Who2com: Collaborative perception via learnable handshake communication," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Paris, France, May 2020, pp. 6876–6883.

[236] Y.-C. Liu, J. Tian, N. Glaser, and Z. Kira, "When2Com: Multi-agent perception via communication graph grouping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4106–4115.

[237] J. Jiang and Z. Lu, "Learning attentional communication for multi-agent cooperation," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, Dec. 2018, pp. 7254–7264.

[238] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun, "V2VNet: Vehicle-to-vehicle communication for joint perception and prediction," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 605–621.

[239] A. Sujith, G. S. Sajja, V. Mahalakshmi, S. Nuhmani, and B. Prasanalakshmi, "Systematic review of smart health monitoring using deep learning and artificial intelligence," *Neurosci. Inform.*, vol. 2, no. 3, Sep. 2022, Art. no. 100028.

[240] P. Sangra, B. Rana, and Y. Singh, "Energy efficiency in IoT-based smart healthcare," in *Proc. 3rd Int. Conf. Comput. Commun. Cyber Security*, Jul. 2022, pp. 503–515.

[241] M. Azimi, A. D. Eslamlou, and G. Pekcan, "Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review," *Sensors*, vol. 20, no. 10, p. 2778, May 2020.

[242] L. Yue, D. Tian, W. Chen, X. Han, and M. Yin, "Deep learning for heterogeneous medical data analysis," *World Wide Web*, vol. 23, pp. 2715–2737, Mar. 2020.

[243] W. Xu, Z. Yang, D. W. Kwan Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, Jan. 2023.

[244] M. Strobel and R. Shokri, "Data privacy and trustworthy machine learning," *IEEE Security Privacy*, vol. 20, no. 5, pp. 44–49, Sep./Oct. 2022.

[245] M. Jegorova et al., "Survey: Leakage and privacy at inference time," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 9090–9108, Jul. 2023, doi: 10.1109/TPAMI.2022.3229593.

[246] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Comput. Surveys*, vol. 54, no. 2, pp. 1–36, Mar. 2021.

[247] C. Dong et al., "Privacy-preserving and byzantine-robust federated learning," *IEEE Trans. Depend. Secure Comput.*, early access, Apr. 5, 2023, doi: 10.1109/TDSC.2023.3264697.

[248] Z. Tian, L. Cui, J. Liang, and S. Yu, "A comprehensive survey on poisoning attacks and countermeasures in machine learning," *ACM Comput. Surveys*, vol. 55, no. 8, pp. 1–35, Dec. 2022.

[249] P. Meena, M. B. Pal, P. K. Jain, and R. Pamula, "6G communication networks: Introduction, vision, challenges, and future directions," *Wireless Pers. Commun.*, vol. 125, pp. 1097–1123, Mar. 2022.

[250] Y. Chen et al. "NetGPT: A native-AI network architecture beyond provisioning personalized generative services." Jul. 2023. [Online]. Available: https://arxiv.org/abs/2307.06148

**Kun Lu** received the B.E. degree from Shanxi University in 2020, and the M.E. degree from Zhejiang University in 2023. He is currently working on large generative models with Huawei Technologies, Ltd., Hangzhou, China. His research interests include machine learning and artificial intelligence.



**Xianfu Chen** (Member, IEEE) received the Ph.D. degree (with Hons.) from Zhejiang University, China, in March 2012. Since April 2012, he has been with the VTT Technical Research Centre of Finland, Oulu, Finland, where he is currently a Senior Scientist. His research interests cover various aspects of wireless communications and networking, with emphasis on human-level and artificial intelligence for resource awareness in next-generation communication networks. He was the recipient of the 2021 IEEE Communications Society Outstanding Paper Award and the 2021 IEEE INTERNET OF THINGS JOURNAL Best Paper Award. He serves as an Editor for IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING as well as *Microwave and Wireless Communications*, an Academic Editor for *Wireless Communications and Mobile Computing*, an Associate Editor for *China Communications*, and served as a member of the First Editorial Board of *Journal of Communications and Information Networks*. He has served as the Guest Editor for several international journals, including *IEEE Wireless Communications Magazine*. He served as a TPC chair/track co-chair/TPC member for a number of IEEE ComSoc Flagship Conferences.



