Semantics-Empowered Space-Air-Ground-Sea Integrated Network: New Paradigm, Frameworks, and Challenges

Siqi Meng, Shaohua Wu, *Member*, *IEEE*, Jiaming Zhang, Junlan Cheng, Haibo Zhou, *Senior Member*, *IEEE*, and Qinyu Zhang, *Senior Member*, *IEEE*

***Abstract*—In the coming sixth generation (6G) communication era, to provide seamless and ubiquitous connections, the space-air-ground-sea integrated network (SAGSIN) is envisioned to address the challenges of communication coverage in areas with difficult conditions, such as the forest, desert, and sea. Considering the fundamental limitations of the SAGSIN including large-scale scenarios, highly dynamic channels, and limited device capabilities, traditional communications based on Shannon information theory cannot satisfy the communication demands. Moreover, bit-level reconstruction is usually redundant for many human-to-machine or machine-to-machine applications in the SAGSIN. Therefore, it is imperative to consider high-level communications towards semantics exchange, called semantic communications. In this survey, according to the interpretations of the term "semantics", including "significance", "meaning", and "effectiveness-related information", we review state-of-the-art works on semantic communications from three perspectives, which are 1) significance representation and protection, 2) meaning similarity measurement and meaning enhancement, and 3) ultimate effectiveness and effectiveness yielding. Sequentially, three types of semantic communication systems can be correspondingly introduced, namely the significance-oriented, meaning-oriented, and effectiveness/task-oriented semantic communication systems. Implementation of the above three types of systems in the SAGSIN necessitates a new perception-communication-computing-actuation-integrated paradigm (PCCAIP), where all the available perception, computing, and actuation techniques jointly facilitates significance-oriented sampling & transmission, semantic extraction & reconstruction, and task decision. Finally, we point out some future challenges on semantic communications in the SAGSIN. This survey provides a comprehensive review on the future semantic communications in the SAGSIN, and elaborates on the performance metrics and techniques related to semantic communications for references and further in-depth investigations.**

***Index Terms*—Space-air-ground-sea integrated network (SAGSIN), semantic communications, perception-communication-computing-actuation integrated paradigm (PCCAIP), data significance, task-oriented communications, deep learning (DL).**

I. Introduction

A. Motivation

The sixth generation (6G) network has been envisioned to provide seamless and ubiquitous services in an extremely reliable and energy-efficient manner with ultra low latency [1]–[3]. However, the current widely used terrestrial network in fifth generation (5G) is facing severe challenges in facilitating such seamless and ubiquitous connections due to the following reasons. Firstly, with the proliferating number of connected devices and the surge of data traffic, the terrestrial network alone cannot satisfy such massive and broad connections. Secondly, almost 90 percent of the Earth area is covered by the sea, forest, and desert, where the terrestrial network cannot provide continuous connection for its harsh environment. To tackle the above inherent challenges in the current 5G network, terrestrial networks should be extended by integrating the nodes deployed in the space, air, and (or) sea to form a space-air-ground(-sea) integrated network (SAGIN or SAGSIN), which has garnered a surge of interest by researchers [4]–[11].

The SAGSIN is comprised of space-based network, air-based network, and sea-based network along with ground-based (terrestrial) network, as shown in Fig. 1. Specifically, space-based network consists of geostationary Earth orbit (GEO) satellites and low Earth orbit (LEO) satellites, which usually form constellations respectively; air-based network is composed of airplanes, unmanned aerial vehicles (UAVs), and high-altitude platforms (HAPs), which usually appear as platoons; sea-based network includes nodes such as maritime base stations, ships, and unmanned ships; ground-based network has been extensively developed with the advent of cell mobile communications, autonomous driving, and Internet of Things (IoT). By connecting the nodes within a certain part via short or long wireless links to construct the abovementioned four types of networks, and integrating the four parts via inter-network links, the SAGSIN can perfectly reap the promised benefits of 6G communications owing to its massive connections among long-distance nodes.

However, the realization of 6G communications with extremely high reliability, very low latency, and high energy efficiency based on the SAGSIN will be affected by several fundamental limitations, which are *large-scale scenarios, highly dynamic channels, and limited device capabilities.* Firstly, the communication distance between the transceiver nodes in the SAGSIN scenarios is considerably large in stark comparison

S. Meng, S. Wu, J. Zhang, J. Cheng, and Q. Zhang are with the Department of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: mengsiqi@stu.hit.edu.cn; hitwush@hit.edu.cn; hitzhangjiaming@163.com; hitsz.chengjunlan@foxmail.com; zqy@hit.edu.cn).

Haibo Zhou is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: haibozhou@nju.edu.cn).

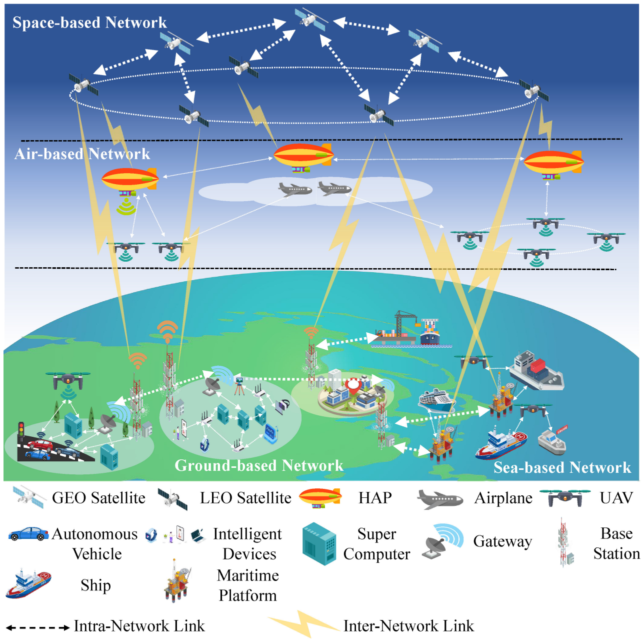


Fig. 1. An Illustration of the SAGSIN.

with terrestrial communications. For instance, the distance from LEO to ground base station is 500-2,000 kilometers, which causes inherent and non-negligible propagation latency. Secondly, the communication channels over transceivers in the SAGSIN vary rapidly compared with terrestrial channels. Taking UAV-LEO-ground multi-hop communications as an example, the LEO orbits around the Earth with significantly high velocity, leading to inherent Doppler shift in UAV-LEO and LEO-ground channels; besides, cosmic environment is so harsh that deep fading will appear due to cosmic ray and atmospheric attenuation. Thirdly, the available communication and computing resources are strictly limited for node devices in the SAGSIN. Specifically, due to hardware and battery limitation on the satellites, UAVs, and so on, transmission rate and computing speed of space/air-ground communications are remarkably lower in comparison to terrestrial communications.

The abovementioned inherent and fundamental characteristics of the SAGSIN impose severe challenge on current traditional communication technologies in facilitating the envisioned 6G communications. On the other hand, traditional communications based on classic Shannon information theory are reaching their limits, which may not support communications in the SAGSIN with such severe limitations. Concretely speaking, since Shannon completed his masterpiece which lays the foundation of information theory in 1948 [12], researchers have contributed much effort to find state-of-the-art coding and modulation schemes that can reach the channel capacity. From near Shannon limit codes such as Turbo codes [13] and low density parity check (LDPC) codes [14], to capacity-achievable channel codes such as Polar codes [15] and Spinal codes [16], the Shannon limit has been reached approximately by these advanced channel coding techniques. From frequency division multiple access (FDMA) applied in the first generation (1G) communications, to non-orthogonal multiple access (NOMA) emerging in 5G communications, the revolution of modulation methods has also enhanced the channel capacity and been reaching the Shannon limits. Therefore, a new revolution on current communication technologies is imperative to satisfy the proliferating need on data exchange in the SAGSIN.

Moreover, in many human-to-machine and machine-to-machine application scenarios in the SAGSIN, the ultimate goal of communications is to complete specific tasks at the machine terminal. Thus, perfect or approximate reconstruction of information on bit level is usually unnecessary. For instance, in the ship detection tasks, the observer can detect the ships based on only the pixels of image related to ships, where reconstruction of the whole image is evidently redundant. Another example is emergency monitoring and rescue tasks, where monitors are expected to only perceive and transmit urgent status information, while missed detection or error transmission of other statuses that are not very urgent are tolerable. Under harsh conditions of the SAGSIN, pursuing perfect or approximate restoration of raw data not only will consume a huge amount of resources but is also not feasible due to limited device capabilities.

Therefore, the above reasons necessitate an idea change from traditional communication technologies to high-level communications towards semantics exchange, named semantic communications [17], for semantic communications aim at recovering only semantics at the receiver instead of reconstructing every code symbol accurately. In [17], Weaver identifies the three levels of communications, which are technical communications, semantic communications, and effectiveness communications. The abovementioned state-of-the-art coding and modulation schemes all make remarkable contributions to technical communications, while deliberately neglecting the semantic and effectiveness levels of communications, since Shannon believes that the technology of symbol transmission can be independent from the semantics that the symbols hold. However, the idea change from technical communications to semantic communications inspires us to reconsider the assumption of semantic independence and take semantic aspects of communications into consideration. This serves as the principal line of this survey.[1]

This survey is dedicated to answering the following three questions regarding to semantic communication systems deployed in the SAGSIN.

Question 1. How should we interpret ”semantics” for semantic communications?

Question 2. How should we design semantic communication systems based on the interpretations of “semantics”?

Question 3. How should we utilize the available while limited resources in the SAGSIN to implement semantic communications?

Overall, the term “semantics” has three interpretations in state-of-the-art research on semantic communications. Firstly, from the perspective of etymology, the word “semantics” originates from the Greek “sēmanticós”, which means “signif-

[1] It is worth noting that in this survey, we consider both semantic level and effectiveness level communications as semantic communications.

icance". Secondly, "semantics" has its common explanation, which is "meaning". Thirdly, since "semantic communications" includes both semantic and effectiveness level, this term can also be interpreted as "effectiveness-related information".

Based on above, we introduce three types of semantic communication systems, which are respectively significance-oriented, meaning-oriented, and effectiveness-oriented semantic communication systems.

(1) *Significance-oriented semantic communication system.* In fact, source data are not equally significant, and thus sampling and transmitting unimportant data will cause waste in resource consumption. Taking the emergency monitoring task as an example, the status message generated later is more significant than the one generated earlier, since the former contains fresher information for the receiver; moreover, the status message which is different from the previous one is more significant, since the status variation usually implies policy or decision change for the task actuator; besides, an abnormal status message is more significant than a normal one, since the environment becomes more urgent due to potential severe consequences. Therefore, from the significance perspective, source data should be sampled, coded, and modulated in a significance-oriented manner, such that only data with more semantics (i.e., more significant data) are reserved and transmitted while those with less semantics (i.e., less significant data) get discarded. How we can measure the "significance" will be elaborated in Section III.

(2) *Meaning-oriented semantic communication system.* After significance-oriented sampling, although the amount of data has been preliminarily reduced, the remaining sampled data still include considerably large amount of redundancy due to semantics-unrelated information. Therefore, semantic extraction from the sampled data should be further adopted at the transmitter in order to eliminate the meaning-unrelated redundancy, such that the receiver can reconstruct data with similar meaning to original data. In the era of Shannon, semantic extraction and reconstruction were tough problems, and researchers had to temporarily neglect the semantic aspects of source data and dedicate themselves to technical communication revolution. Nowadays, the thriving of artificial intelligence (AI), especially the unprecedented proliferation of deep learning (DL) applications, will enable deep feature (semantics) extraction by massive computing resources, and utilizing DL to realize semantic communication systems is becoming a major approach. How DL-based semantic communication systems are designed will be discussed in Section IV.

(3) *Effectiveness-oriented semantic communication system.* After extraction of meaning-related information from sampled data, most of the redundancy is eliminated with only the meaning of data to be transmitted. However, not all of the extracted meaning is useful for the receiver, since there is still redundant information for the ultimate actuation. Moreover, symbol- or semantics-level reconstruction is also unnecessary from effectiveness-oriented perspective, since task actuation process may be intelligently conducted based directly on received messages. Therefore, a task-oriented design for semantic communications can be adopted in order that the sampled data yield ultimate effectiveness more efficiently at the terminal. Specifically, a more intelligent effectiveness-related semantic extractor is introduced to replace common semantic extractor for capturing effectiveness-related information. Moreover, taking full advantage of DL-based techniques, intelligent actuator is adopted which receives the effectiveness-related information and directly outputs the actuation results of the task without reconstruction of original data. By such task-oriented design, the reconstruction process at the receiver gets omitted and thus computing resource consumption will decline. How such task-oriented communication system is constructed can be found in Section V.

Furthermore, in order to implement semantic communication systems in the SAGSIN from the above three perspectives, all the available perception, computing, and actuation techniques should be jointly adopted. Specifically, significance-oriented semantic communication system adopts a perception-communication-integrated design by filtering insignificant information and protecting significant one; meaning-oriented semantic communication system utilizes perception, communication, and computing techniques through DL-based meaning extraction and reconstruction; effectiveness-oriented semantic communication system integrates all the perception, communication, computing, and actuation techniques via effectiveness-related semantic extractor and intelligent actuator without reconstruction. Therefore, a new paradigm called *Perception-Communication-Computing-Actuation Integrated Paradigm (PCCAIP)* can be proposed that guides the design of the above three types of semantic communication systems in the SAGSIN.

B. Related Works

Given that there is a recent surge of research interest on semantic communications, quite a few review and tutorial works on this topic have also emerged in the literature. In the aspects of short briefs, [18] envisions a significance-oriented semantic communication system with sparse and effectiveness-aware sampling, where semantics of data is measured by timeliness metrics and end-to-end mean square error (MSE). Another brief review [19] constitutes a tutorial on the framework of semantic communications by reconsidering semantic information theory. From semantic network perspective, [20] proposes a semantic-aware network architecture by utilizing federated edge learning. Moreover, [21] presents a comprehensive review on DL-based semantic communications. Nevertheless, all these brief reviews on semantic communications focus on only a certain interpretation of "semantics" listed in the previous subsection.

Researchers have also contributed comprehensive tutorial works on semantic communication topics. The first of such tutorial works is [22], where three communication modalities including human-human, human-machine, and machine-machine are addressed by semantic communications using DL-based techniques. In [23], semantic communication theory is reviewed, and DL-based semantic communications for different types of sources are respectively discussed. In [24], a new perspective called reasoning-driven semantic communication system is proposed, and a new semantic representation

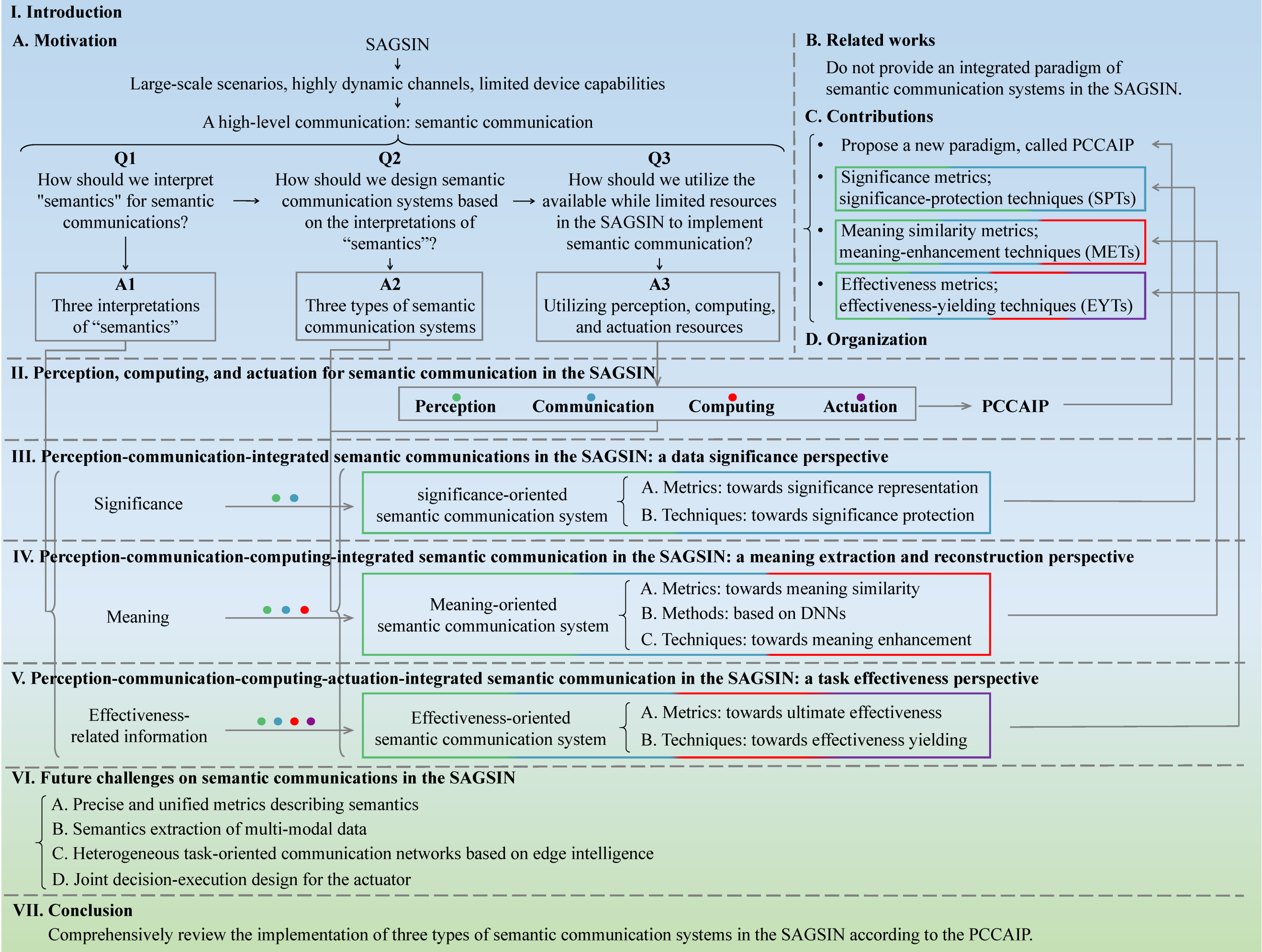

I. Introduction
A. Motivation
SAGSIN
Large-scale scenarios, highly dynamic channels, limited device capabilities
A high-level communication: semantic communication
Q1
How should we interpret "semantics" for semantic communications?
Q2
How should we design semantic communication systems based on the interpretations of "semantics"?
Q3
How should we utilize the available while limited resources in the SAGSIN to implement semantic communication?
A1
Three interpretations of "semantics"
A2
Three types of semantic communication systems
A3
Utilizing perception, computing, and actuation resources
B. Related works
Do not provide an integrated paradigm of semantic communication systems in the SAGSIN.
C. Contributions
Propose a new paradigm, called PCCAIP
Significance metrics; significance-protection techniques (SPTs)
Meaning similarity metrics; meaning-enhancement techniques (METs)
Effectiveness metrics; effectiveness-yielding techniques (EYTs)
D. Organization
II. Perception, computing, and actuation for semantic communication in the SAGSIN
Perception
Communication
Computing
Actuation
PCCAIP
III. Perception-communication-integrated semantic communications in the SAGSIN: a data significance perspective
Significance
significance-oriented semantic communication system
A. Metrics: towards significance representation
B. Techniques: towards significance protection
IV. Perception-communication-computing-integrated semantic communication in the SAGSIN: a meaning extraction and reconstruction perspective
Meaning
Meaning-oriented semantic communication system
A. Metrics: towards meaning similarity
B. Methods: based on DNNs
C. Techniques: towards meaning enhancement
V. Perception-communication-computing-actuation-integrated semantic communication in the SAGSIN: a task effectiveness perspective
Effectiveness-related information
Effectiveness-oriented semantic communication system
A. Metrics: towards ultimate effectiveness
B. Techniques: towards effectiveness yielding
VI. Future challenges on semantic communications in the SAGSIN
A. Precise and unified metrics describing semantics
B. Semantics extraction of multi-modal data
C. Heterogeneous task-oriented communication networks based on edge intelligence
D. Joint decision-execution design for the actuator
VII. Conclusion
Comprehensively review the implementation of three types of semantic communication systems in the SAGSIN according to the PCCAIP.

Fig. 2. The architecture of this survey.

method called semantic language is studied for next generation communications. Moreover, in the recent literature of [25], the authors review semantic communications from perspectives of semantic measures, semantic compression based on semantic rate-distortion theory, semantic transmission via joint source-channel coding (JSCC), and timeliness of semantic communications, respectively. In the recent tutorial work [26], research on semantic communications is reviewed from three perspectives, which are semantic-oriented communications, goal-oriented communications, and semantic-awareness communications, and correspondingly three directions on semantic communications are pointed out, which are semantic extraction, semantic transmission, and semantic metrics. In addition to the above tutorial works, some other tutorials on edge learning [27] and new challenges of 6G [28] also consider semantic communications as competitive solutions. However, these studies does not provide an integrated paradigm of semantic communication system design that satisfies the needs in the SAGSIN.

C. Contributions

To fill the void in literature of semantic communications for the SAGSIN, in this survey, we propose a comprehensive paradigm called PCCAIP which fully addresses fundamental limitations in the SAGSIN. Sequentially, we review recent literature on semantic communications from three perspectives, which are 1) significance representation and protection, 2) meaning similarity measurement and meaning enhancement, and 3) ultimate effectiveness and effectiveness yielding. Specifically, our contributions are listed as follows.

- We propose a new paradigm for the implementation of semantic communications in the SAGSIN called PCCAIP, in which perception, computing, and actuation techniques will be jointly utilized for communications, and thus the design for the three types of state-of-the-art semantic communication systems can be guided by PCCAIP.
- Aiming at designing significance-oriented semantic communication systems, we introduce the significance metrics, including content-agnostic, content-aware, and unified metrics. We then review recent works on *Significance Protection Techniques (SPTs)*, including sampling, coding, and modulation.
- We elaborate on a major approach of semantic communications, namely meaning-oriented semantic communication system, focusing on implementation of DL-based meaning extraction and reconstruction. We provide an

idea of system design, which starts from maximizing meaning similarity metrics, and next trains DL models, and finally utilizes *Meaning-Enhancement Techniques (METs)*.

- We discuss a newly emerging approach of semantic communications called effectiveness/task-oriented semantic communication system, which is guided by PCCAIP. We review metrics that measures the ultimate effectiveness of tasks, and discuss *Effectiveness-Yielding Techniques (EYTs)* facilitating typical services in the SAGSIN.

D. Organization

The rest of this survey is organized as follows. Section II propose the PCCAIP, where utilization of perception, computing, and actuation modules in the SAGSIN is elaborated respectively. In Section III, we introduce significance metrics as semantic measures, and review SPTs including significance-oriented sampling, coding, and modulation. Section IV introduce meaning similarity metrics, DL models, and METs based on meaning extraction and reconstruction. The focus of Section V is task-oriented communications guided by PCCAIP, where effectiveness metrics are reviewed and EYTs are discussed for specific tasks. We then envision some future challenges on semantic communication implementation in the SAGSIN in Section VI, and VII concludes the survey. The architecture of this survey is illustrated in Fig. 2.

II. Perception, Computing, and Actuation for Semantic communications in the SAGSIN

As mentioned in Section I, semantic communication systems play a crucial role in the SAGSIN by enabling efficient information exchange and intelligent actuation. However, relying solely on traditional communication techniques may not meet the requirements of complex tasks due to the limitations in network resources. Therefore, perception, computing and actuation techniques facilitate semantic communications to form the PCCAIP. In the following discussion, we will firstly address these techniques, and then provide a detailed introduction to the PCCAIP.

A. Perception

The integration of space-based, air-based, ground-based, and sea-based networks results in a significant increase in business volume. Multiple types of sensors deployed in these networks enable real-time perception of targets while generating a substantial amount of multimodal data, which pose a great challenge in supporting services and applications with extremely low latency requirements. One feasible solution is to reduce the volume of data to decrease the latency in subsequent communication and processing stages. This is why we need to perceive the significance, meaning, and effectiveness-related information of the source data in the SAGSIN. Capturing significance can be achieved through significance-oriented sampling, while capturing the other two can be accomplished through intelligent semantic extraction, which is typically facilitated by DL, knowledge graphs (KGs), and so on. The significance-oriented sampling and semantic extraction are collectively referred to as perception in this survey.

Perception plays a significant role in various application scenarios. For example, in the emergency response and rescue applications for maritime incidents, ocean buoys deployed in the maritime area are used to collect oceanographic source data such as sea currents, wind speeds, and wave conditions. UAVs deployed above the sea surface capture image data with cameras. During non-urgent situations, the sensors operate at a relatively low sampling frequency to conserve energy and reduce data processing burden. However, in the emergency events such as maritime shipwrecks, oil spills, fires, or drowning, the sensors immediately increase their sampling frequency to keep data freshness and let the server estimate the source status more accurately. The sampled data are further processed to extract semantic information, such as the situation of stranded crew members and ocean current information, which is then transmitted to the emergency rescue center in real-time to facilitate rescue operations.

B. Computing

In the semantic-empowered SAGSIN, computing refers to the AI-related complex operations, such as the feature extraction, semantic coding, and adaptive decision. Therefore, it is imperative to use AI algorithms to extract semantic information and make intelligent decisions, which will release data traffic and thus decrease processing latency. However, AI algorithms rely heavily on sufficient computing resources, and it is unlikely to achieve satisfactory performance when these algorithms are deployed on end devices such as ocean buoys and UAVs due to limitations in computing and caching resources. The SAGSIN encompasses various types of devices, resulting in a large volume of data that usually requires real-time processing and analysis. Edge computing technology has emerged to address this issue.

In recent years, there have been significant academic and industrial interests in on-orbit edge computing based on space-based networks [29], [30], as well as edge storage technologies [31], [32]. A large number of cloud servers, edge servers, and other computing units are distributed in the SAGSIN, providing supports for perception, communication, and actuation in various stages. For instance, in the context of satellite-based edge computing, efficient interconnection is achieved among space-based, and ground-based networks by leveraging LEO satellites as the core. An edge computing platform is built on the satellites, empowering the entire network with on-orbit capabilities for intelligent data collection, processing, and caching [33]. The authors assume a geographically dispersed remote team with diverse members. Through augmented reality (AR) and virtual reality (VR) technologies, remote team members can share a virtual workspace, enabling collaborative working and real-time communications. In this scenario, the on-orbit edge servers are equipped with a semantic encoding network driven by DL models. Team members wear head-mounted displays (HMD) integrated with semantic decoding networks, accessing the shared virtual workspace through VR

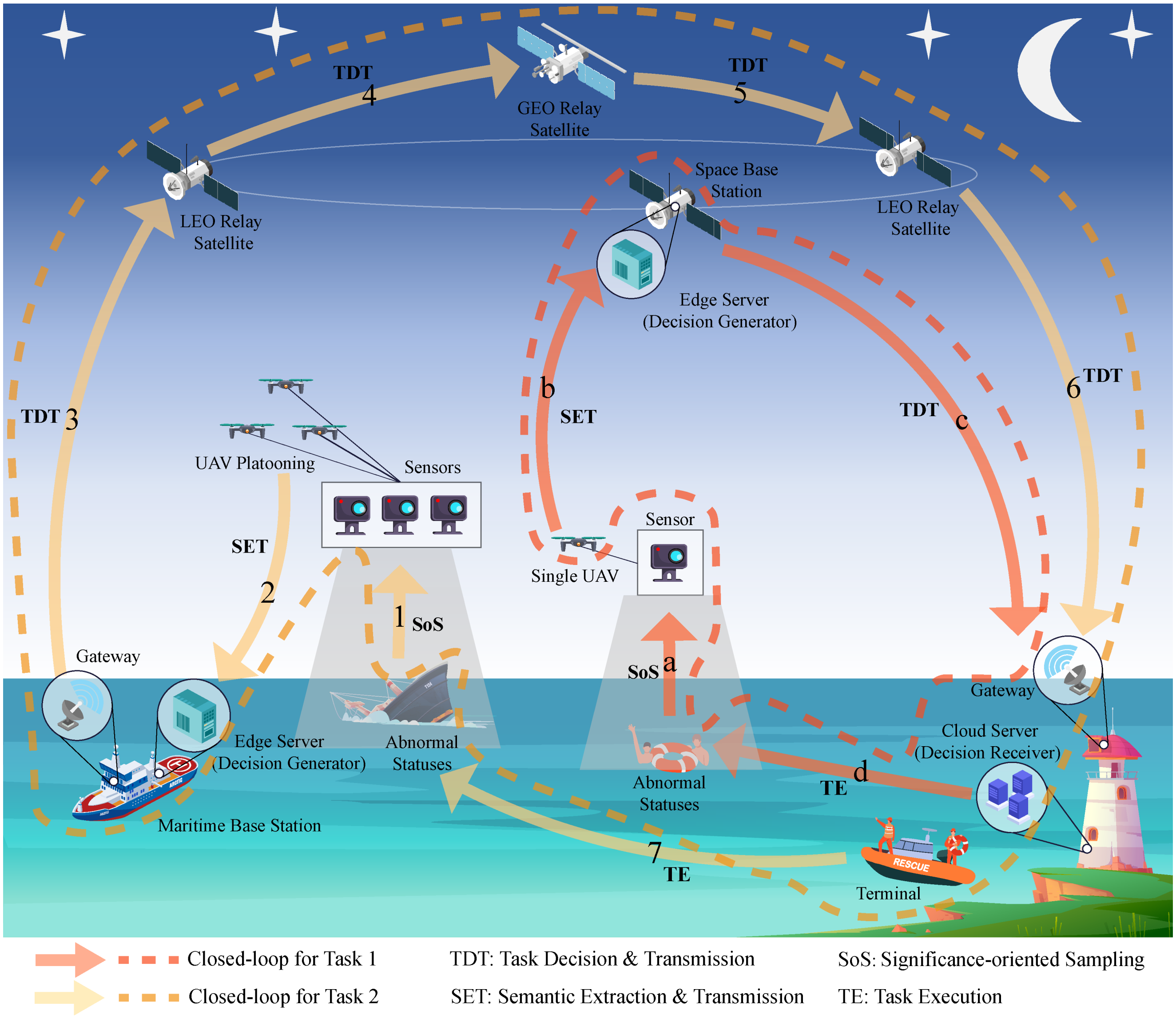


Fig. 3. A real-world scenario of PCCAIP in the SAGSIN. The numbers and letters represent the closed-loop information flows. For task 1, the letters represent the following processes: a. significance-oriented sampling of abnormal statuses from a drowning person; b. semantic extraction and transmission from a single UAV to the space base station; c. task decision transmission from the base station to the cloud server; d. task execution (the terminal will send rescue to the drowning person). For task 2, the numbers represent the following processes: 1. significance-oriented sampling of abnormal statuses from a sinking ship; 2. semantic extraction and transmission from UAV platooning to the maritime base station; 3. task decision transmission from the base station to an LEO satellite; 4. relay from an LEO satellite to the GEO satellite; 5. relay from the GEO satellite to another LEO satellite; 6. task decision transmission from the LEO satellite to the cloud server; 7. task execution (the terminal will send rescue to the sinking ship).

technology. In this example, the satellite-based edge computing platform provides robust support for the combination of VR technology and semantic communications.

C. Actuation

The receiver utilizes the received messages, which are relevant to task performance, to accomplish the task decision. Based on the decision results, the task execution process will exert an influence on the physical world. In this survey, the task decision and task execution comprise the definition of actuation. Taking remote control services as example, the intelligent decision-maker generates and transmits task decision commands based on the received messages, and then the terminal receives the decision commands and executes specific tasks.

With the development of 5G/6G, the variety of task-critical applications in the SAGSIN continues to enrich. For instance, the stability control of UAVs can be modeled as a classical CartPole problem solution. In this scenario, sensors collects environmental data, which will be then processed on the edge server using DL-based autoencoders to extract task-related semantic features. These features are transmitted to the remote control center on the ground to generate control commands, enabling remote control of the UAVs [34].

Another example is found in remote surgeries, which is a typical use case of haptic Internet. Sensors located in remote areas (e.g., mountainous areas and remote seas) collect crucial patient data, which are transmitted to the medical center with the aid of satellites. Surgeons receive real-time auditory, visual, and haptic feedback through a human-machine interface (HMI). Visual feedback is provided using streaming technologies, such as holographic-type communications (HTC), depending on whether surgeons interact with holograms or wear head-mounted devices. Subsequently, surgeons operate haptic devices through the HMI and perform surgical actions based on real-time visual feedback and haptic information transmitted to the robot [35].

D. PCCAIP

According to the interpretation of perception, computation, actuation, we introduce the perception-communication-computing-actuation-integrated framework under the guidance of PCCAIP. This framework aims to maximize the effectiveness of tasks by establishing closed information flow loop, as shown in Fig. 3. Specifically, we consider an emergency detection and rescue scenario in the SAGSIN as an example. Various sensors (e.g., cameras deployed on the UAVs) continuously observe the environment that contains abnormal statuses (e.g., sinking ships or drowning persons), and decide to sample data representing the current statuses when the data are significant enough (e.g. untimeliness, status change, or environment change). After data sampling, these data will be further processed by the computing-intensive semantic extractor deployed on UAVs and then transmitted to remote edge servers (e.g. servers deployed on maritime base station or space base station), by which multi-level data perceptions are completed by distributed computing resources.

Based on the received status data, the edge servers conduct real-time decision of whether or not the terminal (e.g., rescue ship) rescue the emergencies and which emergencies are to be rescued (if rescue abilities are limited), and then transmit the decision commands to remote terrestrial cloud server via LEO/GEO satellite relays. After the cloud server received the decision commands, it instantly conduct task execution (e.g., rescuing the ships/persons, or doing nothing) based on the commands, by which task actuation is completed. Task execution will further determine the statuses and thus the decisions at the next time interval (for example, if a sinking ship gets rescued by the rescue ship, there will be no emergency in the next period of time, and thus the rescue ship does not need to rescue in the next), by which closed-loops for the emergency rescue tasks are formed by the PCCAIP framework.

E. Lessons Learned from This Section

This section elaborates on how the perception, computing, and actuation techniques help realize semantic communications. Overall, perception techniques help select the data with more semantics, computing techniques facilitate intelligent communications in various aspects, and actuation techniques aid in the interaction between system and environment to complete all-inclusive tasks. PCCAIP integrates all these techniques which can realize a closed-loop communication and a joint design to achieve high effectiveness of tasks. Moreover, maritime rescue tasks are adopted as an example to showcase how PCCAIP guides semantic communication system design.

As typical semantic communication frameworks facilitated by PCCAIP, we will introduce significance-oriented, meaning-oriented, and effectiveness/task-oriented communications systems respectively in the following sections. In Section III, we design a significance-oriented semantic communication system which belongs to a perception-communication-integrated semantic communications. In Section IV, we study a meaning-oriented semantic communication system which falls under the category of a perception-communication-computing-integrated semantic communication system. In Section V, we introduce a task-oriented communication system which achieves perception-communication-computing-actuation integration.

III. Perception-Communication-Integrated Semantic Communications in the SAGSIN: A Data Significance Perspective

In this section, we will review the research on semantic communications from the data significance perspective. Here, the term “semantics” is explained as its etymological meaning “significance”, which implies a significance-oriented semantic communication system.

In significance-oriented semantic communications in the SAGSIN, a joint perception-communication framework is adopted as shown in Fig. 4. Specifically, source data are firstly sampled by significance-oriented sampler in order to release the heave data burden, and the sampled data will be coded and modulated for transmission over wireless channel. The receiver demodulates and decodes the received message to recover sampled data. According to the recovery results, the significance of source data varies to affect the evolution of significance metric; sequentially, the variation of significance metric will affect the further decision of sampler and coder/modulator at the transmitter via feedback from the receiver. It is worth noting that both coding and modulation techniques mentioned in this section are traditional technologies based on Shannon information theory.

Considering the fundamental limits of the SAGSIN and the time-sensitivity of various tasks facilitated by SAGSIN, the data freshness/timeliness is naturally significant for semantic communication system due to large communication distance. Therefore, timeliness metrics, including AoI and its variants, are of vital importance in describing the performance of the tasks. Thus, we will first introduce AoI and AoI variants as significance metrics in subsection III-A, and then review the literature on techniques towards significance-protection in subsection III-B.

A. Metrics: Towards Significance Representation

We start our review on significance metrics by AoI, and then we introduce its variants, including nonlinear AoI, age of synchronization (AoS), age of incorrect information (AoII), urgency of information (UoI) and the recently proposed goal-oriented tensor (GoT), to describe multi-dimensional significance (e.g., timeliness, status synchronization cost, and

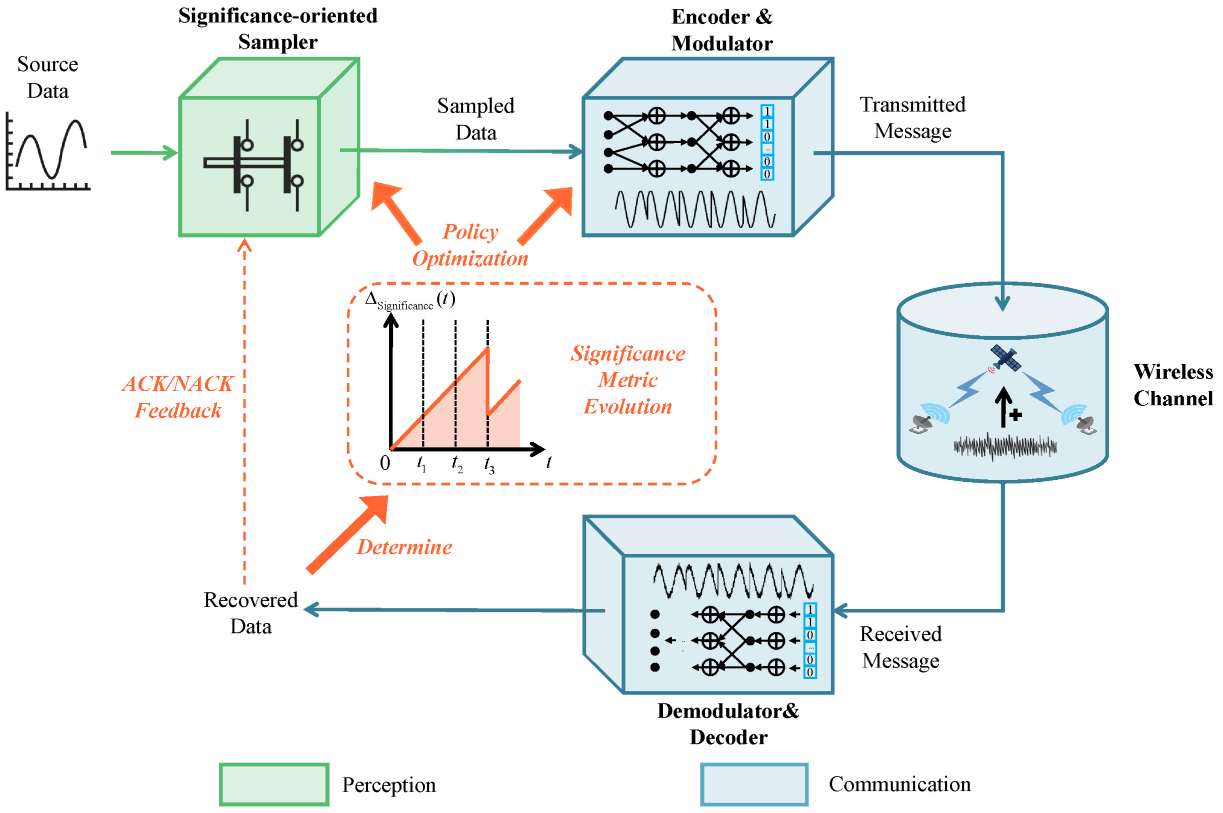


Fig. 4. The framework of perception-communication-integrated semantic communications in the SAGSIN.

environment variation cost). The characteristics of significance metrics reviewed in this survey are listed in Table I.

AoI is firstly proposed in [36], where the authors use this metric to find the optimal information transmission rate in vehicular networks. In [37], the authors further use AoI to evaluate the performance of status update systems. Specifically, AoI is defined as the time elapsed since the latest successfully received status update was generated. The instantaneous AoI expression for the system is

$$\Delta_{\text{AoI}}(t) = t - U(t), \tag{1}$$

where t represents the current time and $U(t)$ represents the generation time of the currently received information. The evolution of AoI is rather simple, that is, only when a status update is successfully received does AoI update as the age of the newly received update, and otherwise AoI continuously increases in a linear manner.

In the SAGSIN, status update messages are exchanged among UAVs, satellites, ships, and ground base stations through wireless channels. High AoI implies long delay and frequent errors of the status update transmission, thereby affecting the significance (or timeliness) of data. Usually, the average AoI or peak AoI is used to measure long-term significance of data. Some research also focuses on the AoI at specific query instants to propose query age of information (QAoI) [44] as an AoI variant.

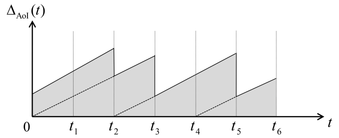
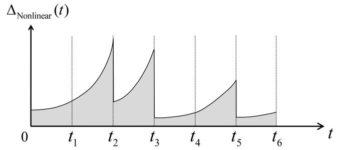
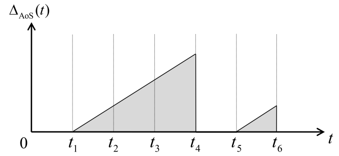
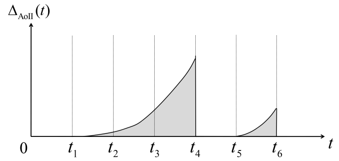
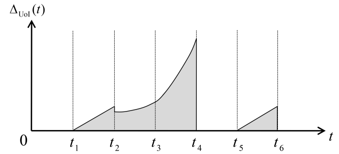
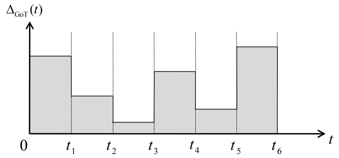
However, the linearly increasing cost of AoI is insufficient in describing the nonlinear significance variation incurred by untimely information. For applications that are more sensitive to time, such as autonomous driving and remote surgery, the performance degradation caused by information aging may not be a linear function of time. For instance, in the state estimation problem of Gaussian linear time-invariant (LTI) systems, if the system is stable, the state estimation error (which directly determine the significance of source data) is a sublinear function of Δ_{AoI}; if the system is unstable, the state estimation error increases exponentially with Δ_{AoI} [45]. Therefore, it is necessary to evaluate the significance of data which is varying nonlinearly in age. Some scholars have proposed value of information (VoI) [38] (also called nonlinear AoI [39]). Nonlinear AoI is a nonlinear function of AoI, that is

$$\Delta_{\text{nonlinearAoI}}(t) = f(t - U(t)), \tag{2}$$

where $f(\cdot)$ is a nonlinear penalty function for linear information age. For example, if the penalty function is chosen as exponential function, a possible evolution of nonlinear AoI is shown in Table I. The choice of penalty function is dependent on how the significance varies in data aging. If the significance increases slowly in time, we adopt functions with decreasing derivatives in time, such as logarithmic function; in contrast, if the significance increases rapidly, we adopt functions with increasing derivatives in time, such as exponential function.

However, both AoI and nonlinear AoI only reflect the time elapsed from the generation of the current information to its

TABLE I
SUMMARY OF SIGNIFICANCE METRICS

Metrics	**Significance Aspect**	**Properties**	**Evolution Example**	**Formula**	**Initial References**
AoI	The staleness of source data	Content-agnostic		(1)	[36], [37]
Nonlinear AoI	The nonlinearly increasing cost caused by staleness of source data	Content-agnostic		(2)	[38], [39]
AoS	The cost caused by asynchronization duration between source data and received data	Content-aware		(3)	[40]
AoII	The (non)linear cost caused by asynchronization (or wrong transmission) duration	Content-aware		(4)	[41]
UoI	The time-variant cost caused by asynchronization, which depends on the current external environment	Content-aware		(5)	[42]
GoT	The customized cost which depends on the environment, the distance between source and received data, and external environment	Unified		(6)	[43]

successful reception, without considering whether the receiver correctly estimates the information from the source. That is, they only focus on the timeliness of the currently transmitting information and ignore the real-time status synchronization between the recovered information and the current source status information. This characteristic of AoI (and nonlinear AoI) is also called content-agnostic. However, synchronization of statuses between transceivers also affects the significance of data, since wrong estimation will lead to biased decision on current status and cause severe consequences. For instance, missed detection of the fire will cause considerably large financial loss due to untimely rescue. Therefore, to reflect whether the receiver fully understands the source information, researchers have proposed a series of content-aware significance metrics including AoS, AoII, and UoI.

In [40] the authors define a novel metric called AoS, which extends the freshness of information to the synchronization duration of information. Unlike AoI, AoS is calculated as the duration since the latest time of perfect inference of the current process status, that is

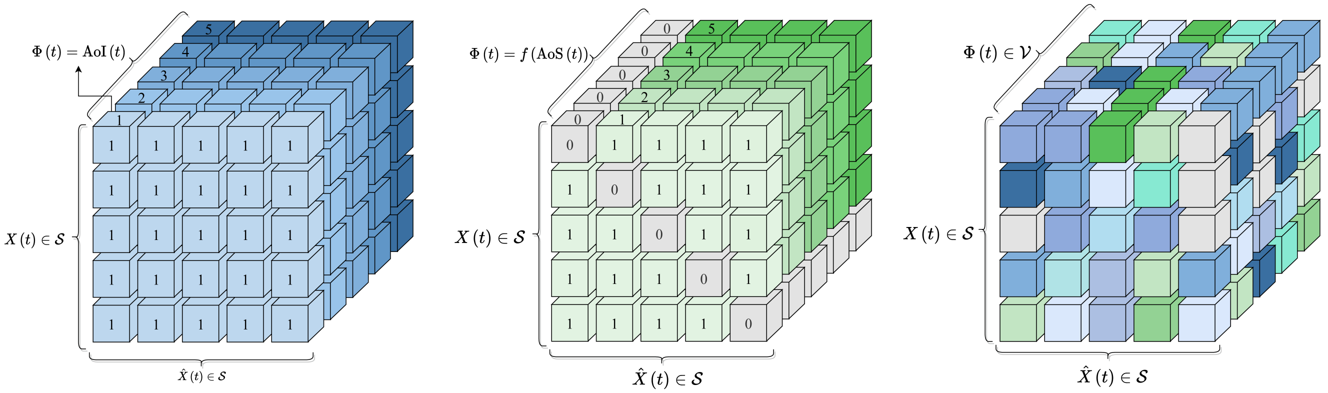
$$\Delta_{\mathrm{AoS}}(t) = t - W(t), \tag{3}$$

where $W(t)$ represents the latest time slot that the statuses between transceivers are synchronized. A possible AoS evolution shown in Table I demonstrates that when statuses are synchronized, AoS is equal to zero, and otherwise AoS linearly increases in asynchronized status duration.

As a general case of AoS, AoII is defined as a function of AoS. Specifically, in [41], AoII is described as a (non)linear function with regard to time duration multiplying a function reflecting the difference between source status and estimated status, that is

$$\Delta_{\mathrm{AoII}}(t) = f(t) \times g(X(t), \hat{X}(t)), \tag{4}$$

where $X(t)$ is the source status at time t, $\hat{X}(t)$ is the estimated result of $X(t)$ predicted by the receiver, $f(t)$ is an increasing time penalty function, and $g(X(t), \hat{X}(t))$ is an information penalty function that reflects the difference between the predicted result of the receiver and the actual status. Since the difference of two identical statuses is zero, the information penalty function $g(X(t), \hat{X}(t))$ should be equal to zero when $X(t) = \hat{X}(t)$. The AoII evolution is similar to that of AoS, that is, if the statuses between transceivers are


$\Phi(t) = \mathrm{AoI}(t)$
1
2
3
4
5
$X(t) \in \mathcal{S}$
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) = f(\mathrm{AoS}(t))$
0
1
2
3
4
5
$X(t) \in \mathcal{S}$
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) \in \mathcal{V}$
$X(t) \in \mathcal{S}$
$\hat{X}(t) \in \mathcal{S}$

(a) An example of how GoT can degenerate to AoI.