**Zhilin Lu** received the M.S. degree from North China Electric Power University, Beijing, China, in 2021. She is currently pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. Her research interests include semantic communication and deep reinforcement learning.
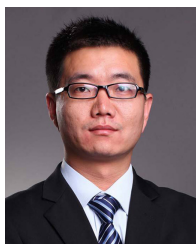


**Rongpeng Li** (Member, IEEE) is currently an Associate Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He was a Research Engineer with the Wireless Communication Laboratory, Huawei Technologies Company Ltd., Shanghai, China, from August 2015 to September 2016. He was a Visiting Scholar with the Department of Computer Science and Technology, University of Cambridge, Cambridge, U.K., from February 2020 to August 2020. His research interest currently focuses on networked intelligence for communications evolving. He received the Wu Wenjun Artificial Intelligence Excellent Youth Award in 2021. He serves as an Editor for *China Communications*.



**Ekram Hossain** (Fellow, IEEE) is a Professor with the Department of Electrical and Computer Engineering, University of Manitoba, Canada (http://home.cc.umanitoba.ca/~hossaina). His current research interests include design, analysis, and optimization of wireless networks with emphasis on next-generation (xG) cellular networks. He received the 2017 IEEE ComSoc TCGCC (Technical Committee on Green Communications & Computing) Distinguished Technical Achievement Recognition Award "for outstanding technical leadership and achievement in green wireless communications and networking". He was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science from 2017 to 2022. He is a member (Class of 2016) of the College of the Royal Society of Canada, and a Fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada. He was elevated to an IEEE Fellow "for contributions to spectrum management and resource allocation in cognitive and cellular radio networks."

**Zhifeng Zhao** (Member, IEEE) received the B.E. degree in computer science, the M.E. degree in communication and information systems, and the Ph.D. degree in communication and information systems from the PLA University of Science and Technology, Nanjing, China, in 1996, 1999, and 2002, respectively. From 2002 to 2004, he acted as a Postdoctoral Researcher with Zhejiang University, Hangzhou, China, where his researches were focused on multimedia next-generation networks and softswitch technology for energy efficiency. From 2005 to 2006, he acted as a Senior Researcher with the PLA University of Science and Technology, where he performed research and development on advanced energy-efficient wireless router, ad-hoc network simulator, and cognitive mesh networking test-bed. From 2006 to 2019, he was an Associate Professor with the College of Information Science and Electronic Engineering, Zhejiang University. He is currently with Zhejiang Lab, Hangzhou, as the Chief Engineering Officer. His research areas include software-defined networks, wireless network in 6G, computing networks, and collective intelligence. He is the Symposium Co-Chair of ChinaCom 2009 and 2010. He was the Technical Program Committee Co-Chair of the 10th IEEE International Symposium on Communication and Information Technology in 2010.

**Honggang Zhang** (Fellow, IEEE) is the Chief Managing Editor of Intelligent Computing, a Science Partner Journal, as well as a Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He was an Honorary Visiting Professor with the University of York, York, U.K., and an International Chair Professor of Excellence with the Université Européenne de Bretagne and Supélec, France. He has coauthored and edited two books: *Cognitive Communications: Distributed Artificial Intelligence (DAI), Regulatory Policy & Economics, Implementation* (John Wiley & Sons) and *Green Communications: Theoretical Fundamentals, Algorithms and Applications* (CRC Press), respectively. His research interests include cognitive radio and networks, green communications, mobile computing, machine learning, artificial intelligence, and Internet of Intelligence. He is a co-recipient of the 2021 IEEE Communications Society Outstanding Paper Award and the 2021 IEEE INTERNET OF THINGS JOURNAL Best Paper Award. He was the leading Guest Editor for the Special Issues on Green Communications of the *IEEE Communications Magazine*. He served as a Series Editor for the *IEEE Communications Magazine* (Green Communications and Computing Networks Series) from 2015 to 2018 and the Chair of the Technical Committee on Cognitive Networks of the IEEE Communications Society from 2011 to 2012. He is the Associate Editor-in-Chief of *China Communications*.