(b) An example of how GoT can degenerate to AoII.

(c) An example of generalized GoT.

Fig. 5. Examples of how GoT can degenerate to existing metrics. By setting the environment states as related to current message ages, the GoT can characterize AoI shown in subfigure (a). By setting the matrices composed of source status dimension and estimated status dimension as symmetric with values on the main diagonal as zeros, the GoT can describe AoII shown in subfigure (b). Naturally, by relaxing these constraints, a generalized metric called GoT can be constructed as shown in subfigure (c).

synchronized, AoII is zero; if they are asynchronized, then AoII increases with the time duration of the asynchronized state in a (non)linear manner.

Furthermore, [42] proposes a new metric called UoI, which is used to measure the nonlinear time-varying significance of state information by considering the impact of external environment. Specifically, UoI is defined as

$$\Delta_{\text{UoI}}(t) = f(t, \mathbf{\Phi}(t)) \cdot g(X(t), \hat{X}(t)), \tag{5}$$

where $\mathbf{\Phi}(t)$ represents the environment states at time t. That is, the time penalty function $f(t)$ is time-variant according to the current environment state $\mathbf{\Phi}(t)$. As shown in Table I, the UoI evolution is rather complex due to time-variant environment. Even the current statuses between transceivers keep asynchronized during time interval $[t_1, t_4]$, the manners of UoI increase are different in $[t_1, t_2]$ (linear) and $[t_2, t_4]$ (exponential), for the data in the latter time interval are more significant due to more urgent environment.

The recent works enrich the concept of significance to propose comprehensive tensor-based metric which unifies the above significance metrics. Specifically, the authors in [43] investigate a novel performance metric called GoT to directly quantify the significance of data considering the impacts of environments, status synchronization, and inherent costs, which is defined as

$$\Delta_{\text{GoT}}(t) = \text{GoT}(X(t), \hat{X}(t), \mathbf{\Phi}(t)), \tag{6}$$

where GoT$(\cdot)$ is a tensor with three dimensions, namely source status, estimated status, and environment states. Each element of GoT reflects the significance value (i.e., inherent cost) depending on the difference of source and estimated statuses between transceivers and current environment state, as shown in Table I, where the significance values in different time intervals are rather distinct. The authors prove that the GoT metric can reduce to existing metrics such as AoI, AoS, AoII, and UoI under certain conditions, as illustrated in Fig. 5.

B. Techniques: Towards Significance Protection

This subsection discusses the design of a significance-oriented semantic communication system within SAGSIN. The main focus is on ensuring appropriate sampling and accurate transmission of information with high significance, referred to as "significance protection". Taking into account the characteristics of SAGSIN, such as large-scale scenarios, highly dynamic channels, and limited device capabilities, we categorize existing research into the design of sampling, coding, and modulation policies, which we term as SPTs. These techniques aim to optimize the significance of transmitted data throughout the signal processing and transmission stages, which are listed in Table II.

1) Significance-oriented sampling policy design: Given the limited device capabilities of the SAGSIN, significance-oriented sampling policy is designed to alleviate the burden on data transmission. Generally, in a status update system shown in Fig. 6, a status update will be sampled only when the significance metric of the system reaches a certain threshold, implying that the current status update is significant enough that the receiver should acquire the current source status. By such significance-oriented sampling, most of data with less importance are filtered by the sampler, with only data with more semantics (i.e., significance) being sampled and transmitted to the receiver. Thus, the pre-processing burden of transmitter gets initially released, saving considerably large perception and communication resources which is severely limited the SAGSIN.

Content-agnostic sampling policy design. Since AoI is a typical timeliness metric, the design of AoI-related sampling policy has been studied widely in the literature. Intuitively, a zero-wait sampling policy, where a new update is sampled just when an acknowledgment (ACK) feedback is received by transmitter, will be naturally AoI-optimal since the newest sampled status update is the freshest and contains the most semantics. However, the authors in [46] point out that zero-wait policy is not always AoI-optimal, especially when the status varies slowly, since most of the status updates are repetitive and thus useless. To solve the AoI-optimal sampling problem

TABLE II
SUMMARY OF SIGNIFICANCE-PROTECTION TECHNIQUES

Perspective	**SAGSIN Issues Addressment**	**Related Metrics**	**Technical Details**		**Related References**
Sampling	Limited device capabilities	AoI, AoS, AoII, Nonlinear AoI	Content-agnostic	AoI-optimal threshold-based sampling with (non)linear age penalty	[39], [46], [47]
				AoI-optimal threshold-based sampling for Wiener or Ornstein-Uhlenbeck sources	[48], [49]
				Multi source joint sampling	[50]
			Content-aware	AoS-optimal threshold-based sampling	[40]
				AoII-optimal threshold-based sampling	[41], [51], [52]
				AoII-optimal sampling under random delay	[53]
				AoII-optimal sampling with retransmission	[54]
Coding	Highly dynamic channels, large-scale scenarios	AoI	Neglecting propagation delay	HARQ-IR	[55], [56]
				Truncated HARQ-CC for LDPC code	[57]
				HARQ-based Polar coded system	[58]
			Considering propagation delay	Codeblock assignment	[59]
				HARQ-IR	[60]–[62]
				HARQ-based Spinal coded system	[63]
				HARQ-based Polar coded system	[64]
Modulation	Limited device capabilities	AoI	NOMA/OMA		[65], [66]
			HARQ-CC aided NOMA		[67]
			NOMA in S-IoT network		[68], [69]

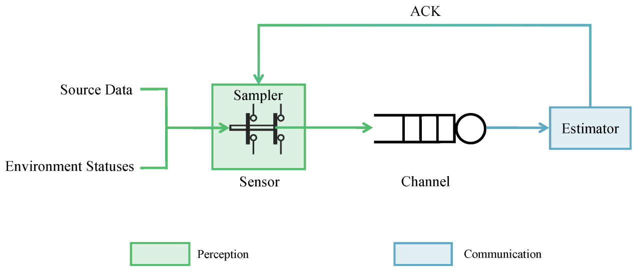


Fig. 6. A general system model of status update system facilitated by significance-oriented sampling. It is assumed that source data, environment statuses, and channel are all time-variant, affecting the significance of source data and thus determining whether to sample the current update. Moreover, the channel is usually assumed to be a queuing system with/without preemption, which will also affect the optimal sampling policy.

for such cases, [46] adopts a penalty function to measure the dissatisfaction level on data staleness, and formulate the age penalty minimization problem as a semi Markov decision process (MDP), which is solved by a divide-and-conquer approach. The zero-wait policy is proven to be optimal when the variance of log-normal distribution for service time is small; in contrast, the policy of waiting is optimal when the variance is large (i.e., a heavy tail distribution) or the penalty function increases rapidly with age growing. Similarly, the AoI-optimal sampling policy under energy constraint is solved in [47]. The choice of nonlinear age penalty and its impact on optimal sampling policy can also be seen in [39].

With regard to AoI-optimal sampling policy for specific distributions of source data, [48] discusses the optimal sampling policies for remote estimation of a Wiener process for different metrics including AoI and MSE. Specifically, the AoI-optimal sampling policy follows a threshold structure, that is, only when AoI reaches a threshold does the sampler generate a new status update for transmission. Meanwhile, the MSE-optimal sampling policy is equivalent to AoI-optimal one only when sampling times are independent with Wiener process; however, when the sampling times are determined by the knowledge of the Wiener process, the AoI-optimal sampling policy will no longer achieve the minimum estimation error as compared to MSE-optimal one. Another similar work is [49], where the optimal sampling policy of a Ornstein-Uhlenbeck source is studied serving as a general case of the study in [48]. The sampling optimization problem is formulated as a time-continuous MDP, and the solution is also proven to be a threshold structure related to instantaneous estimation error (which is related to MSE).

The above research focuses on sampling design of single source. For the case of multiple sources, [50] studies the total-peak-AoI-optimal and total-average-AoI-optimal sampling strategies, respectively. To simplify the solution of overall scheduling-sampling policy for multi-source case, the authors firstly prove that a maximum age first (MAF) scheduling policy provides the best age performance, and the optimal sampling policy for each source can be solved easily by a dynamic programming algorithm based on the MAF policy. The total-peak-AoI-optimal sampling policy is proven to be zero-wait, while the total-average-AoI-optimal sampling policy can be solved by water filling, which holds a threshold structure similar to [46], [48], [49].

Content-aware sampling policy design. Although AoI-oriented sampling policy do selects more significant data for transmission, this design is inherently content-agnostic, since the sampler selects only fresh status updates without knowing the content of data. In order to design a content-aware significance-oriented sampling policy, we should resort to AoI variants such as AoS and AoII, etc.

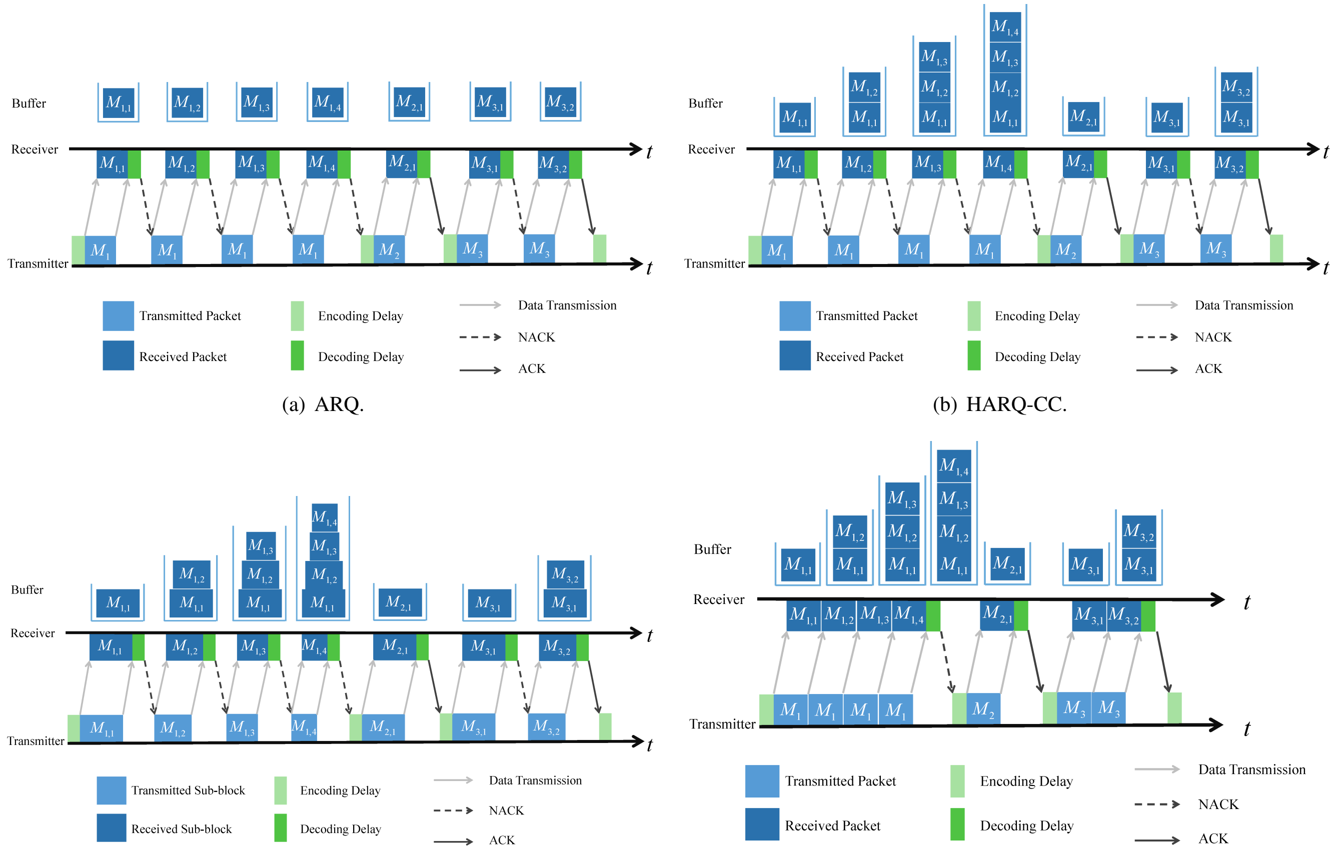
A major content-aware significance metric is AoII, and thus we review AoII-related data sampling optimization. In [40], the AoS-optimal and AoI-optimal sampling rate are respectively solved under a multi-source scenario. As a general case of AoS, the authors in [41] discuss AoII-optimal sampling policy with and without power constraints respectively. Under no power constraint, the optimal policy is "always update", which also achieves the minimum AoI and MSE simultaneously; however, the optimal policy is complex under a power constraint, which is solved by constraint MDP. [51] further studies AoII-optimal sampling policy under power constraint over an unreliable channel, where a threshold-based policy is proven to be optimal and adopted to reduce the complexity of global optimization problem. The authors in [52] point out the semantic (i.e., significance) characteristics of AoII, and showcase the superiority of significance-oriented AoII-optimal policy as compared to MSE-optimal and AoI-optimal ones. Considering the non-trivial transmission delay which is a fundamental characteristic in the SAGSIN, [53] solves AoII-optimal sampling policy when the status updates experience random delay, which also holds a threshold structure related to the maximum transmission delay. Also, a recent work [54] introduces limited retransmission with resource constraint to enhance the AoII performance.

2) Significance-oriented coding policy design: Considering the large-scale scenarios and highly dynamic channels of the SAGSIN, the implementation of significance-oriented coding policy necessitates the allocation of data rates, striking a trade-off between system efficiency and reliability. In particular, in scenarios involving bad channel conditions or a large volume of significant data, introducing more redundancy can improve the successful transmission probability. However, in contrast, when the channel conditions are favorable or there are less important data, using fewer codewords can still yield satisfactory results. As a widely used retransmission protocol, hybrid automatic repeat request (hybrid ARQ or HARQ), including HARQ with chase combining (HARQ-CC) and HARQ with incremental redundancy (HARQ-IR), accomplishes code rate allocation by dynamically adapting the number of retransmissions or/and the redundancy length in each retransmission, taking into account the factors such as channel conditions, delay constraints, and data significance. The differences among simple ARQ protocol, HARQ-CC protocol, and HARQ-IR protocol are illustrated in Fig. 7.

Coding policy design without considering propagation delay. Most of the works in this field have only considered the transmission delay and neglected the propagation delay, i.e., the only delay elements that affect AoI evolution is the code length. For example, [55] considers a scenario where the sensor transmits the collected data to a central entity over an unreliable link, and coding policies with low latency are designed to protect significant data. HARQ is used for the transmission of updates and the relationship between average timeliness and physical layer decisions (i.e., whether to retransmit the old data or send a new data) is analyzed. The analysis reveals a trade-off between average feedback rate and average timeliness. Specifically, for a constrained set of HARQ code word lengths, refining the code length can improve the average timeliness at the receiver. The work formulates the average age as the objective function and finds out the block allocation vector that minimizes the average age under the constraint of the average feedback rate. The results show that HARQ can greatly outperform ARQ and fixed-length schemes with no retransmission when the size of the incremental redundancy sub-block is properly chosen. The HARQ protocol considered in [55] is called HARQ-IR. Different from the ARQ scheme where the same packet is retransmitted after a failed decoding until ACK reception, the HARQ-IR scheme combines previously received symbols in decoding process to enhance successful decoding probability [70].

In [56], the authors utilize the HARQ-IR protocol in the status update system over a noisy channel. In case of decoding failure at the receiver, IR bits are transmitted to increase the successful decoding probability in the future. If decoding remains unsuccessful for a specified duration, a new status update is transmitted as a replacement. Conversely, when decoding is successful, the transmitter enters an idle state for a certain period after successful transmission, prior to sending a new update. The research focuses on optimizing the code word and IR length for each update, along with the waiting time, with the objective of minimizing the long-term average AoI over the binary symmetric channel (BSC).

In recent years, there has been an increase in research focusing on analyzing and comparing AoI based on retransmission protocols for specific state-of-the-art channel coding techniques. The authors in [57] firstly investigate the AoI performance of specific coding schemes. Specifically, the authors analyze the average AoI and energy cost in LDPC coded status update system with and without ARQ using a fixed redundancy scheme. Different coding policies, including the non-ARQ, classical ARQ, truncated ARQ, and truncated HARQ-CC schemes, are analyzed and compared. Among them, truncated ARQ involves repeated transmission of the current update until a maximum number of transmissions or a successful



(a) ARQ.

(b) HARQ-CC.

(c) HARQ-IR.

(d) Proactive HARQ-CC.

Fig. 7. The comparison of ARQ, HARQ-CC, and HARQ-IR. Specifically, the main difference between ARQ and HARQ is whether receiver uses the previous packets (or sub-blocks) stored in the buffer in the current decoding. Moreover, the packets transmitted in HARQ-CC are the same in the retransmission rounds, and the receiver simply combine the received packets using maximum ratio combining method. However, the packets transmitted in HARQ-IR (also called sub-blocks) are different in the retransmission rounds (usually with different code lengths), and the receiver combine the received sub-blocks into a whole vector for decoding. Hence, HARQ-IR usually achieves high coding gain than HARQ-CC, while the protocol design of the former is naturally higher. Note that the subfigures (b) and (c) both show reactive HARQ protocols where a retransmission occurs only if a negative acknowledgment (NACK) feedback is received by transmitter, while subfigure (d) shows a proactive HARQ protocol with chase combining. In proactive HARQ, the transmitter will continuously retransmit the packets no matter whether the message is successfully received or not. Unless otherwise specified, the mentioned HARQ protocols in this paper are all reactive HARQ.

reception is achieved. Truncated HARQ-CC combines failed packets with soft information from the initial transmission for decoding attempts, and thus eliminates the need for additional encoding schemes as it uses the same codeword for retransmission. Additionally, the accumulation effect of signal-to-noise ratio (SNR) in each HARQ retransmission improves the successful decoding probability. These inherent characteristics empower truncated HARQ-CC to strike a balance between AoI and energy consumption, achieving the best average age and moderate energy cost among the considered schemes.

As for AoI performance of status update system coded by capacity-achieving codes, the average AoI performance of Polar coded status updates over the additive white Gaussian noise (AWGN) channel is first investigated in [58]. Similar to the conclusion drawn by [57], compared to non-ARQ schemes, the HARQ-CC scheme can effectively reduce the average AoI, especially in low SNR regions. To further optimize the AoI of the system, two methods are proposed to optimize the design SNR and puncture length, both of which are important construction parameters of Polar codes.

Coding policy design considering propagation delay. Instead of only concentrating on transmission delay, some research further considers other types of delays (including propagation delay) and specific application scenarios, thereby providing a more precise depiction of the transmission process in the SAGSIN [59]–[62]. The authors in [59] conducted a comprehensive analysis of the AoI for two types of HARQ techniques: reactive HARQ and proactive HARQ. [59] derives unified closed-form expressions for the average AoI and average peak AoI for the two techniques under various types of delay, including coding delay, transmission delay, propagation delay, decoding delay, and feedback delay. Based on the derived explicit expressions, an optimization problem is formulated to minimize the AoI by investigating the optimal codeblock assignment strategy in the finite block-length (FBL) regime. The numerical results show that proactive HARQ offers advantages in terms of both age performance and system robustness.

In [60], the authors focus on the scenarios characterized by non-trivial propagation delays, such as space communications and satellite communications. The authors study AoI performance in a communication system with non-trivial propagation delays, where status updates are transmitted to the receiver through a binary erasure channel (BEC). In order to mitigate the effect of erasures on timeliness performance, an HARQ-IR scheme with a predetermined maximum number of retransmissions is employed. It is shown that a critical upper threshold related to propagation delay exists, within which retransmissions benefits the AoI. In cases where the propagation delay is longer than the threshold, allocating all the symbols to the first transmission is the optimal coding policy. Following [60], in [61], the author considers a satellite-based IoT system, where IoT devices observe physical processes and transmit status updates to a monitor node through an error-prone channel with significant propagation delay. [61] applies HARQ-IR scheme to the satellite-based IoT system and investigates the age-optimal redundancy allocation problem under reliability constraints. The results also demonstrate the superiority of HARQ-IR in minimizing AoI below a certain propagation delay threshold. In SAGSIN scenarios, [62] presents a novel fast HARQ-IR protocol which omits the decoding and feedback operations of the standard HARQ-IR scheme in the first few rounds. Under the constraint of finite block lengths, it is also demonstrated that the fast HARQ-IR scheme outperforms in the age-optimal rate allocation problem when the propagation delay is shorter than a certain threshold.

Taking into account specific channel coding schemes, the authors in [63] first consider a Spinal coded timely status update system based on HARQ with all practical delay elements. An upper bound of average AoI for Spianl codes is derived. Moreover, an HARQ transmission scheme is optimized in two steps to minimize the AoI. Firstly, the authors optimize the puncture pattern of Spinal codes and propose a transmission scheme based on incremental tail transmission puncture (ITTP). Secondly, an optimal transmission scheme is solved based on coarse-grained ITTP, where the number of symbols in each round is refined.

Based on the previous work [58], the authors in [64] investigate a Polar coded status update system that accounts for encoding, transmission, propagation, decoding, and feedback delays. The average AoI of the proposed system with different transmission protocols is analyzed. Based on the analysis, the authors focus on the optimization for Polar-coded HARQ design. Specifically, an effective algorithm is devised to optimize the design SNR in code construction and the code length in HARQ-CC transmission. Additionally, a greedy algorithm is utilized to optimize the code lengths for each transmission and the number of transmissions in Polar-coded HARQ-IR.

3) Significance-oriented modulation policy design: Modulation techniques is crucial in the wireless communication system, for they can enhance the reliability, spectral efficiency, and data transmission rate. However, the modulation technique design encounters challenges in efficiently utilizing limited spectrum resources to ensure timeliness in the SAGSIN. To address this, NOMA has emerged as a promising solution, which enables multiple users to share time-frequency resources by using power domain multiplexing. By adopting power allocation (i.e., assigning varying power levels to different users) which is a modulation policy, NOMA facilitates simultaneous message transmission of multiple users. Allocating more power to transmitted data of higher significance not only ensures reliable transmission of significant data but also enhances the timeliness of the system by directly influencing data rates and system capacity.

Various techniques are employed to enhance the timeliness performance of NOMA systems. [65] considers the problem of energy-efficient scheduling for minimizing the AoI in an opportunistic NOMA/orthogonal-multiple-access (NOMA/OMA) downlink broadcast wireless network. The initial step involves formulating a resource allocation problem to minimize the average AoI in the network. To address energy-efficiency considerations, the work takes into account both a long-term average power constraint and a maximum power constraint. Then the Lyapunov framework is employed to approximate the original problem as a queue stability problem. The obtained single time-slot optimization problem is further decomposed into power allocation and user scheduling subproblems. An efficient piece-wise linear approximation is utilized to solve the non-convex power allocation subproblem. Unlike the approach discussed in [65] that focuses solely on NOMA and OMA transmission schemes, [66] proposes an adaptive NOMA/OMA/cooperative-SWIPT-NOMA transmission scheme. In the NOMA scheme, the base station (BS) is responsible for allocating power to each user. However, in the cooperative-SWIPT-NOMA scheme, the BS not only allocates power to each user but also determines the power splitting coefficient, which determines the amount of harvested energy. The authors focus on a short packet communication system within a wireless network, where a BS transmits timely status updates to the users. The objective is to minimize the weighted average AoI by judiciously selecting appropriate multiple access techniques and corresponding power allocations. The resource allocation problem for the adaptive transmission scheme is formulated as an MDP and solved by iteration algorithms. Similar to [66], the authors in [67] investigate the AoI performance in a downlink wireless communication system, while HARQ-CC aided NOMA technology is employed in [67]. Furthermore, [67] considers a more flexible scenario where both power allocation decisions and retransmission decisions are made adaptively to minimize the system average AoI in each time slot. The authors notice the unfairness between users causes by the average-AoI-optimal policy. To address this issue, they formulate a problem to minimize the user's maximum expected AoI, ensuring fairness by preventing the farther user with weak channel gains from being deprived of timely service. It can be proven that a stationary optimal policy exists for the transformed MDP problem.

The aforementioned works primarily focus on ground-based wireless communication scenarios, while [68], [69] consider the satellite-based IoT (S-IoT) network. In particular, the authors in [68] propose an AoI-minimal resource allocation scheme in NOMA-based S-IoT downlink network. In this system, the satellite sends timely status updates to multiple user equipment (UE) devices. To ensure the freshness of the status updates in the network, an AoI optimization problem is formulated, taking into account long-term average power, peak power, and throughput constraints. The problem is then transformed into a series of online power allocation problems using Lyapunov optimization tools. Considering the limited computing resources of satellites, the particle swarm optimization (PSO) algorithm is employed to solve the non-convex optimization problem with linear computational complexity, named NOMA-AoI scheme. In the PSO algorithm, each particle means a possible solution which is the power allocation factors. By iteratively updating the velocities and positions of particles based on their historical information and the globally best information, the PSO algorithm explores the search space to find the optimal power allocation strategy. Based on [68], the authors in [69] further consider the network stability constraint in AoI optimization problem under NOMA-based S-IoT network. The problem is then transformed into three queue stability problems using the Lyapunov framework, which also converts the optimization problem into a series of single time slot deterministic optimization problems. The weights for the queue backlog and channel conditions are derived using the ListNet algorithm, a machine-learning-based approach, to obtain an optimized power allocation order with linear complexity. Similar to [68], the proposed NOMA long-term AoI minimization power allocation problem is also addressed using the PSO algorithm.

C. Lessons Learned from This Section

This section focuses on significance-oriented sampling, coding, and modulation policies to release heavy burden caused by massive source data and ensure that only significant data are sampled and transmitted for further processing and actuation. In fact, in the SAGSIN scenarios, although data generated by distributed edge devices are massive and multi-modal, not all of the data are significant enough, especially when the source data vary slowly as compared to data transmission. Therefore, as the first step of semantic signal processing, significance-oriented sampling filters most of unimportant source data, saving communication resources for sequential coding and modulation. On the other hand, even the sampled "significant enough" data may have different level of importance according to the varying environment. Therefore, significance-oriented coding and modulation policy designs provide unequal protection for sampled data with different level of importance. Specifically, more significant data (e.g., status updates with lower age) are given more coding redundancy or more power allocation to ensure their successful acknowledgment at the receiver. By significance-oriented semantic communication system, implementation of semantic communication systems in the SAGSIN (which will be elaborated in the following two sections) will be more energy-efficient.

IV. Perception-Communication-Computing-Integrated Semantic Communications in the SAGSIN: A Meaning Extraction and Reconstruction Perspective

In this section, we will review another major semantic communication system, called meaning-oriented semantic communication system. Since the primary purpose of such meaning-oriented system design is to achieve perfect recovery of the original data, this meaning-oriented semantic communication system is characterized by a strict symmetrical structure as shown in Fig. 8.

In the SAGSIN, even after the semantic-aware sampling mentioned in Section III, there are still a large volume of data due to numerous devices and massive connections. To reduce the transmitting data, it is necessary to perform semantic extraction from the sampled data. Here, the term "semantics" is interpreted as "meaning". However, the meaning of data is quite implicit to the extent that traditional codes cannot easily capture it. Hopefully, because of the similarity, we may associate meaning extraction with the process of human language learning. Humans gradually grasp semantics through extensive "training" such as teacher instruction, reading books, conversations with others, and so on. Also, a fact is that the original intention of the design for artificial neural networks (ANNs) is to mimic the human brain. Taking the deep neural network (DNN) as an example, through training on large amounts of data, it can learn something from the data to some extent just like humans. Thus, using DNNs to construct the system codec has become the most popular alternative nowadays to "learn" the semantics from original data.

Based on above, the meaning-oriented semantic communication system adopting a joint perception-communication-computing framework is established, which is illustrated in Fig. 8. Aiming at conducting intelligent meaning extraction, coding, and modulation, we assume that the input data have been sampled (in a significance-oriented manner). Specifically, DL-based meaning extractor captures the semantic information of sampled data to generate semantic representation, which serves as input of a JSCC encoder and modulator to generate transmitted messages for transmission. At the receiver, after demodulation and JSCC decoding, the semantic representation is recovered, which will be processed by DL-based meaning reconstructor to recover the raw sampled data. During the training of such large-scale neural network, edge servers and cloud data centers must be utilized to provide knowledge bases and facilitate distributed learning, since edge devices are usually limited in computing resources in the SAGSIN scenarios. Also, the knowledge bases distributed among edge/cloud servers can be synchronized via data sharing between transceivers.

In the following subsections, we will firstly elaborate on the metrics for representing the system performance in terms of the "meaning similarity" in Section IV-A. Next, in Section IV-B, we will provide a detailed description of the DNNs used to construct the system codec, which strongly intersects with the field of computer science. Finally, we discuss the related techniques for meaning enhancement in Section IV-C.

A. Metrics: Towards Meaning Similarity

In this subsection, we review the metrics for the meaning-oriented semantic communication system. A perfect metric can maximally represent the "meaning similarity" between the recovered data at the receiver and the original one at the

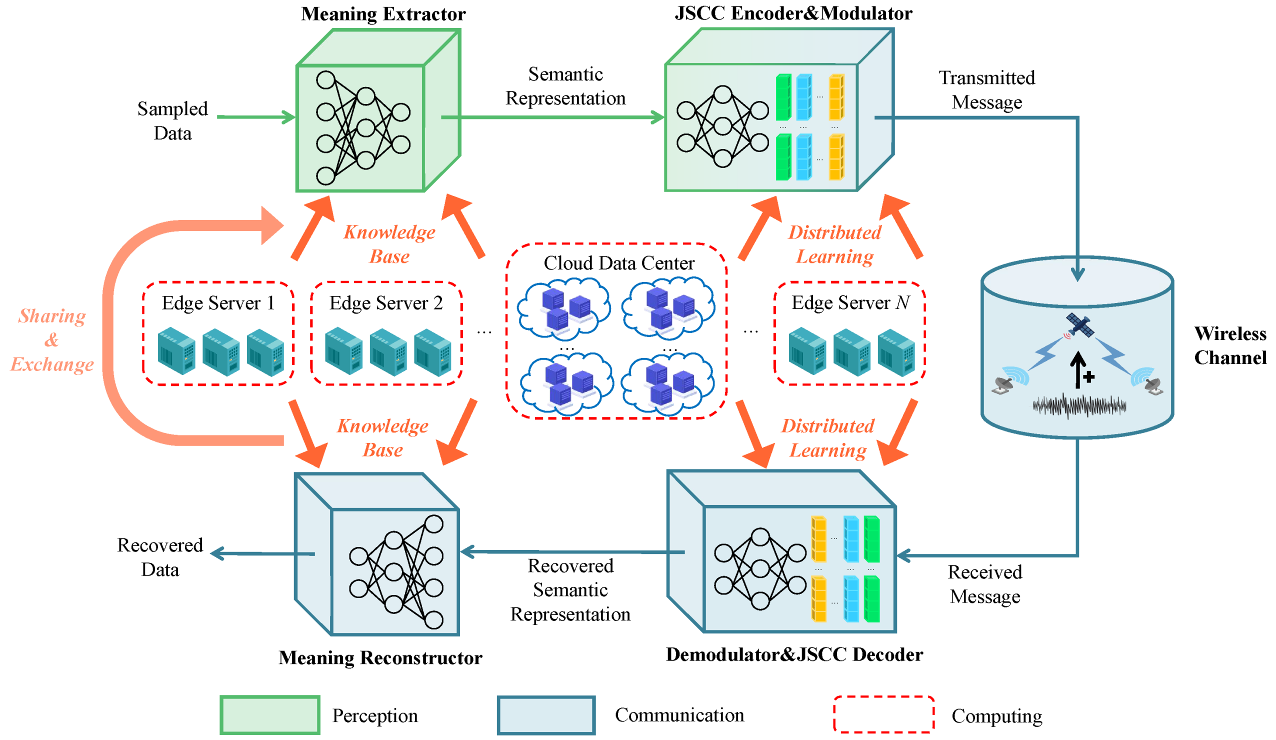


Fig. 8. The framework of joint perception-communication-computing semantic communications in the SAGSIN.

transmitter. We classify these metrics according to the source type and acquisition method. For the former, we consider text, image/video, and speech sources. As for the latter, based on using whether explicit formulas or neural networks (NNs) to obtain metric values, we analyze objective metrics and learning-based metrics. Please note that the NNs for obtaining metrics is different from the one for building the semantic communication system. The comparison of all the involved metrics is presented in Table III, and these metrics play crucial guiding roles for the system design in Section IV-C.

1) text sources: Here, the "meaning similarity" can be interpreted as the similar degree of the meaning contained in the texts (which is composed of words and sentences) of the transmitter and receiver.

Objective metrics. Similar to the bit error rate (BER) in traditional communication systems, the word error rate (WER) is a fundamental metric to measure the text differences. It is calculated by

$$\mathrm{WER} = \frac{S+D+I}{N}, \tag{7}$$

where S, D, I denote the numbers of word substations, deletions, and insertions respectively, and N is the word number of the original text. Unlike BER, WER can be greater than 1 due to a large number of insertions.

WER focuses on the similarity of individual words, while the bilingual evaluation understudy (BLEU) score [71] simultaneously considers the differences between n-length combinations of consecutive words (named n-grams) in two texts. As a typical method for measuring translation quality in machine translation, BLEU is now commonly used to describe the recovery performance in semantic communication systems. It is commonly used in logarithmic form as

$$\log \mathrm{BLEU} = \min(1 - \frac{l_{\hat{s}}}{l_s}, 0) + \sum_{n=1}^{N} w_n \log p_n, \tag{8}$$

where $l_{\hat{s}}$ and l_s are the lengths of recovered texts $\hat{s}$ and original texts s, w_n is the weight of n-grams, N is the total considering number of grams, p_n is the n-grams score, which is defined as

$$p_n = \frac{\sum_{k=1}^{K_n} \min(C_k(\hat{s}), C_k(s))}{\sum_{k=1}^{K_n} \min(C_k(\hat{s}))}, \tag{9}$$

where K_n is the number of elements in the n-th gram, and $C_k(\cdot)$ is the frequency count function for the k-th element in the n-th gram. The BLEU score falls within the range of 0 to 1, and a higher score indicates the greater text similarity.

Besides, consensus-based image description evaluation (CIDEr) [72] is used to assess the consistency and quality between the reference and generated image descriptions, and thus it can be seen as a metric for evaluating the similarity between texts. Moreover, the metric of semantic similarity (MSS) [73] can measure both the meaning similarity between texts and the correctness of the recovered text (i.e., whether it is a fluent statement), and the contribution weights of the two aspects for the metric can be adjusted.

The aforementioned metrics are all based on the text words,

TABLE III
SUMMARY OF METRICS BASED ON MEANING SIMILARITY

Sources	**Metrics**		**Explanation**	**Formula**	**Initial Reference**
Text	Objective	WER	Focusing on individual words	(7)	/
		BLEU	Focusing on consecutive words	(8)	[71]
		CIDEr	Reflecting the consistency and quality	($*$)	[72]
		MSS	Measuring the similarity and correctness	($*$)	[73]
	Learning-based	BERT score	A precision score by contextual understanding	(10)	[74]
		Sentence similarity	Focusing on complete sentences	(13)	[75]
Image/ Video	Objective	PSNR	Focusing on the pixel grayscale values	(14)	/
		SSIM	Considering the luminance, contrast and structure	(16)	[76]
		MS-SSIM	Analyzing SSIMs at different scales	(21)	[76]
	Learning-based	LPIPS	Starting from human perception	(22)	[77]
		FID	A biased estimator	(23)	[78]
		KID	An unbiased estimator	(24)	[79]
		AKD	Analyzing facial keypoints	($*$)	[80]
Speech	Objective	SDR	Comparing signal quality and distortion	(26)	[81]
		PESQ	Calculating speech quality scores	($*$)	[82]
		MCD	Measuring spectral distortion	($*$)	[83]
	Learning-based	FDSD	Extension based on FID	(27)	[84]
		KDSD	Extension based on KID	(28)	[84]
		cFDSD	FDSD under specific feature distributions	($*$)	[84]
		cKDSD	KDSD under specific feature distributions	($*$)	[84]

and their exploration for text semantics is far from sufficient. Taking the BLEU as an example, sentences “I am available on the weekend.” and “I am not busy during the weekend.” have the same semantics, but the BLEU score for them is not 1. To explore deep semantics of sentences (such as synonyms and expressions with similar meanings), we should resort to learning-based metrics.

Learning-based metrics. The bidirectional encoder representations from Transformers (BERT) model [85] is a pre-trained language model that learns rich language representations from a vast amount of text data through large-scale unsupervised training.

The BERT score [74] is a metric built upon the BERT model to evaluate text similarity, utilizing the contextual understanding learned by the BERT neural network. [74] mentions three types of BERT scores, which are recall, precision, and F1 scores[2]. Considering the need to measure the meaning similarity, the BERT score here specifically refers to the precision score. We assume that the representation vector generated by the embedding model for the original text $\langle x_1, x_2, \ldots, x_k \rangle$ is denoted as $\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \rangle$, and the representation vector for the recovered text $\langle \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m \rangle$ is denoted as $\langle \hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \ldots, \hat{\boldsymbol{x}}_m \rangle$. The BERT score is given by

$$\mathrm{BERT} = \frac{\sum\limits_{\hat{x}_j \in \hat{x}} \mathrm{idf}(\hat{x}_j) \max\limits_{x_i \in x} \boldsymbol{x}_i^{\mathrm{T}} \hat{\boldsymbol{x}}_j}{\sum\limits_{\hat{x}_j \in \hat{x}} \mathrm{idf}(\hat{x}_j)}, \tag{10}$$

where $\mathrm{idf}(\cdot)$ is the importance weighting function. Given M sentences from the test corpus $\{x^{(i)}\}_{i=1}^{M}$, $\mathrm{idf}(\cdot)$ is defined as

$$\mathrm{idf}(w) = -\log \frac{1}{M} \sum_{i=1}^{M} \mathrm{I}[w \in x^{(i)}], \tag{11}$$

where $\mathrm{I}[\cdot]$ is an indicator function. To increase the score readability, we adjust the scale of the BERT score in relation to its empirical lower bound b. The rescaled value $\hat{\mathrm{BERT}}$ of BERT is

$$\hat{\mathrm{BERT}} = \frac{\mathrm{BERT} - b}{1 - b}, \tag{12}$$

$\hat{\mathrm{BERT}}$ is between 0 and 1, and a higher score implies the better text similarity.

Sentence similarity [75] is another metric based on the BERT model that effectively evaluates the meaning similarity

In table III, formulas for some metrics are omitted by * due to space limits in the main text. Readers can go to corresponding initial references for details.

[2] For the definitions of these scores, readers can go to Section V-A for more details.

between two sentences. It is given by

$$\text{match}(\hat{s}, s) = \frac{\boldsymbol{B}_{\boldsymbol{\Phi}}(s) \cdot \boldsymbol{B}_{\boldsymbol{\Phi}}(\hat{s})^{\mathrm{T}}}{\left\|\boldsymbol{B}_{\boldsymbol{\Phi}}(s)\right\| \left\|\boldsymbol{B}_{\boldsymbol{\Phi}}(\hat{s})\right\|}, \tag{13}$$

where $\boldsymbol{B}_{\boldsymbol{\Phi}}(\cdot)$ is the BERT model to map the sentence to its semantic vector space. Like the rescaled BERT score, sentence similarity also ranges from 0 to 1. When the similarity between two sentences is maximum, the sentence similarity value is equal to 1.

The above two metrics are both based on the BERT model, which leverages the learning from billions of sentences to comprehend certain semantic information within the sentences. Therefore, they can provide a more accurate evaluation of the meaning similarity between texts than objective metrics.

2) Image (Video) sources: A digital image is composed of a finite number of pixels, while a video is composed of a continuous sequence of images, with each image referred to as a "frame". Therefore, when discussing metrics, images and videos can be grouped together for analysis, and we take image sources as an example next. The semantics of images heavily rely on context, making it challenging to directly define the "meaning similarity". For instance, without providing contextual information, we cannot determine whether "a yellow triangle" is more similar to "a green triangle" or "a yellow square".

Objective metrics. Assume that the original and recovered images are respectively $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. The matrix dimensions depend on the image properties, such as two dimensions for grayscale images, and three dimensions for colorful images.

The peak signal-to-noise ratio (PSNR), comparing each corresponding pixel of two images, is one of the typical metrics to measure the differences between images. The PSNR of a grayscale image is defined as

$$\text{PSNR}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 20 \lg \frac{\text{MAX}}{\sqrt{\text{MSE}(\boldsymbol{x}, \hat{\boldsymbol{x}})}}, \tag{14}$$

where MAX is the maximum grayscale value (typically 255 for an 8-bit image), and $\text{MSE}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ (mean squared error) is given by

$$\text{MSE}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \left\|\boldsymbol{x} - \hat{\boldsymbol{x}}\right\|_2^2. \tag{15}$$

As for colorful images, one common approach is to calculate the MSE for each color channel separately and then take the average, which is used in (14) to get the PSNR. Obviously, PSNR cannot provide an accurate similarity measurement of colorful images.

The structural similarity index (SSIM) [76], considering the similarity of the luminance, the contrast and the structure, is effective in evaluating the colorful image similarity. It is defined as

$$\text{SSIM}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = [l(\boldsymbol{x}, \hat{\boldsymbol{x}})]^{\alpha} \cdot [c(\boldsymbol{x}, \hat{\boldsymbol{x}})]^{\beta} \cdot [s(\boldsymbol{x}, \hat{\boldsymbol{x}})]^{\gamma}, \tag{16}$$

where $l(\boldsymbol{x}, \hat{\boldsymbol{x}})$, $c(\boldsymbol{x}, \hat{\boldsymbol{x}})$, and $s(\boldsymbol{x}, \hat{\boldsymbol{x}})$ represent the similarity of the luminance, contrast, and structure comparisons, respectively. They are given by

$$l(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{2\mu_{\boldsymbol{x}}\mu_{\hat{\boldsymbol{x}}} + C_1}{\mu_{\boldsymbol{x}}^2 + \mu_{\hat{\boldsymbol{x}}}^2 + C_1}, \tag{17}$$

$$c(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{2\sigma_{\boldsymbol{x}}\sigma_{\hat{\boldsymbol{x}}} + C_2}{\sigma_{\boldsymbol{x}}^2 + \sigma_{\hat{\boldsymbol{x}}}^2 + C_2}, \tag{18}$$

$$s(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{\sigma_{\boldsymbol{x}\hat{\boldsymbol{x}}} + C_3}{\sigma_{\boldsymbol{x}}\sigma_{\hat{\boldsymbol{x}}} + C_3}, \tag{19}$$

where $\mu_{\boldsymbol{x}}$, $\sigma_{\boldsymbol{x}}^2$ and $\sigma_{\boldsymbol{x}\hat{\boldsymbol{x}}}^2$ are the mean of $\boldsymbol{x}$, variance of $\boldsymbol{x}$, and the covariance between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, respectively, and C_1, C_2, and C_3 are the constants to avoid instability. If we set $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, we obtain a simplified expression, which is given by

$$\text{SSIM}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{(2\mu_{\boldsymbol{x}}\mu_{\hat{\boldsymbol{x}}} + C_1)(2\sigma_{\boldsymbol{x}\hat{\boldsymbol{x}}} + C_2)}{(\mu_{\boldsymbol{x}}^2 + \mu_{\hat{\boldsymbol{x}}}^2 + C_1)(\sigma_{\boldsymbol{x}}^2 + \sigma_{\hat{\boldsymbol{x}}}^2 + C_2)}. \tag{20}$$

SSIM is a value between 0 and 1, and a larger value indicates smaller differences between images.

The multi-scale structural similarity index (MS-SSIM) [76] is an extension of SSIM. It applies multiple downsampling operations on the original image to obtain images at different scales. Then, the SSIMs for these images are calculated separately, and finally a weighted summation operation of these SSIMs is performed, that is

$$\begin{aligned} \text{MS-SSIM}(\boldsymbol{x}, \hat{\boldsymbol{x}}) =& [l_M(\boldsymbol{x}, \hat{\boldsymbol{x}})]^{\alpha_M} \\ & \cdot \prod_{j=1}^{M} [c_j(\boldsymbol{x}, \hat{\boldsymbol{x}})]^{\beta_j} \cdot [s_j(\boldsymbol{x}, \hat{\boldsymbol{x}})]^{\gamma_j}, \end{aligned} \tag{21}$$

where α_M, β_j, and γ_j are used to adjust the relative importance for different components. MS-SSIM takes into account local details and global structures of images, thereby generally providing a better evaluation of the image quality compared to SSIM.

Considering the need for the accurate measurement of meaning similarity between images, the aforementioned three metrics are simple and shallow, and thus cannot capture the subtle differences perceived by humans.

Learning-based metrics. The learned perceptual image patch similarity (LPIPS) [77] learns image semantics from the human perception perspective through DNNs. It evaluates image similarity by calculating the distance between the feature representations of images extracted from pretrained DNNs such as the SqueezeNet, AlexNet and VGG. LPIPS is defined as

$$\text{LPIPS} = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(f_{hw}^l - \hat{f}_{hw}^l\right) \right\|_2^2, \tag{22}$$

where H_l and W_l denote the height and width of layer l, f^l and $\hat{f}^l$ are the normalized latent feature maps by layer l of the specific neural network for two images, h and w represent the (h, w)-th element of the feature map, and $\odot$ denotes the scale operation with scale vector w_l. By comparing the feature representations, LPIPS (lower values being better) can capture higher-level visual information.

Another category of metrics measures the meaning similarity between the generating image (created by the generative adversarial network (GAN) based on latent features from the real image distribution) and the real image, including the Fréchet Inception distance (FID) [78] and the kernel Inception distance (KID) [79], which utilize the pre-trained model from the Inception network for feature representations. FID is defined as

$$\begin{aligned}\text{FID} &= d^2\left((\boldsymbol{m},\boldsymbol{C}),(\boldsymbol{m}_w,\boldsymbol{C}_w)\right)\\ &= \|\boldsymbol{m}-\boldsymbol{m}_w\|_2^2 + \text{Tr}\left(\boldsymbol{C}+\boldsymbol{C}_w-2(\boldsymbol{C}\boldsymbol{C}_w)^{\frac{1}{2}}\right),\end{aligned} \tag{23}$$

where $\boldsymbol{m}$ and $\boldsymbol{C}$ are the mean vector and covariance matrix in the feature space of the generating image distribution, $\boldsymbol{m}_w$ and $\boldsymbol{C}_w$ are the mean vector and covariance matrix of the real image distribution, and $\text{Tr}(\cdot)$ indicates the trace of a square matrix. During the derivation of FID, [78] assumes that both the real and generated images follow the normal distribution (which is an ideal assumption), and the final FID is a biased estimator. As an improvement, [79] proposed KID, which is an unbiased estimator without the assumption of a normal distribution. It is given by

$$\begin{aligned}\text{KID} =&\frac{1}{m(m-1)}\sum_{i\neq j}^{m} k(\boldsymbol{x}_i,\boldsymbol{x}_j)\\ &+\frac{1}{n(n-1)}\sum_{i\neq j}^{n} k(\boldsymbol{y}_i,\boldsymbol{y}_j)\\ &-\frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(\boldsymbol{x}_i,\boldsymbol{y}_j),\end{aligned} \tag{24}$$

where m and n are respectively the number of samples in the generating and real images, $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ are the 2048-dimensional vectors obtained from the Inception network for the generating image, $\boldsymbol{y}_i$, $\boldsymbol{y}_j$ are the ones for the real image, and $k(\cdot)$ is the polynomial kernel defined as

$$k(\boldsymbol{x},\boldsymbol{y}) = \left(\frac{1}{d}\boldsymbol{x}^{\text{T}}\boldsymbol{y}+1\right)^3, \tag{25}$$

where d is the representation dimension, which equals to 2048 here.

Smaller values of FID and KID indicate greater meaning similarity between images. They are consistent with human perception and can reveal semantic differences between images at a deeper level.

In addition, considering the video conferencing scenario, [80] proposes the average keypoint distance (AKD) based on the keypoints (contain certain semantic information) extracted by a pretrained facial landmark detector to evaluate the differences between images.

The learning-based metrics above go beyond the pixel level and instead extract deeper semantic information from images using different pre-trained models. Thus, they can effectively measure the meaning similarity between images.

3) Speech sources: The semantics in speeches, which encompass not only the vocabulary and grammar but also the intonation, timbre, and speed of speech, are significantly complex and subjective. The same sentence with different intonations can convey different meanings. There are two evaluation methods, respectively evaluating the meaning similarity of the texts contained in the speeches or the speeches themselves. For the former, we can use all the aforementioned metrics for text sources. For the latter, we focus on the "meaning similarity" of speeches, which can be interpreted as the similarity degree of the inferred meanings derived from the speeches at the transmitter and receiver. The metrics related to the second method are introduced below.

Objective metrics. Assume that the original and recovered speeches are respectively $\boldsymbol{s}$ and $\hat{\boldsymbol{s}}$. Similar to the SNR, the signal-to-distortion ratio (SDR) [81] was initially proposed to calculate the ratio between the signal quality and distortion. As an extension, the SDR can be used to measure the similarity between two speeches [86].

$$\text{SDR}(\boldsymbol{s},\hat{\boldsymbol{s}}) = 10\lg\frac{\|\boldsymbol{s}\|^2}{\|\boldsymbol{s}-\hat{\boldsymbol{s}}\|^2}, \tag{26}$$

where $\|\boldsymbol{s}\|^2 = \langle\boldsymbol{s},\boldsymbol{s}\rangle$ represents the energy of $\boldsymbol{s}$. A larger SDR indicates a higher similarity between speeches.

Another metric called perceptual evaluation of speech quality (PESQ) [82] calculates the speech quality score by comparing the differences between speeches, which was established as Recommendation ITU-T P.862 in 2001. The specific process of calculating the PESQ score for two speeches mainly includes two stages: the alignment routine and the perceptual model. The former consists of the computation of the overall system gain, the intelligent reflecting surface (IRS) filtering, the time alignment (including the envelope-based alignment, the fine time alignment, the utterance splitting, and the perceptual realignment), and so on. The latter includes the IRS-receive filtering, the short-term fast fourier transform (FFT), the calculation of the pitch power densities, the realignment of bad intervals, the computation of the PESQ score, and so on. The range of the PESQ score is -0.5 to 4.5, and the larger value reveals the higher similarity between speeches. The calculation process of the PESQ is quite complex, and due to limited space, we have provided only a brief description here. For more details, see, e.g., [82] and the references therein.

Furthermore, the Mel cepstral distortion (MCD) [83], which measures the spectral distortion between synthesized and target speeches, can be regarded as a metric to evaluate the similarity between speeches. The smaller the MCD is, the closer the spectra between speeches are.

The aforementioned metrics do not sufficiently represent the semantic information of speeches, making it difficult to accurately evaluate the similarity between speeches at the semantic level.

Learning-based metrics. Based on the FID and KID, the unconditional Fréchet DeepSpeech distance (FDSD) and unconditional kernel DeepSpeech distance (KDSD) are proposed in [84] to measure the meaning similarity between speeches. By utilizing the pre-trained DeepSpeech2 model from the NVIDIA OpenSeq2Seq library, the KDSD and FDSD are able to capture the semantic information in the high-level speech feature space. The computational differences between the FDSD and FID are only reflected in the source and network types. According to (23), the FDSD is defined as

$$\begin{aligned}\text{FDSD} =&d^2\left((\boldsymbol{\mu}_{\boldsymbol{s}},\boldsymbol{\Sigma}_{\boldsymbol{s}}),(\boldsymbol{\mu}_{\hat{\boldsymbol{s}}},\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}})\right)\\ &= \|\boldsymbol{\mu}_{\boldsymbol{s}}-\boldsymbol{\mu}_{\hat{\boldsymbol{s}}}\|_2^2\\ &\quad+\text{Tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{s}}+\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}}-2(\boldsymbol{\Sigma}_{\boldsymbol{s}}\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}}))^{\frac{1}{2}}\right),\end{aligned} \tag{27}$$

where $\boldsymbol{\mu}_{\boldsymbol{s}}$, $\boldsymbol{\Sigma}_{\boldsymbol{s}}$ are the mean vector and covariance matrix in the feature space of the original speech distribution, and $\boldsymbol{\mu}_{\hat{\boldsymbol{s}}}$, $\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}}$ are the ones of the recovered speech distribution. $\mathrm{Tr}(\cdot)$ indicates the trace of a square matrix. Moreover, the KDSD is given by

$$\begin{aligned}\text{KDSD} =& \frac{1}{K(K-1)} \sum_{i \neq j}^{K} k(\boldsymbol{s}_i, \boldsymbol{s}_j) \\ &+ \frac{1}{\hat{K}(\hat{K}-1)} \sum_{i \neq j}^{\hat{K}} k(\hat{\boldsymbol{s}}_i, \hat{\boldsymbol{s}}_j) \\ &+ \sum_{i=1}^{K} \sum_{j=1}^{\hat{K}} k(\boldsymbol{s}_i, \hat{\boldsymbol{s}}_j),\end{aligned} \tag{28}$$

where K, $\hat{K}$ are respectively the number of samples in the original and recovered speeches. $\boldsymbol{s}_i$, $\boldsymbol{s}_j$ are the vectors obtained from the DeepSpeech2 model for the original speech, and $\hat{\boldsymbol{s}}_i$, $\hat{\boldsymbol{s}}_j$ are the ones for the recovered speech, and $k(\cdot)$ is the polynomial kernel defined as (25). Additionally, [84] also proposes the conditional Fréchet DeepSpeech distance (cFDSD) and conditional kernel DeepSpeech distance (cKDSD) to evaluate the meaning similarity between speeches under the specific linguistic feature distribution condition.

Compared to objective metrics, learning-based metrics utilize NNs to assist in capturing the semantic information, which aid in better representing the meaning similarity of speeches.

Remark 1. *"Perfect" metrics are believed to completely represent the "meaning similarity". However, there is not a unified definition for "semantics". Hopefully, with the development of semantic information theory, the metrics may also evolve accordingly.*

Remark 2. *The aforementioned objective metrics are not differentiable. Also, the learning-based metrics cannot be expressed in an explicit form, making it impossible to determine the differentiability as well. Therefore, they cannot be used to design the loss function in the training of semantic communication systems. The inconsistency between the objective (loss) function in the training phase and the evaluation metrics in the testing phase restricts the performance of meaning-oriented semantic communication systems. One obvious solution is to focus on researching differentiable evaluation metrics. However, these metrics should represent the "meaning similarity" as accurately as possible at the same time. One possible solution is to leverage the reinforcement learning (RL), where non-differentiable metrics can be used to guide learning processes and train high-performing systems by maximizing long-term rewards [87], [88].*

B. Methods: Based on DNNs

Shannon's separation theorem [12] proves that under the condition of infinite block lengths, the separate source-channel coding (SSCC) scheme can achieve the same performance as the JSCC scheme. Due to its modularity and ease of implementation, the SSCC scheme is widely applied in traditional communication systems, with source codes and channel codes being proposed in parallel on separate tracks.

However, affected by the typical characteristics of large-scale scenarios, highly dynamic channels, and limited device capabilities in the SAGSIN, the communication systems have extremely stringent requirements on latency, while the computing resources of IoT devices are limited. Moreover, in the semantic communication systems, redundant information is removed through the semantic extraction (a kind of perception) before encoding. All these facts lead to the prevalence of finite block lengths (even short ones) in the SAGSIN. Thus, the SSCC scheme cannot achieve the Shannon limit, and the JSCC one regains attention.

The mainstream approach in constructing semantic communication systems currently involves utilizing DNNs as source (semantic) and channel codecs, and training them jointly, called deep learning based JSCC (DeepJSCC), which is first applied in [89] to implement semantic communications. Taking the transmitter as an example here, DNNs sequentially perform operations such as semantic extraction and encoding on the sampled data. The output of the network can be directly transmitted into the channel. Therefore, it is necessary to review typical DNNs which can be used in JSCC.

In semantic communication systems, commonly used DNN-related models primarily include four categories: convolutional neural Networks (CNNs), recurrent neural networks (RNNs), Transformer and generative adversarial networks (GANs). These networks become the base of the system in Section IV-C.

1) CNN-related models: CNN is primarily used for space relevant sources, particularly image ones. The development of CNN can be traced back to 1962 when Hubel and Wiesel introduced the concept of Receptive fields [90]. In 1980, [91] proposed a neural network including convolutional and pooling layers. The first formal CNN model, called LeNet-5 [92], was proposed in 1998. It laid the foundation for the basic structure of CNNs: convolutional, pooling, and fully connected layers. However, due to limited computing resources and the vanishing gradient caused by sigmoidal activation functions, LeNet-5 did not receive the deserving attention.

AlexNet [93] marks the triumphant return of CNN, which consists of eight layers, including five convolutional ones and three fully connected ones. Compared to LeNet-5, AlexNet has the following four main improvements: First, a deeper network architecture is used to learn more intricate semantic features. Second, the sigmoidal activation function is replaced by the rectified linear unit (ReLU) one, alleviating the vanishing gradient. Third, researchers utilize two data augmentation methods and dropout technique to reduce overfitting. Fourth, overlapping pooling further enhances the richness of feature maps and extracts more detailed semantic information. Later, ZFNet [94] (a fine-tuned version of AlexNet), VGGNet [95] (a "very deep" version of AlexNet), and GoogLeNet [96] (also known as Inception-V1) are successively proposed.

The aforementioned CNNs are primarily used for image classification tasks, and thus the network output is the classification result. However, for image semantic segmentation tasks, the network should output an image of the same size as the input one and classify each pixel of the image into semantic categories (e.g., person, car, and tree). Based on this purpose, [97] proposes the fully convolutional network (FCN), which consists entirely of convolutional layers.

The development history of CNNs implies a belief: increasing the network depth does not decrease the model performance. A straightforward example is that adding a layer with the "identity mapping" does not change the model performance, but makes the network deeper. However, "Is learning better networks as easy as stacking more layers? [98]" Researchers discovered the "degradation" phenomenon through experiments, revealing a negative answer to the question. In order to make the network fully learn the data semantic features, activation functions are set to be nonlinear, to the extent that deep layers cannot simply perform the linear transformation (e.g., "identity mapping"). One solution is to build a "shortcut connection" between certain layers to strike better balance between nonlinear and linear transformations, forming a residual block. This is the ResNet [98].

Recently, some new variants of CNNs have been also proposed consecutively, including DenseNet [99] (employing dense connection structures), PSPNet [100] (utilizing pyramid pooling), and PatchGAN [101] (commonly used as a discriminator in GANs).

2) RNN-related models: RNN is mainly used for time relevant sources, particularly text and speech ones. In 1982, a Hopfield network was introduced by [102] as a precursor to RNNs. Later, a Jordan network was proposed in [103]. In [104], a Jordan network was improved, resulting in a Elman network, which is also known as the standard RNN. It consists of an input layer, a hidden layer, and an output layer. The neurons in the hidden layer are recurrently connected over time, allowing the output from the previous time step to be fed as input to the current one, thus enabling the network to remember and utilize historical information. This is why RNNs can process sequential data and learn semantic features from the context.

When the input sequence is too long, the time steps of the standard RNN become large, which may lead to the vanishing/exploding gradients during training. The solution is to utilize techniques like gradient clipping or more stable RNN variants such as long short-term memory (LSTM) networks. LSTM [105] effectively addresses the issue of long-term dependencies in standard RNNs by introducing memory units and gate mechanisms (an input gate, a forget gate and an output gate). It can efficiently propagate and express information in long sequences without neglecting/forgetting useful information from earlier time steps. In fact, only a portion of the input sequence is crucial (i.e., having more semantic information), and it needs to be remembered for a long time, while the remaining one can be appropriately forgotten. Therefore, the concept of "long short-term memory" was proposed in [105].

Because standard RNNs only consider past information of the input sequence, [106] introduced two independent RNN modules (forward and backward processing) simultaneously to capture both past and future context information, resulting in the bidirectional RNN (BiRNN). Afterwards, [107] combined the essence of LSTM and BiRNN to create bidirectional LSTM (BiLSTM) networks. Furthermore, [108] proposed deep BiLSTM networks to achieve better extraction and representation of semantic features. Gated recurrent unit (GRU) networks [109], a variant of LSTM, controls the flow of information by using an update gate and a reset gate. It is simpler than the LSTM model and can achieve comparable performance to LSTM with higher training efficiency.

3) Transformer-related models: The Transformer [110] was initially proposed in the field of natural language processing (NLP) (which is for text sources). Its primary purpose is to address the limitations of RNNs in handling sequential data, such as long-term dependencies and computational efficiency. However, due to its powerful capabilities, the Transformer has been successfully applied to other domains, including computer vision (image sources) and speech processing (speech sources).

The Transformer primarily comprises an encoder and a decoder. The encoder utilizes multiple (typically 6) encoding layers to convert input sequences into semantic space representations. Each encoding layer consists of a multi-head self-attention sub-layer and a position-wise fully connected feed-forward network sub-layer. The former is employed to establish global dependencies, while the latter performs nonlinear transformations. Residual connections are applied around each of the two sub-layers, followed by layer normalization. In comparison to the encoder, the decoder introduces an extra masked multi-head self-attention sub-layer at each decoding layer. By setting the attention weights of future positions to negative infinity, the masking operation prevents leakage of future information.

The Transformer is powerful due to the following aspects: Firstly, the multi-head self-attention mechanism divides the model into multiple heads to form subspaces and allows the model to automatically learn the relationships between different positions in the input sequence, capturing contextual information more effectively. Secondly, the self-attention mechanism allows for "parallel" computation of the encoder and decoder, effectively harnessing the parallel computing capabilities of modern hardware accelerators such as graphics processing units (GPUs). Thirdly, the Transformer effectively models long-term dependencies in sequences, alleviating the issue of gradient vanishing/exploding in RNNs when processing long sequences. Fourthly, the Transformer structure is simple, highly modular, and easy to extend and modify.

In addition, there are several variants of the Transformer proposed recently, such as the vision Transformer (ViT) [111] (for image classification tasks), the detection Transformer (DETR) [112] (for object detection tasks), the shifted window (swin) Transformer [113] (reducing computational complexities by the window self-attention mechanism), and so on.

4) GAN-related models: In practical applications, the GAN can be applied to various types of sources. The classic GAN was first proposed in [114], which consists of a generator and a discriminator. The purpose of the generator is to generate instances that appear natural and realistic, closely resembling the original data. And the role of the discriminator is to determine whether a given instance is real or fake. Clearly, they have an ”adversarial” relationship. The generator and discriminator are trained alternately, progressing together in the "game" and ideally reaching a dynamic equilibrium (i.e., Nash equilibrium). At this point, the generated instances from the generator are indistinguishable from real ones. There are two key points to note: Firstly, the ultimate goal is to obtain a well-trained generator, while the discriminator is an additional benefit. Secondly, there are no specific constraints on DNNs used for the generator and discriminator, and thus they can be CNNs, RNNs, or even Transformers.

The initiall GAN is an unsupervised learning model, and its training process does not require labeled data. Conditional GANs (CGANs) [115] bring GANs back into the realm of supervised learning, alleviating the problem of unstable training in GANs. Deep convolutional GANs (DCGANs) [116] are the first successful implementation of GANs that use CNNs instead of fully connected layers. This significantly reduces the number of network parameters while greatly improving the quality of generated data. Wasserstein GANs (WGANs) [117] improve the stability of model training by introducing the Wasserstein distance as a loss function. Self-attention GANs (SAGANs) [118] introduce an attention mechanism that enables the network to better understand the semantic information. The introduction of BigGANs [119] reveals that GAN network training can also benefit from large-scale data and computing resources. BigGANs achieve unprecedented levels of image generation quality. Recently, some new variants of GANs have been proposed, such as StyleGANs [120] (controlling the style and features of generated images), HiFi-GAN [121] (focusing on generating high-fidelity audio signals), and so on.

Remark 3. *There are also other neural networks used in JSCC schemes. Graph convolutional networks (GCNs) [122], which are commonly utilized to extract semantic features from non-Euclidean structured images (e.g., social networks and knowledge graphs), are used in [123] to construct joint source-channel codecs. [87], [88], [124]–[126] employs reinforcement learning networks (RLNs) to construct semantic communication systems.*

Remark 4. *Semantic communication systems can be also constructed by DL-based SSCC (DeepSSCC) schemes, which consist of two categories. The first one uses traditional codecs for channel coding, and only the source (semantic) codec are trained [127]–[130]. The source and channel codecs of the second one are both DNNs, but trained separately [131], [132].*

C. Techniques: Towards Meaning Enhancement

From a macro perspective, the meaning-oriented semantic communication system proposed in this section integrates the perception, communication, and computing parts. The coordinated collaboration of these three parts contributes to better application of the system in various scenarios within the SAGSIN. Now, let's narrow our focus to some specific details and consider the practical works of system implementation. It is evident that DNNs serve as the core to build semantic codecs. However, considering the objective realities such as large-scale scenarios, highly dynamic channels, and sensitive communication security issues in the SAGSIN, it is necessary to incorporate additional targeted "designs" for the components of semantic communication systems. We name these "designs" as METs and categorize them into four major classes, which respectively focus on the channel adaptation, rate control, interpretability, and security. In the following, we will review the existing related works for each category, and the summary of these techniques is shown in Table IV.

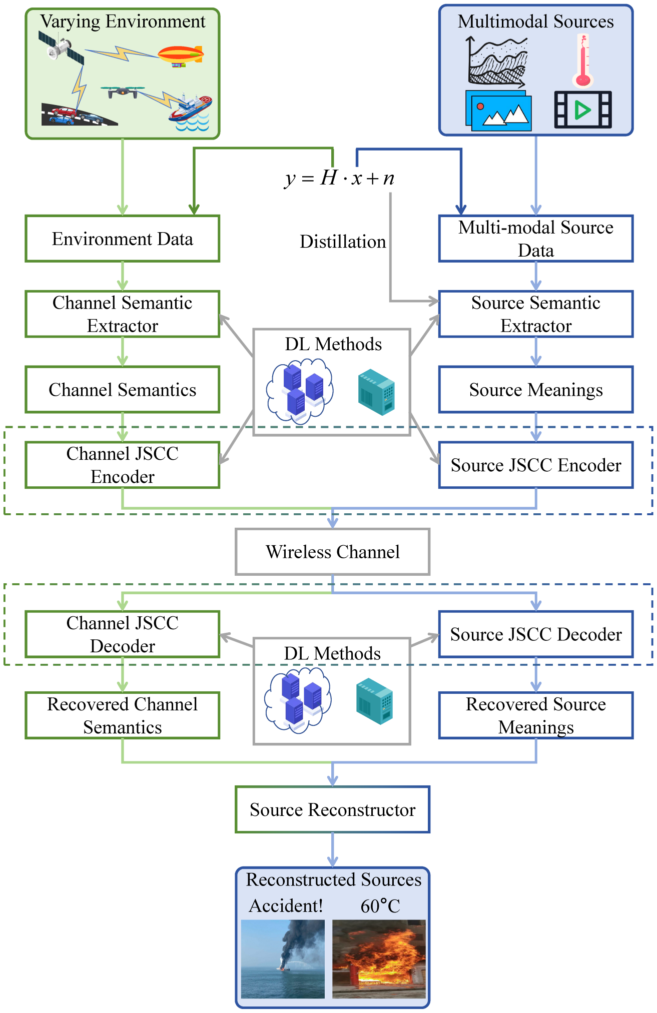


Fig. 9. An illustration of typical channel-related METs, including channel semantic extraction (CSI involved) and noise resisting. It is worth noting that the source and channel JSCC encoders (decoders) are usually integrated by one module in the literature.

1) Channel-related METs: Complex and highly dynamic channels in the SAGSIN pose challenges for channel modeling in system design. Meanwhile, in order to ensure high recovery performance, DL-based systems require similar channel conditions during both training and testing phases. Therefore, it is necessary to conduct targeted channel-related analysis and design, such as extraction of channel semantics, noise-resistant mechanisms, and the capturing & utilization of channel state information (CSI). The main procedures of realizing these channel-related METs are demonstrated in Fig. 9.

Analyzing and extracting the channel semantics to assist source information encoding is a new research direction [133],

TABLE IV
SUMMARY OF MEANING-ENHANCEMENT TECHNIQUES

Techniques	SAGSIN Issue Addressment	Technical Details		Related References
Channel-related METs	Highly dynamic channels	Extracting channel semantics	Role as a new research direction	[133], [134]
			Concept of "channel semantics"	[135]
		Designing noise-resistant mechanisms	Joint semantic-noise coding	[88]
			Literal semantic noise and adversarial semantic noise	[136]
		Capturing and utilizing CSI	Channel ModNet	[137], [138]
			Attention feature module	[139]
			Response network	[140]
			Encoding and modulation match with CSI	[141]
			Training with CSI	[142]
Rate-control-related METs	Highly dynamic channels	Retransmission-based	SC-RS-HARQ, SCHARQ	[130]
			SVC-HARQ	[80]
			IK-HARQ	[143]
		Non-retransmission-based	Channel feedback	[144]
			Policy network	[145]
			Rate adaptive transmission	[146], [147]
			Rate allocation network	[148]
			Adaptive density learning module	[127]
			Rate controller	[126]
Interpretability-related METs	Limited device capabilities	KG	Interpretable semantic information detection algorithm	[149]
			Transformer-based knowledge extractor	[150]
			Semantic representation based-on KG	[73]
		ProbLog	Conversion from NPM to a symbolic graph	[151]
			ProbLog-based KB	[152]
		Conceptual space	Abstracting semantic measurement into geometric space	[153]
Security-related METs	Security	Understanding attack mechanisms	Destructive physical layer semantic attacks	[154], [155]
			Backdoor (Trojan) attacks	[156]
		Designing defense mechanisms	Counter-eavesdropping DeepJSCC model	[157]
			Adversarial encryption training scheme	[131]
			Privacy filter	[158]
		New security-related metrics	SOP, DFP	[159]

[134]. In [135], the authors introduces the concept of "channel semantics", which encompasses parameter semantics (e.g., angle of departure, angle of arrival, number of paths, and so on) and environmental semantics (e.g., the layout, shape, and category of objects in the images). The system obtains environmental information in the SAGSIN through devices such as cameras and radars, and extracts channel semantics by DL techniques. The system considering channel semantics exhibits superiority in terms of meaning similarity metrics such as WER and FDSD.

To address the impact of noise in complex channels, a confidence-based distillation mechanism is established in [88] to achieve the "joint semantic-noise coding (JSNC)". The distillation time in the semantic distillation mechanism of

JSNC can be automatically adjusted to accommodate the fluctuating channel conditions. Taking the example of the AWGN channel with 10 dB SNR, JSNC achieves a 24% lower WER compared to the Transformer baseline. Furthermore, considering the unique "semantic channel" (the information flow from semantic representation to semantic recovery) of the semantic communication system, the authors in [136] classify semantic noise for text sources into "literal semantic noise" and "adversarial semantic noise", and propose a robust DL-based semantic communication system (R-DeepSC) to resist them. Specifically, a calibrated self-attention mechanism is utilized to counteract the former, and an adversarial training method is employed to combat the latter. Compared to other baselines considering physical noise only, R-DeepSC achieves superior BLEU and BERT scores at different SNRs.

Effectively capturing and utilizing the rich CSI in the SAGSIN will bring extra recovery performance gains. The authors of [137] and [138] both design a plugin-in CSI modulation module, called "Channel ModNet". [137] inserts it into the DeepJSCC decoder and proposes an adaptive semantic communication (ASC) based on the overfitting of the source and channel. This design adapts well to different SNRs in block fading channels. Reaching the same PSNR, ASC saves bandwidth up to 41% compared to the benchmark. In [138], the authors propose a wireless image transmission Transformer (WITT) model, utilizing the Swin Transformer and incorporating "Channel ModNet" into the codec. Compared to two baselines, namely the CNN-based DeepJSCC and the traditional BPG-LDPC SSCC, WITT exhibits superior PSNR and MS-SSIM at a wide range of (particularly low) SNRs. [139] proposes an improved version of the classical DeepJSCC, called the adaptive DeepJSCC. Specifically, an "attention feature (AF) module" is utilized to interact with codec DNNs, which can process the input CSI to adapt to different channel conditions. Experimental results show that the adaptive DeepJSCC achieves better PSNR than the classical DeepJSCC under a wide range of SNRs. In [140], the authors introduce a "response network" enabled nonlinear transform source-channel coding (NTSCC) framework to explicitly construct response functions and directly embed CSI into the parameters of the DeepJSCC codec. Under the AWGN channel with 10 dB SNR, this architecture achieves higher PSNR than traditional transmission schemes with standardized image codecs across all channel bandwidth ratios. Besides, a joint coding-modulation (JCM) scheme based on BPSK modulation for digital semantic communications is proposed in [141], matching the encoding and modulation processes with CSI better and achieving larger PSNR. Also, the lite distributed DL-based semantic communication system (L-DeepSC) in [142] is assisted by CSI during training, reducing the impact of fading channels on the transmission and increasing BLEU scores.

2) Rate-control-related METs: Rate control is another measure to address the issues of highly dynamic channels for the SAGSIN. In the JSCC encoding process, allocating flexible rates to the extracted semantics allows for better adaptation to time-variant channels. Specifically, based on whether the retransmission protocols are adopted to prompt the correct decoding of the current codeword after decoding failure, rate-control techniques can be classified into retransmission-based ones and non-retransmission-based ones.

A typical retransmission-based technique is the HARQ. Recently, a few variants of HARQ have been proposed in semantic communication systems. It is worth noting that the HARQs here are improved versions to adapt to the meaning-oriented semantic communication system, which is the main difference compared to the ones discussed in Section III-B. In [130] and [80], HARQ-IR used in traditional communication systems is improved to apply to semantic communication systems. The authors in [130] first propose SC-RS-HARQ, which belongs to the first category of DeepSSCC mentioned in Remark 4. The semantic codec is based on Transformers, while the channel codec utilize Reed-Solomon (RS) codes. Because HARQ-IR is only based on RS codes, the authors directly incorporate it into SC-RS-HARQ. To further improve system performance, they improve HARQ-IR and apply it to the DeepJSCC architecture, resulting in SCHARQ. Simulation results show that in a wide range of (especially low) SNRs, SCHARQ achieves lower WER compared to SC-RS-HARQ. [80] proposes a semantic video conferencing (SVC) network and improves HARQ-IR specifically for SVC, resulting in SVC-HARQ. The semantic error detector is employed to ascertain whether an incremental transmission is necessary for the received frame. Compared to traditional communication systems, SVC-HARQ exhibits lower AKD at high BERs. In [143], the authors propose a progressive semantic HARQ scheme based on the incremental knowledge (IK-HARQ). An encoding mechanism with adaptive semantic rate control is employed to dynamically adjust rates based on the contextual semantic complexity and channel conditions. Compared to other benchmarks, IK-HARQ outperforms in terms of BLEU in a wide range of SNRs.

It is worth noting that non-retransmission does not imply non-feedback. In [144], no retransmission mechanism is employed, but the authors utilize feedback signals from the current codeword to guide the encoding of subsequent codewords, and introduce an autoencoder-based JSCC scheme, called DeepJSCC-f. They employ layered autoencoders and leverage channel output feedback to achieve variable codeword length transmission. The simulation results demonstrate the benefits of feedback in improving system performance. Compared to the non-feedback DeepJSCC and traditional communication systems, DeepJSCC-f reaches higher PSNR at different SNRs.

The following works on rate control are all without retransmission or feedback. [145] proposes an adaptive JSCC for wireless image transmission. The "policy network" is introduced to activate or freeze visual features by dynamically generating binary masks, thereby achieving rate control. In scenarios with a high SNR and lower information content in the image, this model can learn reasonable strategies with less bandwidth. [146] and [147] propose the same "rate adaptive transmission" mechanism to apply to embedding vectors in latent representation of the encoder side. This enables the model to learn to allocate limited bandwidth resources in order to maximize overall performance. The model in [146] exhibits superior performance in terms of PSNR and MS-

SSIM compared to DeepJSCC and traditional communication systems at various SNRs. In [147], the proposed architecture achieves up to 50% savings in channel bandwidth costs compared to traditional communication systems while maintaining the same PSNR and MS-SSIM performances. Besides, [148] introduces a variable-length semantic-channel coding (VL-SCC) method, where a “rate allocation network (RAN)” is utilized to estimate the optimal code length and enhancing the PSNR and LPIPS performance. The forward-adaption-based (FA-based) autoencoder scheme proposed in [127] improves performance in MS-SSIM and PSNR by introducing an “adaptive density learning module” at the transmitter to adaptively construct the codebook. [126] proposes a reinforcement learning based adaptive semantic coding (RL-ASC) model, which use a “rate controller” to adaptively adjust the code length. Under different rates, RL-ASC exhibits superior performance to benchmarks in terms of PSNR, SSIM, FID, and KID.

3) Interpretability-related METs: One of the widely criticized flaws of DL is lack of interpretability. The black-box nature makes it hard for people to understand the internal decision process of models, hindering model improvement and optimization. Therefore, enhancing the interpretability of semantic communication systems by specific means (e.g., knowledge graph (KG) and probabilistic logic programming language (ProbLog)) can help further improve communication performance in the SAGSIN. Meanwhile, enhancing interpretability also facilitates the efficient learning of NNs by accelerating the convergence of complex models. This can better address the characteristic of limited device capabilities in the SAGSIN.

KG is a graphical structure used to describe the relationships between entities. It is an extension and development of a knowledge base (KB). KGs contain a series of semantic triplets in the form of “entity-relation-entity.” As an advantage compared to KBs, KGs provide richer semantic associations and contextual information. Therefore, with the backdrop of building the meaning-oriented semantic communication system, KGs gain increasing attention recently. [149] proposes an interpretable semantic information detection algorithm by using KG triplets as semantic symbols. Also, an efficient semantic correction algorithm is also introduced by extracting inference rules from the KG. The KG-based semantic communication system in [149] exhibits high sentence similarity in a wide range of (particularly low) SNRs. The authors in [150] present a KG-enhanced semantic communication framework. In particular, they design a Transformer-based knowledge extractor in the decoder side to extract relevant factual triples based on the received noisy signal, thereby enhancing the semantic decoding capability. The experimental results demonstrate that regardless of channel types, in low SNRs, the semantic communication system based on knowledge extractor always yields a BLEU improvement of over 5% for the model. An interpretable semantic communication framework for text data transmission is considered in [73]. The semantic information is represented through a collection of semantic triples based on a KG. In addition, the authors also define a new metric, namely MSS, which comprehensively measures the accuracy and completeness of the reconstructed data. The framework proposed in [73], compared to traditional communication systems, can reduce the transmission data by 41.3% and improve the overall MSS by two times.

ProbLog is a programming language that combines logic programming with probability theory, extending traditional logic programming. It represents knowledge in the form of logical clauses and allows specifying probability facts and rules within these logical clauses, enabling the representation and inference of knowledge involving uncertainty and probability. ProbLog has been widely applied in various fields, including machine learning (ML), NLP, and so on. In [151], the authors propose a semantic protocol model (SPM) to enhance the interpretability of the principles and impact of the NN-based protocol models (NPMs). Specifically, ProbLog is employed to convert the NPM into a symbolic graph. The experimental results confirm that the SPM performs closely to the NPM in performance while occupying only 0.02% of the memory. [152] introduces a KB-based on the ProbLog to facilitate semantic information exchange at the semantic communication level.

In addition, [153] proposes semantic communications based on the “conceptual space.” Conceptual spaces are similar to KBs but abstract the semantic information measurement into a geometric space, defined as distances in the conceptual space. This approach also enhances the interpretability of the model.

4) Security-related METs: Communication security, as a prerequisite for accurately transmitting information, has always been an essential concern, especially in the SAGSIN. This comprehensive network consists of numerous nodes, handling massive services at any given moment. When two specific nodes engage in the end-to-end communications, it is undesirable to be eavesdropped on the information by other unrelated nodes. Therefore, focusing on the security issues under the background of semantic communications becomes an emerging research direction [160]. A typical semantic communication system with eavesdropping is shown in Fig. 10.

The first step of designing techniques to defend against attacks is understanding the attack mechanisms. A few works validate the impact of certain specific attacks on the performance of the semantic communication system, demonstrating the necessity of researching semantic communication security. [154] proposes a model called SemBLK, which can learn to generate destructive physical layer semantic attacks for end-to-end semantic communication systems in a black-box setting. Semantic perturbations are generated by introducing a surrogate semantic encoder in the SemBLK. Experimental results demonstrate the destructive and imperceptible nature of black-box attacks in the SemBLK by comparing metrics such as PSNR and SSIM with benchmarks. Similarly, in the recent work [155], a semantic perturbation generator called SemAdv is trained for physical-layer attacks that can pollute the images with specific semantics in an imperceptible, controllable, and input-agnostic manner. Besides, the authors in [156] demonstrate that the meaning-oriented semantic communication system exhibits weak resistance against backdoor (Trojan) attacks. The experimental results indicate that backdoor attacks are stealthy and selective.

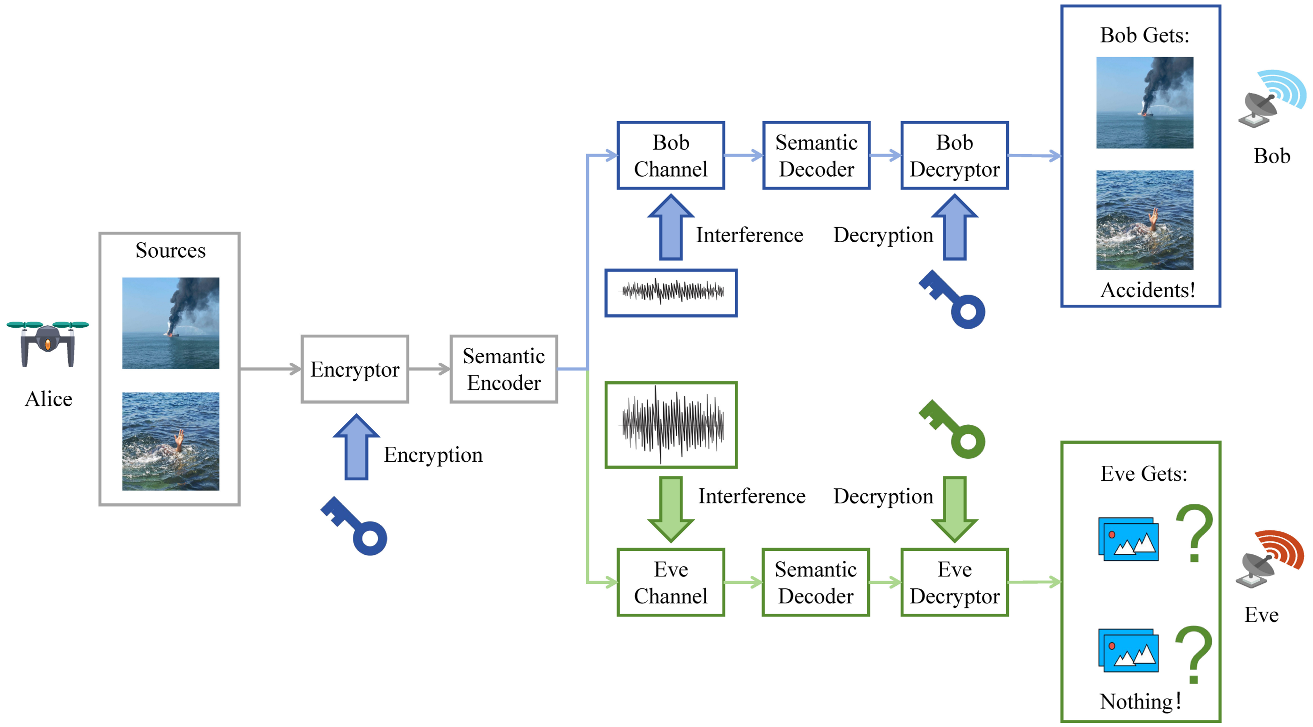


Fig. 10. An illustration of typical security communication in the SAGSIN scenario. Alice and Bob share the key such at the source data encrypted by Alice can be decrypted by Bob and cannot be decrypted by those who do not have the key. Eve aims at eavesdropping the transmitted data while suffering from severe interference and not knowing the key. Thus, Eve cannot recover the data sent from Alice.

The second step is to design defense mechanisms to enhance the semantic security. [157] proposes a counter-eavesdropping DeepJSCC model, called DeepJSCEC. It can resist the eavesdroppers' chosen-plaintext attacks without assuming their channel conditions. Compared to the SSCC encryption schemes of traditional communication systems, this model exhibits comparable or even superior PSNR, SSIM, and MS-SSIM performance at different SNRs. An encrypted semantic communication system (ESCS) is proposed in [131]. Specifically, the authors introduce an adversarial encryption training scheme, which enables the model to achieve high communication accuracy in both encrypted and unencrypted modes. Under the support of ESCS, at high SNRs, the BLEU score of Bob (receiver) approaches 1, while that of Eve (attacker) is less than 0.2. [158] proposes a StyleGAN-based semantic communication framework, which utilizes a privacy filter and a KB to replace private information with natural features in the KB, ensuring communication security. Compared to the benchmark (a kind of DeepJSCC), the model has a smaller and more stable LPIPS value under different SNRs.

Some new security-based metrics used in semantic communications are also proposed in the literature. In [159], the authors examine and contrast traditional security methods, such as physical layer security, covert communications, and encryption, in terms of semantic information security. They introduce two metrics, which are semantic secrecy outage probability (SOP) for physical layer security techniques, and detection failure probability (DFP) for covert communication techniques. With these metrics as the ultimate goal, the security-related design becomes clearer.

Remark 5. *Finally, the following works, not classified into the above four categories, have also made significant contributions to semantic communications. Many of these have been used as benchmarks of the aforementioned references. We categorize them based on the source types: text [89], [128], [161]–[164], image/video [124], [125], [165]–[174], and speech [86], [132], [175]–[177]. Readers can delve into the details.*

D. Lessons Learned from This Section

This section provides a detailed review of the meaning-oriented semantic communication system. The metrics reflecting the "meaning similarity" between the original and reconstructed data serve as key guidelines for the system design, and it is worth noting that the accurate representation of the "meaning similarity" for the utilized evaluation metrics is the basis to improve the superior system performance. As the fundamental method for system design, DNNs are considered as the absolute core of the meaning-oriented communication system. When designing the system, considering the limited device capabilities in the SAGSIN, incorporating meaning-enhancement techniques into DNNs is a powerful approach to further improve the system performance. As a summary, this section provides an idea for designing the meaning-oriented semantic communication system: starting from maximum meaning similarity metrics, establishing DNN-related models, and then proposing targeted meaning-enhancement techniques.

V. Perception-Communication-Computing-Actuation-Integrated Semantic Communications in the SAGSIN: A Task Effectiveness Perspective

In this section, we interpret the term "semantics" as "effectiveness-related information" and review a newly proposed approach of semantic communications called effectiveness-oriented semantic communications (also called task-oriented semantic communications), which focus on the effects yielded by the task actuation process by adopting a joint reconstructor & decision-maker design.[3]

Before the review, we firstly elaborate on the reason for introduction of task-oriented communications. Traditionally, both technical communications and meaning-oriented semantic communications adopt separate reconstructor and decision-maker design, where the receiver are expected to firstly reconstruct the estimation of source data by inverse process of transmitter, and secondly generate task decision results based on the estimation of data. The reason why reconstruction process is imperative is that the receiver cannot distinguish which part of received data is task-oriented, and thus all the source data must be firstly reconstructed for further decision. In stark comparison, task-oriented communications jointly design the reconstructor and decision-maker, where an intelligent module at the receiver implicitly processes the received data and directly produces task decision results.

This joint design has a salient characteristic of non-symmetry, since no explicit reconstruction process corresponding to pre-processing (e.g., encoding or modulating) occurs at the receiver. Joint reconstructor & decision-maker design is feasible for receiver side to generate moderate decision results, because the received data contain all the information that is needed for tasks, and thus reconstruction is unnecessary for decision process except when the task itself is to reconstruct source data. Moreover, such non-symmetric design of joint reconstructor & decision-maker module is proven to be more simpler in computation than separate design by state-of-the-art literature to be elaborated in Section V-B, which perfectly addresses the issue of limited computing resources in the SAGSIN.

On the other hand, since the goal of task-oriented communications is just yielding perfect decision at receiver side for task actuation at the terminal, semantic extractor at the transmitter can be further refined to be also task-oriented. Specifically, the semantic extractor introduced in Section IV can extract all of the semantic information contained by source data. However, the extracted semantic information is not all useful for task decision. Therefore, a more intelligent task-oriented semantic extractor can be introduced, which extracts only effectiveness-related information (i.e., task-related information) of source data to generate effectiveness-related semantic representation for further transmission. Since the decision-maker at the receiver needs only task-related information for decision, this effectiveness-related information extractor is feasible for task-oriented communications. Furthermore, the amount of data gets thoroughly declined as compared to extracting all the semantic information, which also tackles the challenge of limited communication capabilities in the SAGSIN.

In task-oriented communications, a radical departure of communication goal from reconstruction to enhancing ultimate effectiveness represents that the actuation process at the terminal along with abovementioned perception, communication, and computing processes is taken into account in network design, by which a new PCCAIP is developed as demonstrated in Fig. 11. Similar to Section IV, we assume that the raw data have been sampled by intelligent sampler. Specifically, in order to reduce the redundancy caused by effectiveness-unrelated information, we utilize an effectiveness-related information extractor as semantic extractor at the transmitter to remove unrelated information and generate effectiveness-related representation, which will be further coded and modulated by JSCC encoder and modulator to generate transmitted message. At the receiver, different from Section III and Section IV where received messages are reconstructed or recovered by demodulator, decoder, and reconstructor, only one intelligent DL-based actuator receives messages as input and outputs directly the task decision results. Similarly, edge servers and cloud data centers provide knowledge bases and support the distributed learning process of the whole network. By intelligent actuator, the actuation of task will affect the physical world at the transmitter via feedback and effectiveness is thus yielded, which is a closed-loop framework.

In the following subsections, we firstly introduce metrics to measure the ultimate effectiveness of task-oriented communications; then, we detail the EYTs for the task-oriented semantic communication system according to specific task categories, including image processing services, remote control services, and other computing-intensive services.

A. Metrics: Towards Ultimate Effectiveness

Since reconstruction of symbols or semantics no longer exists in task-oriented communications, we cannot use end-to-end metrics, which are based on the distance between reconstructed messages/semantics and the original ones, to measure the performance. Instead, metrics that directly evaluate the ultimate performance of the tasks should be considered, such as the accuracy and preciseness.

1) Accuracy of tasks: The tasks related to classification, such as image/video classification, image retrieval, and audio recognition, necessitates the correctness or accuracy of output results. For these classification tasks, the samples can be coarsely divided into four categories, which are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Denote the number of classified samples as n, and a specific category of samples as n_{TP}, n_{TN}, n_{FP}, or n_{FN}, respectively. Generally, the precision rate $\mathcal{R}_{\text{precision}}$ is defined as

$$\mathcal{R}_{\text{precision}} = \frac{n_{\text{TP}}}{n_{\text{TP}} + n_{\text{FP}}}, \tag{29}$$

and the recall rate $\mathcal{R}_{\text{precision}}$ is defined as

$$\mathcal{R}_{\text{recall}} = \frac{n_{\text{TP}}}{n_{\text{TP}} + n_{\text{FN}}}. \tag{30}$$

[3] In this section, the actuation process mentioned in the references is in fact only decision process mentioned in Section II.

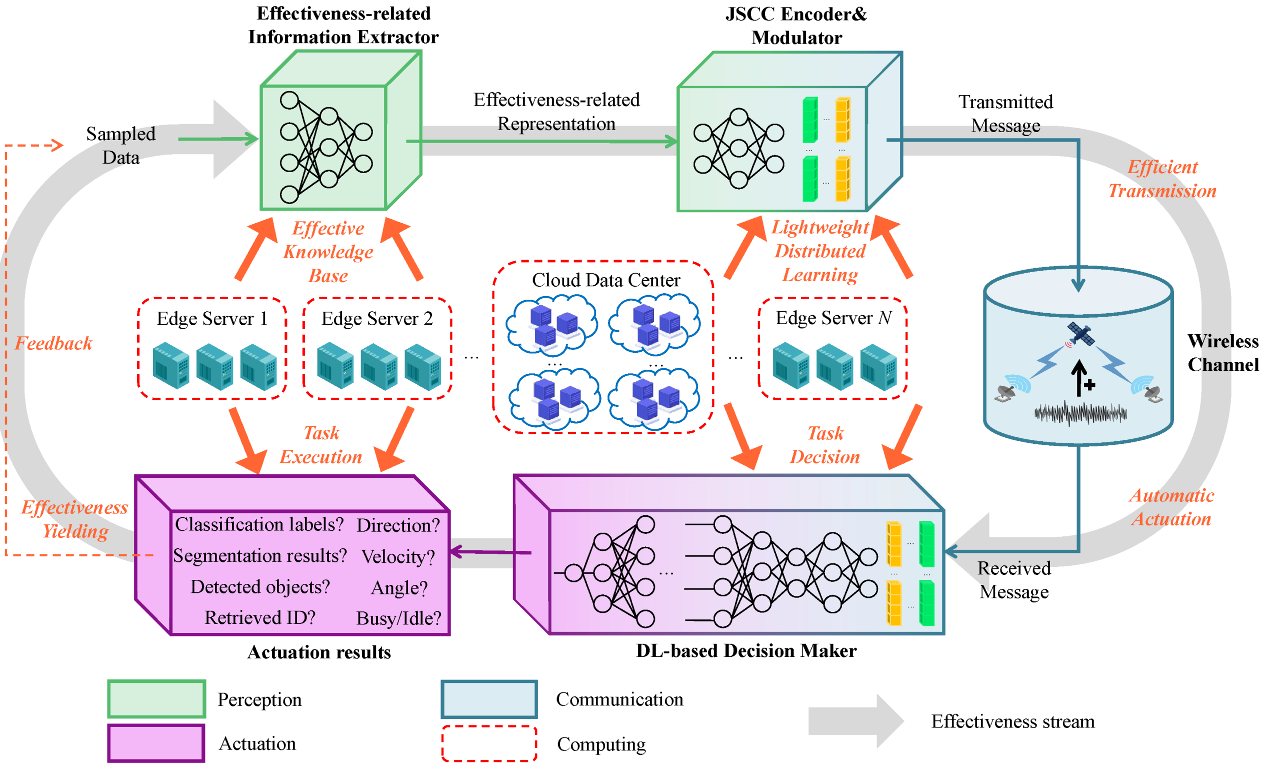


Fig. 11. The framework of perception-communication-computing-actuation-integrated semantic communications in the SAGSIN.

The classification accuracy $\mathcal{R}_{\text{acc}}$ is defined as the ratio of correctly classified samples including TPs and FNs, that is

$$\mathcal{R}_{\text{acc}} = \frac{n_{\text{TP}} + n_{\text{FN}}}{n_{\text{total}}}, \tag{31}$$

where $n_{\text{total}} = n_{\text{TP}} + n_{\text{TN}} + n_{\text{FP}} + n_{\text{FN}}$ is the number of total samples.

The classification accuracy defined in equation (31) describes how many samples of one category is correctly classified, which is also called top-1 accuracy. As a general case of top-1 accuracy, top-k accuracy is defined as the ratio of samples whose real category (also called ground truth) belongs to the categories with highest k degree of confidence.[4]

In specific classification tasks where the negative impact of FPs is comparable to that of TNs, $\mathcal{R}_{\text{accs}}$ is sufficient for serving as ultimate effectiveness metric. However, in the SAGSIN scenario, the cost yielded by different classification results may be remarkably distinct, implying that the impact of FPs and TNs should be considered as different in more general situations. By adopting this idea, the F-measure metric is proposed [178] to endow different weight for FPs and TNs by a hyper parameter β, defined as

$$\mathcal{R}_{\text{Fmeasure}} = \frac{(1+\beta^2)\mathcal{R}_{\text{precision}}\mathcal{R}_{\text{recall}}}{\beta^2\mathcal{R}_{\text{precision}} + \mathcal{R}_{\text{recall}}}. \tag{32}$$

F1-score is defined as a special F-measure when $\beta = 1$, that is

$$\mathcal{R}_{\text{F1score}} = \frac{2\mathcal{R}_{\text{precision}}\mathcal{R}_{\text{recall}}}{\mathcal{R}_{\text{precision}} + \mathcal{R}_{\text{recall}}}. \tag{33}$$

As a more comprehensive accuracy metric, the weighted F1-score combines the accuracy of all categories of samples, which is expressed as

$$\mathcal{R}_{\text{wF1}} = \sum_{i=1}^{m} p_i \mathcal{R}_{\text{F1score},i}, \tag{34}$$

where m is the number of categories, p_i is the ratio of category i among all samples, and $\mathcal{R}_{\text{F1score},i}$ is the F1-score of category i calculated by (33). $\mathcal{R}_{\text{wF1}}$ can serve as an ultimate effectiveness metric for multi-classification tasks.

For object detection and image segmentation tasks, we are interested in specific classified results in the pixel level. That is, the accuracy of these tasks are determined by the ratio of correctly classified pixels in each image, i.e., whether the detected area or segmented area overlaps the ground truth area. The intersection over union (IoU) metric is defined as the ratio of the intersected area by the union area between the task output results n_{output} and the ground truth n_{gt}, that is

$$\mathcal{R}_{\text{IoU}} = \frac{S_{n_{\text{output}} \cap n_{\text{gt}}}}{S_{n_{\text{output}} \cup n_{\text{gt}}}}, \tag{35}$$

where S represents the area scaled by the number of pixels.

It is worth noting that the ground truth is usually labeled manually, which represents the perception of human on the

[4] The degree of confidence is usually the direct output of DL-based classifier, which represents the probabilities of all the categories. In the cases where the number of categories is small, top-1 accuracy is usually adopted. When the number of categories is large, in order to avoid feature overlapping among categories (a sample has similar degree of confidence on several categories), top-k accuracy is more common.

task results. Because these detection and segmentation tasks necessitate evaluation from human during task decision, the IoU between classification and real results (i.e., approximate perception of human) is competent for these tasks.

2) Preciseness of tasks: The tasks related to operation and control, such as automatic driving and UAV route tracking, are subject to the preciseness of the commands generated by the decision-maker, for only the precise and exact commands based on the environment and current situation will yield positive effect on the tasks, while inappropriate commands will hinder the effectiveness of tasks. For these operation tasks, the most determinant factor is system stability, which means that the generated commands must render the state of the system convergent.

Here we take wireless network control system (WNCS) as an example to introduce the preciseness metrics that measure the ultimate effectiveness. Consider a discrete-time process denoted by $X(t) \in \mathbb{R}^n$ which is controlled by command $U(t) \in \mathbb{R}^n$ and interrupted by a noise $Z(t)$ normally distributed with zero mean and covariance matrix $\mathbf{N}$. The process $X(t)$ is assumed as a linear time-invariant system, that is [179], [180]

$$X(t+1) = \mathbf{A}X(t) + \mathbf{B}U(t) + Z(t), \tag{36}$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ are the state transition matrix and control coefficient matrix, respectively.

One of the most common preciseness metric is the MSE of process $X(t)$, with a tacit assumption that the goal of control task is maintaining $X(t)$ as close as zero. Usually we utilize the long-term average MSE as preciseness metric of the WNCS, denoted by

$$\mathcal{R}_{\text{MSE}} = \frac{1}{T}\sum_{t=0}^{T} Q(t), \tag{37}$$

where $Q(t) = E(X(t)X(t)^H)$ represents the MSE of each time slot, and $(\cdot)^H$ represents conjugate transpose operation.

B. Techniques: Towards Effectiveness Yielding

In the task-oriented semantic communication system, the extraction, dissemination, and ultimate actuation is remarkably distinct from the meaning-oriented semantic communication system discussed in Section IV. Task-oriented semantic communication systems are designed towards task actuation at the terminal, and thus the resource consumption of perception, communication, and computing get reduced, which satisfy the resource-efficient demand in the SAGSIN. In this subsection, we review the EYTs applied in the task-oriented communications which facilitate typical services in the SAGSIN, such as image processing, remote control, and other computing-intensive services. A summary of these EYTs can be seen in Table V.

1) EYTs for image processing services: Image processing tasks include image classification, objective detection, image segmentation (semantic segmentation included), and image retrieval, etc, and the main differences among these tasks are illustrated in Fig. 12. These services require only for accurate results at the terminal instead of exact restoration of image itself. Therefore, a joint transmitter-actuator training method can perfectly and efficiently conduct these image processing tasks in the SAGSIN with limited computing resources, benefiting from lite model size and high computing speed.

Image classification services. In [181], the authors study a task-oriented image classification system from a freshness perspective. In this system, the transmitter encodes the images as analog code symbols (i.e., code symbols in a float type) for transmission, and the intelligent classifier deployed at the receiver directly yields the category of a specific image based on received symbols without reconstruction of image. Taking the age of task information (AoTI) as objective function, the optimal arrival rate and analog code length are respectively solved to achieve highest classification accuracy. The authors in [182] theoretically analyze a task-oriented communication system for classification based on robust information bottleneck theory, by which both high relevance of information on tasks and low data distortion are guaranteed, achieving high inference accuracy on images. An adaptive compression method of the task-oriented communication system for classification from a resource allocation perspective is proposed in [183]. According to the relevance on tasks, the compression ratio and resource consumption are jointly optimized by a two-stage iteration algorithm. Simulation results demonstrate that the proposed adaptive compression scheme saves 80% transmission data while maintaining the classification accuracy, implying that both high ultimate effectiveness and high efficiency is achieved.

For the sake of realization of task-oriented communications in the SAGSIN scenarios based on digital modulation, [184] proposes a possible solution which utilizes a spiking-neural-network-based (SNN-based) semantic communication system for image classification task. Compared with other NNs such as CNN, RNN which have been introduced in Section IV-B, a remarkable feature of SNN [201] is that the output of last layer is only a vector $\mathbf{z}$ consisting of binary elements, i.e., $\mathbf{z} \in \{0,1\}^n$, where n is the number of neurons at the last layer of SNN. This remarkable feature implies that SNN directly outputs digital symbols that can be modulated by constellations such as BPSK without quantization procedure, reducing distortion of effectiveness-related information caused by extra quantization. [184] considers a task-oriented semantic codec equipped with SNN over the BSC or BEC, and provides simulation results on the classification accuracy with regards to transition/erasure probability.

To enhance the robustness of task-oriented communication system against adversarial attacks, the authors in [185], [186] focus on anti-semantic-noise semantic communications for robust image classification. The GAN is utilized for generating image data set polluted by semantic noise which cannot be distinguished by human but may mislead the intelligent classifier to yield wrong category results. The authors introduce mask auto-encoder to encounter the negative effect of semantic noise, and propose a semantic codebook method similar to [202], [203] in order to extract only the most significant features (i.e., the most task-relevant semantics) for transmission. By these means the classification accuracy does not decline because of semantic noise compared with noise-

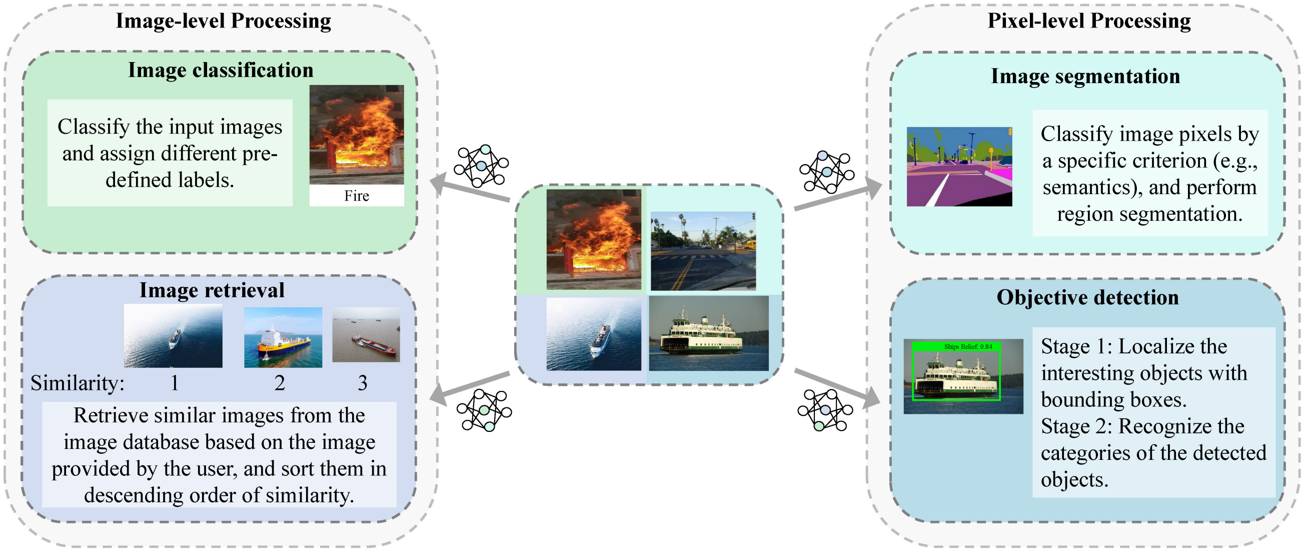


Image-level Processing
Image classification
Classify the input images and assign different pre-defined labels.
Fire
Image retrieval
Similarity: 1 2 3
Retrieve similar images from the image database based on the image provided by the user, and sort them in descending order of similarity.
Pixel-level Processing
Image segmentation
Classify image pixels by a specific criterion (e.g., semantics), and perform region segmentation.
Objective detection
Stage 1: Localize the interesting objects with bounding boxes.
Stage 2: Recognize the categories of the detected objects.

Fig. 12. A comparison of typical image processing tasks that can be facilitated by SAGSIN.

free scenarios.

Objective detection services. [187] considers task-oriented edge video transmission and detection. Specifically, a temporal entropy model and a temporal-spatial integrated model are respectively proposed to extract effectiveness-related information from both current frame and previous frame, according that video sources are temporally continuous and thus the information involved by previous frame can be neglected for feature extraction in the current frame. The ultimate effectiveness in specific tasks, e.g., precision (based on IoU thresholds) of pedestrian distribution detection and multi-device joint objective detection, gets enhanced by task-oriented model design while achieving better rate-effectiveness trade off.

Image segmentation services. [188] studies Internet-of-vehicles-based (IoV-based) task-oriented communication networks for real-time image segmentation. The cars in the front row capture, process, and transmit environment images in real time by swin Transformer encoder, and the cars in the back row conduct image segmentation tasks for received images. The intelligent task-oriented transmitter-actuator pair achieves higher mIoU performance than traditional data-oriented codec pair, showing the superiority of ultimate effectiveness of task-oriented design.

Image retrieval services. The authors in [189] consider image retrieval issue at the edge, where one edge device captures and transmits images to the edge server, and the server conduct retrieval task to predict the identification of the images. Specifically, two task-oriented schemes namely SSCC and JSCC schemes are utilized for effectiveness-oriented feature extraction, coding, and modulation. Moreover, an fully-connected-network-based classifier at receiver directly conduct retrieval tasks without image reconstruction. Numerical results show that JSCC scheme achieves better task accuracy in a higher actuation speed. Further considering the multi-device scenario at the edge, [190] studies collaborative image retrieval problem based on edge computing. Taking two-device case as an example, edge devices capture images and conduct JSCC-based effectiveness-oriented feature extraction independently, and the JSCC outputs (i.e.,effectiveness-related information representation) are transmitted over a shared multiple access channel using NOMA or OMA modulation. The task actuator at the edge server conducts joint image retrieval task by simultaneously processing information from two edge devices. Such collaborative task actuation achieves performance gain in total retrieval accuracy as compared to separate task actuation design.

2) EYTs for remote control services: In remote control applications under the support of WNCS shown in Fig. 13, command generator (i.e., the sensor) should transmit the commands through uplink channel to remote terminal (i.e., the controller) for actuation, and the command messages may be corrupted by noise and interference of highly dynamic downlink channels in the SAGSIN. In order to resist the effect of dynamic erroneous channels and yield high ultimate effectiveness, the commands generated based on the received messages are expected to be precise at each time slot, i.e., commands should maintain the stability and convergence of the controlled process.

From a transmission control perspective, [191] constructs a task-oriented multi-device communication system with multi-modal sources and transmission control. Specifically, in the considered human activity recognition task, sources including room environment and human acceleration information are firstly captured by several end devices and then transmitted over wireless channel. Then, a DL-based processor generates a command of whether to empower the monitoring cameras for video transmission based on the captured sources. By controlling the dissemination process of video sources, the task-oriented system achieves both high total transmission rate and moderate human activity recognition accuracy, which implies a positive ultimate effectiveness.

For multi-device control services, the authors in [192] aim at task-oriented data compression in a multi-agent scenario, con-

TABLE V
SUMMARY OF EFFECTIVENESS YIELDING TECHNIQUES FOR DIFFERENT TASKS

Tasks	**SAGSIN Issue Addressment**	**Related Metrics**	**Specific Services**	**Technical Details**	**Related References**
Image processing	Limited device capabilities	Accuracy of tasks	Image classification	AoTI-based rate-accuracy joint optimization	[181]
				Information bottleneck theory	[182]
				Adaptive compression method	[183]
				SNN-based digital modulation	[184]
				GAN to resist semantic noise, semantic codebook method	[185], [186]
		Accuracy, IoU	Object detection	Temporal-spatial integrated model	[187]
		IoU	Image (semantic) segmentation	Collaborative task decision in internet of vehicles	[188]
		Accuracy of tasks	Image retrieval	JSCC encoder and FCN actuator	[189]
				Collaborative task decision for multi-device scenario	[190]
Remote control	Highly dynamic channels	Accuracy of tasks	Transmission control	Multi-device joint communication-control design	[191]
		Accuracy, MSE	Multi-device scheduling	Task-oriented data compression based on Dec-POMDP	[192]
				Scheduling for multi-modal tasks based on MDP	[193]
		MSE	Practical implementation of remote control	CartPole problem solution	[34]
Other computing-intensive tasks	Limited device capabilities	Accuracy, traditional metrics (e.g., rate, error probability)	Mixed reality	Scene construction according to specific angle of view	[194]
			UAV-based image transmission	Personal-saliency-related semantic triplet	[195]
			Identification	Centralized identification in multi-device scenario	[196]
			Question answering	Multi-user VQA	[197], [198]
				Question answering with memory	[199]
			Sentiment analysis	Semantic triplets	[200]

ducting a joint communication and control optimization. The optimization problem is modeled as a decentralized partially observable MDP (Dec-POMDP) to minimize the distortion caused by compression (a mean absolute error (MAE) metric which is similar to MSE) under constraint of transmission rate. A state-aggregation for information compression algorithm (SAIC) is proposed as near-optimal solution for the optimization problem, yielding low MAE and thus high ultimate effectiveness of the task-oriented communications. Moreover, [193] considers scheduling problem among multi-modal users over time-varying channel, where some of the users conduct task-oriented status updating while others conduct traditional

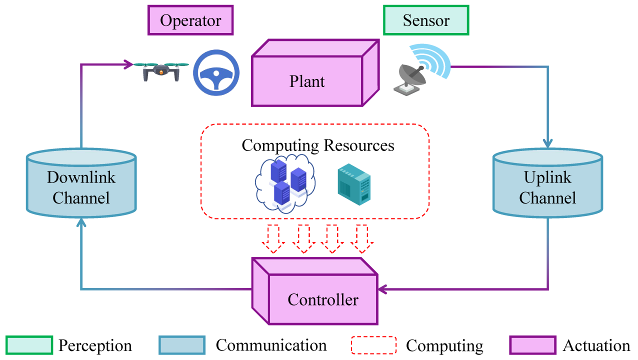


Fig. 13. An illustration of WNCS supported by computing resources distributed in the SAGSIN.

data transmission. In order to jointly optimize the ultimate effectiveness of such multi-user system, both the significance of information (i.e., AoII) for task-oriented user data and the throughput of data-oriented user data should be considered. The joint AoII-throughput optimization problem is formulated as an MDP, and solution verifies a better trade-off between the task accuracy and throughput.

As a practical example of remote control services which can be deployed in the SAGSIN, [34] study both semantic-level (meaning-level) and effectiveness-level communications in remote control over unreliable channel based on dynamic feature compression. The authors consider a classic CartPole control task actuated by two agents, where one agent serves as observer which captures and encodes the CartPole status image into symbols implicitly representing semantic features, and the other agent serves as actuator which generates a binary command of “left” and “right” based on the received semantic symbols. The problem can be modeled as a POMDP, and different levels of communications can be achieved using different reward function. The authors point our that by setting the reward function as minus MSE, the communication system is semantics-oriented; by setting the reward function as the Q-function learned by the actuator, the communication system is effectiveness-oriented, since the Q-function comprehensively reflects the cost of each decision.

3) EYTs for other computing-intensive services: In this part, we review some recent work on task-oriented communication systems designed for other tasks that also consume considerably large amount of computing resources. By a task-oriented design, the resource consumption from communication, computing, and task actuation can get further reduced while ensuring effectiveness of tasks, satisfying the energy-efficiency need for the SAGSIN with limited computing resources.

For instance, [194] studies information sharing between mixed reality (MR) devices (e.g., wearable devices) supported by task-oriented communications. The transmitter user captures and transmits the scene information from his angle of view via semantic extractor, and the receiver user receives the information and utilize a generative network to reproduce the scene from the angle of view of receiver user. Simulation results show that the transmission rate and successful transmission probability at receiver side get enhanced by task-oriented designs. Another example is [195], in which the personal saliency of users in task-oriented communication networks is considered. The authors consider UAV-based image transmission scenarios, where UAVs conduct semantic extraction for captured images to generate a semantic triplet, and personalized user decides whether to download the whole image based on the preference score of the received semantic triplet. By formulating and solving multi-user resource allocation problem, high comprehensive utility values (i.e., the match scores between personal query and received triplets) of all the users can be achieved, hence yielding high ultimate effectiveness. Moreover, an identification-task-oriented multi-user communication system is studied in [196], where cars in the IoV extract the environment semantics and transmit to the central server, and the server conducts identification task for each car based on the integrated semantics from all the cars to ensure high identification accuracy. For tasks where asking and answering occur between transceivers, [197], [198] propose multi-user DL-enabled semantic communication systems for visual question answering (VQA) tasks, and [199] proposes a context-based semantic communication system with memory for question answering tasks, which all increase correct answering ratio of the tasks. Besides, [200] proposes a semantic-triplet-based semantic communication system to complete sentiment analysis and question answering tasks, and the proposed system achieves at least 43.5% and 52% accuracy gains respectively on the two tasks.

C. *Lessons Learned from This Section*

This section elaborates on how a task-oriented communication system is designed by non-symmetric transceivers. Specifically, the core of such non-symmetric transceiver design is to reduce the resource consumption of source processing, data transmission and task decision. An effectiveness-related information extractor is adopted at transmitter side to release the burden of data amount, while an intelligent actuator is utilized to substitute traditional separate reconstructor-actuator design and thus reduce computing resource consumption as compared to reconstruction-based methods. We believe that the task-oriented communication is a promising idea for semantic communications in the SAGSIN for its energy efficiency and design brevity.

VI. Future Challenges on Semantic Communications in the SAGSIN

In this section, we point out some major challenges and future directions on implementation and enhancement for semantic communications in the SAGSIN scenarios.

A. *Precise and Unified Metrics Describing Semantics*

Currently, there still lack precise and unified semantics metrics that can accurately and comprehensively measure multi-dimensional significance/meaning/effectiveness of the source data. On the one hand, the current semantic metrics, especially meaning similarity metrics, can only describe the structural or learned implicit similarity between original and received messages, which is rather inaccurate in "semantic domain" description. One of the major reasons is that semantic information theory is still in its infant age. Although there has been some state-of-the-art works [204] to refine classical semantic information theory proposed by Weaver [205], semantic communication theory is naturally abstract and vague in semantic representation, since it is based on the logical probability which cannot be explicitly calculated. Therefore, scholars are forced to seek alternative metrics that describe semantics from various perspectives, which indirectly while partly define how much semantic information is received by receiver.

On the other hand, the existing semantic metrics are only suitable for a certain type/modal of data, which lack unity and generality. For instance, AoI and its variants reflect only the freshness aspect of semantics, and meaning similarity metrics such as BERT and SSIM are customized for specific modal of sources. On the other hand, considering that massiveness and multi-modality are fundamental characteristics of data in the SAGSIN scenario, the metrics that lack comprehensiveness will challenge the design and optimization on such multi-modal and multi-dimensional source data, since one-dimensional metrics are not suitable for describing the significance/meaning/effectiveness of data of other modals. Therefore, it is imperative to seek a class of precise and unified semantic metrics for semantic communication system design of multi-modal and multi-dimensional data. Our previous works [43], [206] initially make some efforts on unifying the significance metrics to propose the GoT metric as a comprehensive goal-oriented semantic metric, which describes the significance by involving the effect of all the natural cost, actuation cost, and the potential cost caused by actuation. In order to propose comprehensive meaning similarity or effectiveness metrics, some problems should be further solved. For instance, a distance-based metric in a certain domain can be considered to unify current meaning-related metrics that respectively reflect the similarity of image, text, and speech sources. Moreover, to measure the effectiveness yielded by source data (or sampled data), we may quantify the potential cost in the life cycle of a certain message, including cost of total transmission delay, cost of total energy consumption, and cost reduced by a precise actuation (or cost increased by a wrong actuation), and a weighted "unified effectiveness metric" can be finely designed.

B. Semantic Extraction of Multi-modal Data

Semantic extraction is always the initial and essential step in DL-based semantic communications, since the performance of extractor directly affects the performance of semantics transmission and further source reconstruction. Nevertheless, current research focuses on semantic extraction for only specific modal of source data, which cannot directly extract the multi-dimensional semantics from multi-modal data widely distributed in the SAGSIN. In fact, scholars have recently discussed the semantic extraction for multi-modal image-text-fused source data in VQA task [197], while they separately extract the image semantics and text semantics by ResNet and LSTM network, respectively. This separate extractor design for different modal of sources must adopt two or more exclusive NNs, and training such semantic extractor will consume considerably large computing resources, which may not be feasible for edge devices and servers that have access to limited computing resources in the SAGSIN. Therefore, the issue of extracting semantics of massive multi-modal data with only one intelligent NN has to be addressed.

For instance, videos are a typical type of multi-modal data with both image and audio sources. Instead of separating audio information from image frames and extracting the semantics respectively, we may consider image frames and audio information at the same time as semantic-related. We can cut audio signal frame by frame, and utilize state-of-the-art DL techniques such as Transformer to simultaneously extract semantic information of both audio signal and image frame in the current time slot. Because of semantic-related audio and image sources, the parameters of NNs will get reduced as compared to separate design, for semantic representation includes the relation of image and audio of current frame, and thus representation length for current frame is shorter. By integrated extractor design for multi-modal data, a lite semantic coder can be realized and may be directly trained and deployed on small IoT devices in the SAGSIN.

C. Heterogeneous Task-oriented Communication Networks Based on Edge Intelligence

In the SAGSIN, massive and heterogeneous tasks are actuated in a distributed manner by edge devices. Especially, even for the same goal, several cooperative devices may execute different specific intelligent tasks. In the current literature, task-oriented communication network design usually consider homogeneous task executed cooperatively by edge devices [207], [208], which cannot be directly applied for heterogeneous tasks scenario, since the former design consider each device as achieving same goal. The only exceptions are [162], [209], where both semantic communication users and technical communication users work to achieve heterogeneous goals (i.e., both semantic and bit-level reconstruction). However, in the practical applications, tasks are more complex and different users may require different level of quality of service (QoS) or quality of experience (QoE) in task actuation, which calls for more intelligent design of heterogeneous task-oriented communication networks based on edge intelligence.

We may consider a joint detection-classification heterogeneous task which is executed by multiple edge devices distributed in the SAGSIN. Specifically, some of devices detect the objects, while the others classify them based on both their own perception images and detection results by detecting devices. Since the classifier input is multi-modal, i.e., consisting of both detection results and perception images, we have two methods for classifier implementation. Firstly, we can separate the classifier into perception-only and detection-based-only devices, respectively processing the perception images and detection results. Another way is that we assume that all the classifier devices receive both inputs to execute classification tasks. In order to jointly train these detection and classification devices, federated edge learning can be utilized as a solution. However, considering that the goals of devices are different, we can firstly conduct clustered training for devices with the same task or goal in the first training phase, and secondly fine-tune the parameter among devices with different goals in the second training phase. By such two-phase training, heterogeneous tasks can be actuated cooperatively to yield high ultimate effectiveness.

D. Joint Decision-Execution Design for the Actuator

In the SAGSIN, task execution at the terminals will directly affect the environment statuses, which is the direct effectiveness yielded by the decision commands generated from actuator. However, the current execution modules (or techniques) are principally assumed as separated from the decision-maker design, which is still way from the envisioned PCCAIP in Fig. 3 of Section II. For instance, the current task-oriented image classification decision-maker only gives the category results on a certain image without considering how the classification results facilitate task execution to affect the external environment. The only exception is [34], where the intelligent actuator both gives decision results (i.e., left or right) and conducts task execution to complete CartPole game. The real tasks facilitated by the SAGSIN is far more complex and heterogeneous than simple CartPole control, which necessitates more intelligent and comprehensive techniques of joint decision-execution realization for the actuator in order to directly yield ultimate effectiveness.

Here we study an abnormal status observation and rescue task to demonstrate the advantage of such joint decision-actuation design. The observers distributed in the network monitor the environment, and when detecting abnormal statuses, the observer will transmit them using SPTs, METs, or EYTs. The terminals receive the statuses using a joint reconstruction & task-decision & task-execution technique to directly conduct rescue task, e.g., sending rescuers to dangerous areas for sinking ships. The state-of-the-art works reviewed in Section V mainly focus on the monitoring steps of this task in order to enhance accuracy of the observation. However, in joint decision-execution design, we are not only interested in the accuracy of observation but also the affect of the decision (which is based on observation results) caused by task execution, e.g., recovery of abnormal statuses due to timely rescue. When several observed objects are detected as abnormal at the same time but the terminal can only support one of the object to be rescued, the actuator should be finely designed to find out the decision yielding the most ultimate effectiveness via considering the different level of significance of all the objects. In such cases, we must resort to joint design instead of only enhancing the accuracy of observation.

VII. Conclusion

In this survey, we have comprehensively reviewed the implementation of three types of semantic communication systems in the SAGSIN according to the PCCAIP. Starting with significance-oriented semantic communication systems, we introduce metrics measuring the significance of source data and review SPTs including sampling, coding, and modulation policy designs aiming at enhancing the performance in capturing data significance. Then, we review meaning-oriented semantic communication systems by elaborating on meaning similarity metrics, DNN-based methods, and METs focusing on semantic reconstruction performance improvement. Next, we discuss a newly emerging semantic communication approach called effectiveness-oriented semantic communication systems by introducing effectiveness metrics and studying EYTs intended to increase ultimate effectiveness of task actuation. Finally, we propose several research challenges on implementation of semantic communications in the SAGSIN which will draw further insight research interests.

References

[1] B. Aazhang, M. Juntti, R. Kantola, P. Kyösti, S. LaValle, C. Lima, M. Matinmikko-Blue, T. Ojala, A. Pouttu, A. Pärssinen, S. Yrjola, P. Ahokangas, H. Alves, M.-S. Alouini, J. Beek, H. Benn, M. Bennis, J. Belfiore, E. Strinati, and E. Peltonen, "Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper)," *6G Flagship University of Oulu Finland*, Sep. 2019. I-A

[2] X. Cheng, Z. Huang, and L. Bai, "Channel nonstationarity and consistency for beyond 5G and 6G: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1634–1669, Jun. 2022. I-A

[3] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 11 891–11 915, Mar. 2021. I-A

[4] H. Guo, J. Li, J. Liu, N. Tian, and N. Kato, "A survey on space-air-ground-sea integrated network security in 6G," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 53–87, Nov. 2022. I-A

[5] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2714–2741, May 2018. I-A

[6] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. Mendoza Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani, E. Lagunas, and B. Ottersten, "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2633–2672, Aug. 2022. I-A

[7] S. Gu, Q. Zhang, and W. Xiang, "Coded storage-and-computation: A new paradigm to enhancing intelligent services in space-air-ground integrated networks," *IEEE Wireless Communications*, vol. 27, no. 6, pp. 44–51, Dec. 2020. I-A

[8] T. Hong, M. Lv, S. Zheng, and H. Hong, "Key technologies in 6G SAGS IoT: Shape-adaptive antenna and radar-communication integration," *IEEE Network*, vol. 35, no. 5, pp. 150–157, Nov. 2021. I-A

[9] D. Liu, J. Zhang, J. Cui, S.-X. Ng, R. G. Maunder, and L. Hanzo, "Deep learning aided routing for space-air-ground integrated networks relying on real satellite, flight, and shipping data," *IEEE Wireless Communications*, vol. 29, no. 2, pp. 177–184, Apr. 2022. I-A

[10] L. Bai, R. Han, J. Liu, J. Choi, and W. Zhang, "Relay-aided random access in space-air-ground integrated networks," *IEEE Wireless Communications*, vol. 27, no. 6, pp. 37–43, Dec. 2020. I-A

[11] Y. Wang, Z. Su, J. Ni, N. Zhang, and X. Shen, "Blockchain-empowered space-air-ground integrated networks: Opportunities, challenges, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 160–209, Dec. 2021. I-A

[12] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948. I-A, IV-B

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proceedings of ICC '93 - IEEE International Conference on Communications*, vol. 2, May 1993, pp. 1064–1070 vol.2. I-A

[14] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962. I-A

[15] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes," in *2008 IEEE International Symposium on Information Theory*, Jul. 2008, pp. 1173–1177. I-A

[16] J. Perry, P. A. Lannucci, K. Fleming, H. Balakrishnan, and D. Shah, "Spinal codes," *ACM Sigcomm Computer Communication Review*, vol. 42, no. 4, pp. 49–60, Aug. 2012. I-A

[17] W. Weaver, “Recent contributions to the mathematical theory of communication,” *ETC: A review of general semantics*, vol. 10, no. 4, pp. 261–281, Jul. 1953. I-A, I-A

[18] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, B. Soret, T. Soleymani, and K. H. Johansson, “Semantic communications in networked systems: A data significance perspective,” *IEEE Network*, vol. 36, no. 4, pp. 233–240, Oct. 2022. I-B

[19] K. Niu, J. Dai, S. Yao, S. Wang, Z. Si, X. Qin, and P. Zhang, “A paradigm shift toward semantic communications,” *IEEE Communications Magazine*, vol. 60, no. 11, pp. 113–119, Aug. 2022. I-B

[20] G. Shi, Y. Xiao, Y. Li, and X. Xie, “From semantic communication to semantic-aware networking: Model, architecture, and open problems,” *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, Aug. 2021. I-B

[21] X. Luo, H.-H. Chen, and Q. Guo, “Semantic communications: Overview, open issues, and future research directions,” *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210–219, Jan. 2022. I-B

[22] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, and K. Huang, “What is semantic communication? a view on conveying meaning in the era of machine intelligence,” *Journal of Communications and Information Networks*, vol. 6, pp. 336–371, Jan. 2021. I-B

[23] Z. Qin, X. Tao, J. Lu, and G. Y. Li, “Semantic communications: Principles and challenges,” 2022. [Online]. Available: http://arxiv.org/abs/2201.01389 I-B

[24] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, “Less data, more knowledge: Building next generation semantic communication networks,” 2022. [Online]. Available: http://arxiv.org/abs/2211.14343 I-B

[25] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, “Beyond transmitting bits: Context, semantics, and task-oriented communications,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, Nov. 2022. I-B

[26] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, “Semantic communications for future Internet: Fundamentals, applications, and challenges,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213–250, Nov. 2022. I-B

[27] W. Xu, Z. Yang, D. W. K. 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 Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 9–39, Jan. 2023. I-B

[28] M. Chafii, L. Bariah, S. Muhaidat, and M. Debbah, “Twelve scientific challenges for 6G: Rethinking the foundations of communications theory,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 868–904, Feb. 2023. I-B

[29] Z. Wang, Z. Zhou, H. Zhang, G. Zhang, H. Ding, and A. Farouk, “AI-based cloud-edge-device collaboration in 6G space-air-ground integrated power IoT,” *IEEE Wireless Communications*, vol. 29, no. 1, pp. 16–23, Feb. 2022. II-B

[30] S. Yu, X. Gong, Q. Shi, X. Wang, and X. Chen, “EC-SAGINs: Edge-computing-enhanced space–air–ground-integrated networks for internet of vehicles,” *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5742–5754, Apr. 2021. II-B

[31] S. Gu, Q. Zhang, and W. Xiang, “Coded storage-and-computation: A new paradigm to enhancing intelligent services in space-air-ground integrated networks,” *IEEE Wireless Communications*, vol. 27, no. 6, pp. 44–51, Dec. 2020. II-B

[32] S. Gu, Y. Wang, N. Wang, and W. Wu, “Intelligent optimization of availability and communication cost in satellite-UAV mobile edge caching system with fault-tolerant codes,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1230–1241, Dec. 2020. II-B

[33] Z. Zhang, W. Zhang, and F.-H. Tseng, “Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques,” *IEEE network*, vol. 33, no. 1, pp. 70–76, Jan. 2019. II-B

[34] P. Talli, F. Pase, F. Chiariotti, A. Zanella, and M. Zorzi, “Semantic and effective communication for remote control tasks with dynamic feature compression,” in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2023, pp. 1–6. II-C, V, V-B2, VI-D

[35] ITU-T, “Representative use cases and key network requirements for network 2030,” FG-NET2030-Sub-G1, Tech. Rep. FG-NET2030-Sub-G1, Jan. 2020. II-C

[36] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, “Minimizing age of information in vehicular networks,” in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Jun. 2011, pp. 350–358. III-A, I

[37] S. Kaul, R. Yates, and M. Gruteser, “Real-time status: How often should one update?” in *2012 Proceedings IEEE INFOCOM*, Mar. 2012, pp. 2731–2735. III-A, I

[38] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “Age and value of information: Non-linear age case,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 326–330. III-A, I

[39] Y. Sun and B. Cyr, “Sampling for data freshness optimization: Non-linear age functions,” *Journal of Communications and Networks*, vol. 21, no. 3, pp. 204–219, Jun. 2019. III-A, I, II, III-B1

[40] J. Zhong, R. D. Yates, and E. Soljanin, “Two freshness metrics for local cache refresh,” in *2018 IEEE International Symposium on Information Theory (ISIT)*, Aug. 2018, pp. 1924–1928. I, III-A, II, III-B1

[41] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “The age of incorrect information: A new performance metric for status updates,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2215–2228, Jul. 2020. I, III-A, II, III-B1

[42] X. Zheng, S. Zhou, and Z. Niu, “Urgency of information for context-aware timely status updates in remote control systems,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7237–7250, Jul. 2020. I, III-A

[43] A. Li, S. Wu, and S. Sun, “Goal-oriented tensor: Beyond AoI towards semantics-empowered goal-oriented communications,” May 2023. I, III-A, VI-A

[44] F. Chiariotti, J. Holm, A. E. Kalør, B. Soret, S. K. Jensen, T. B. Pedersen, and P. Popovski, “Query age of information: Freshness in pull-based communication,” *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 1606–1622, Jan. 2022. III-A

[45] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, “Age of information: An introduction and survey,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1183–1210, Mar. 2021. III-A

[46] Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. E. Koksal, and N. B. Shroff, “Update or wait: How to keep your data fresh,” in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, Apr. 2016, pp. 1–9. III-B1, II, III-B1, III-B1

[47] B. Zhou and W. Saad, “Optimal sampling and updating for minimizing age of information in the Internet of things,” in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018, pp. 1–6. II, III-B1

[48] Y. Sun, Y. Polyanskiy, and E. Uysal, “Sampling of the Wiener process for remote estimation over a channel with random delay,” *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1118–1135, Aug. 2019. II, III-B1, III-B1, III-B1

[49] T. Z. Ornee and Y. Sun, “Sampling and remote estimation for the Ornstein-Uhlenbeck process through queues: Age of information and beyond,” *IEEE/ACM Transactions on Networking*, vol. 29, no. 5, pp. 1962–1975, May 2021. II, III-B1, III-B1

[50] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, “Age-optimal sampling and transmission scheduling in multi-source systems,” in *Proceedings of the 20th ACM International Symposium On Mobile Ad Hoc Networking and Computing (MOBIHOC ‘19)*, Jul. 2019, pp. 121–130. II, III-B1

[51] Y. Chen and A. Ephremides, “Minimizing age of incorrect information for unreliable channel with power constraint,” in *2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021, pp. 1–6. II, III-B1

[52] A. Maatouk, M. Assaad, and A. Ephremides, “Semantics-empowered communications through the age of incorrect information,” in *ICC 2022 - IEEE International Conference on Communications*, May 2022, pp. 3995–4000. II, III-B1

[53] Y. Chen and A. Ephremides, “Minimizing age of incorrect information in the presence of timeout,” 2022. [Online]. Available: http://arxiv.org/abs/2207.02926 II, III-B1

[54] K. Bountrogiannis, A. Ephremides, P. Tsakalides, and G. Tzagkarakis, “Age of incorrect information with hybrid arq under a resource constraint for n-ary symmetric markov sources,” 2023. [Online]. Available: http://arxiv.org/abs/2303.18128 II, III-B1

[55] S. C. Bobbili, P. Parag, and J.-F. Chamberland, “Real-time status updates with perfect feedback over erasure channels,” *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5363–5374, Jul. 2020. II, III-B2, III-B2

[56] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, “On timely channel coding with hybrid ARQ,” in *2019 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2019, pp. 1–6. II, III-B2

[57] M. Xie, Q. Wang, J. Gong, and X. Ma, “Age and energy analysis for LDPC coded status update with and without ARQ,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 388–10 400, Apr. 2020. II, III-B2, III-B2

[58] J. You, S. Wu, Y. Deng, J. Jiao, and Q. Zhang, “An age optimized hybrid ARQ scheme for Polar codes via Gaussian approximation,” *IEEE Wireless Communications Letters*, vol. 10, no. 10, pp. 2235–2239, Jul. 2021. II, III-B2, III-B2

[59] A. Li, S. Wu, J. Jiao, N. Zhang, and Q. Zhang, “Age of information with hybrid-ARQ: A unified explicit result,” *IEEE Transactions on Communications*, vol. 70, no. 12, pp. 7899–7914, Oct. 2022. II, III-B2, III-B2, III-B2

[60] Y. Wang, S. Wu, D. Li, J. Jiao, and Q. Zhang, “Age-optimal IR-HARQ design in the presence of non-trivial propagation delay,” in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2019, pp. 1–6. II, III-B2, III-B2, III-B2

[61] D. Li, S. Wu, Y. Wang, J. Jiao, and Q. Zhang, “Age-optimal HARQ design for freshness-critical satellite-IoT systems,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2066–2076, Dec. 2020. II, III-B2, III-B2, III-B2

[62] D. Li, S. Wu, J. Jiao, N. Zhang, and Q. Zhang, “Age-oriented transmission protocol design in space-air-ground integrated networks,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 5573–5585, Jan. 2022. II, III-B2, III-B2

[63] S. Meng, S. Wu, A. Li, J. Jiao, N. Zhang, and Q. Zhang, “Analysis and optimization of the HARQ-based Spinal coded timely status update system,” *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6425–6440, Aug. 2022. II, III-B2

[64] Y. Deng, S. Wu, J. You, J. Jiao, N. Zhang, and Q. Zhang, “Optimizing age of information in Polar coded status update system,” *IEEE Internet of Things Journal*, pp. 1–1, Jun. 2023. II, III-B2

[65] B. R. Sharan, S. Deshmukh, S. R. B. Pillai, and B. Beferull-Lozano, “Energy efficient AoI minimization in opportunistic NOMA/OMA broadcast wireless networks,” *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 1009–1022, Dec. 2021. II, III-B3, III-B3

[66] S. Wu, C. Guo, Z. Deng, J. Jiao, N. Zhang, and Q. Zhang, “Optimizing age of information in adaptive NOMA/OMA/cooperative-SWIPT-NOMA system,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 11 125–11 138, Jul. 2022. II, III-B3, III-B3

[67] S. Wu, Z. Deng, A. Li, J. Jiao, N. Zhang, and Q. Zhang, “Minimizing age-of-information in HARQ-CC aided NOMA systems,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 2, pp. 1072–1086, Sep. 2022. II, III-B3, III-B3, III-B3

[68] S. Liao, J. Jiao, S. Wu, R. Lu, and Q. Zhang, “Age-optimal power allocation scheme for NOMA-based S-IoT downlink network,” in *ICC 2021-IEEE International Conference on Communications*, Jun. 2021, pp. 1–6. II, III-B3, III-B3, III-B3, III-B3

[69] J. Jiao, H. Hong, Y. Wang, S. Wu, R. Lu, and Q. Zhang, “Age-optimal downlink NOMA resource allocation for satellite-based IoT network,” *IEEE Transactions on Vehicular Technology*, Apr. 2023. II, III-B3, III-B3

[70] Z. Shi, H. Ding, S. Ma, K.-W. Tam, and S. Pan, “Inverse moment matching based analysis of cooperative HARQ-IR over time-correlated Nakagami fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3812–3828, Aug. 2017. III-B2

[71] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, Jul. 2002, pp. 311–318. IV-A1, III

[72] R. Vedantam, C. Lawrence Zitnick, and D. Parikh, “CIDEr: Consensus-based image description evaluation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Jun. 2015, pp. 4566–4575. IV-A1, III

[73] Y. Wang, M. Chen, T. Luo, W. Saad, D. Niyato, H. V. Poor, and S. Cui, “Performance optimization for semantic communications: An attention-based reinforcement learning approach,” *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2598–2613, Jul. 2022. IV-A1, III, IV, IV-C3, IV-C3

[74] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, “BERTSCORE: Evaluating text generation with BERT,” 2019. [Online]. Available: https://arxiv.org/abs/1904.09675 III, IV-A1, IV-A1

[75] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, “Deep learning enabled semantic communication systems,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, Apr. 2021. III, IV-A1

[76] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, Nov. 2003, pp. 1398–1402. III, IV-A2, IV-A2

[77] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Jun. 2018, pp. 586–595. III, IV-A2

[78] 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,” *Advances in neural information processing systems*, vol. 30, Dec. 2017. III, IV-A2, IV-A2

[79] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, “Demystifying MMD GANs,” 2018. [Online]. Available: https://arxiv.org/abs/1801.01401 III, IV-A2, IV-A2

[80] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, “Wireless semantic communications for video conferencing,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 230–244, Nov. 2022. III, IV-A2, IV, IV-C2, IV-C2

[81] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, Jun. 2006. III, IV-A3

[82] “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” *ITU-T recommendation P.862*, Feb. 2001. III, IV-A3, IV-A3

[83] R. Kubichek, “Mel-cepstral distance measure for objective speech quality assessment,” in *Proceedings of IEEE pacific rim conference on communications computers and signal processing*, vol. 1, May 1993, pp. 125–128. III, IV-A3

[84] M. Bińkowski, J. Donahue, S. Dieleman, A. Clark, E. Elsen, N. Casagrande, L. C. Cobo, and K. Simonyan, “High fidelity speech synthesis with adversarial networks,” 2019. [Online]. Available: https://arxiv.org/abs/1909.11646 III, IV-A3, IV-A3

[85] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional Transformers for language understanding,” in *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2019), Vol. 1*, Jan. 2019, pp. 4171–4186. IV-A1

[86] Z. Weng and Z. Qin, “Semantic communication systems for speech transmission,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2434–2444, Jun. 2021. IV-A3, 5

[87] K. Lu, R. Li, X. Chen, Z. Zhao, and H. Zhang, “Reinforcement learning-powered semantic communication via semantic similarity,” 2021. [Online]. Available: https://arxiv.org/abs/2108.12121 2, 3

[88] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, and H. Zhang, “Rethinking modern communication from semantic coding to semantic communication,” *IEEE Wireless Communications*, vol. 30, no. 1, pp. 158–164, Feb. 2023. 2, 3, IV, IV-C1

[89] N. Farsad, M. Rao, and A. Goldsmith, “Deep learning for joint source-channel coding of text,” in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Apr. 2018, pp. 2326–2330. IV-B, 5

[90] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *The Journal of physiology*, vol. 160, no. 1, p. 106, Jan. 1962. IV-B1

[91] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, Apr. 1980. IV-B1

[92] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998. IV-B1

[93] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017. IV-B1

[94] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, Sep. 2014, pp. 818–833. IV-B1

[95] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014. [Online]. Available: https://arxiv.org/abs/1409.1556 IV-B1

[96] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9. IV-B1

[97] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on*

computer vision and pattern recognition, Jun. 2015, pp. 3431–3440. IV-B1

[98] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Jun. 2016, pp. 770–778. IV-B1, IV-B1

[99] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Jul. 2017, pp. 4700–4708. IV-B1

[100] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Jul. 2017, pp. 2881–2890. IV-B1

[101] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Jul. 2017, pp. 1125–1134. IV-B1

[102] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982. IV-B2

[103] M. Jordan, “Serial order: a parallel distributed processing approach. Technical report, June 1985-March 1986,” California Univ., San Diego, La Jolla (USA). Inst. for Cognitive Science, Tech. Rep., May 1986. IV-B2

[104] J. L. Elman, “Finding structure in time,” *Cognitive science*, vol. 14, no. 2, pp. 179–211, Mar. 1990. IV-B2

[105] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. IV-B2, IV-B2

[106] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997. IV-B2

[107] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM networks,” in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 4, Jul. 2005, pp. 2047–2052. IV-B2

[108] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE international conference on acoustics, speech and signal processing*, May 2013, pp. 6645–6649. IV-B2

[109] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” 2014. [Online]. Available: https://arxiv.org/abs/1406.1078 IV-B2

[110] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, Dec. 2017. IV-B3

[111] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2020. [Online]. Available: https://arxiv.org/abs/2010.11929 IV-B3

[112] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, Aug. 2020, pp. 213–229. IV-B3

[113] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin Transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, Oct. 2021, pp. 10 012–10 022. IV-B3

[114] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, vol. 27, Dec. 2014, pp. 2672–2680. IV-B4

[115] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014. [Online]. Available: https://arxiv.org/abs/1411.1784 IV-B4

[116] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2015. [Online]. Available: https://arxiv.org/abs/1511.06434 IV-B4

[117] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*, Jul. 2017, pp. 214–223. IV-B4

[118] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-attention generative adversarial networks,” in *International conference on machine learning*, May 2019, pp. 7354–7363. IV-B4

[119] A. Brock, J. Donahue, and K. Simonyan, “Large scale GAN training for high fidelity natural image synthesis,” 2018. [Online]. Available: https://arxiv.org/abs/1809.11096 IV-B4

[120] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Jun. 2019, pp. 4401–4410. IV-B4

[121] J. Kong, J. Kim, and J. Bae, “Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 022–17 033, Dec. 2020. IV-B4

[122] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016. [Online]. Available: https://arxiv.org/abs/1609.02907 3

[123] Z. Lu, Y. Xiao, Z. Sun, Y. Li, G. Shi, X. Chenk, M. Bennis, and H. Poor, “Adversarial learning for implicit semantic-aware communications,” 2023. [Online]. Available: http://arxiv.org/abs/2301.11589 3

[124] C. K. Thomas and W. Saad, “Neuro-symbolic artificial intelligence (AI) for intent based semantic communication,” in *GLOBECOM 2022-2022 IEEE Global Communications Conference*, Dec. 2022, pp. 2698–2703. 3, 5

[125] ——, “Neuro-symbolic causal reasoning meets signaling game for emergent semantic communications,” 2022. [Online]. Available: https://arxiv.org/abs/2210.12040 3, 5

[126] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, “Towards semantic communications: Deep learning-based image semantic coding,” 2022. [Online]. Available: http://arxiv.org/abs/2208.04094 3, IV, IV-C2

[127] J. Huang, D. Li, C. Huang, X. Qin, and W. Zhang, “Joint task and data oriented semantic communications: A deep separate source-channel coding scheme,” *IEEE Internet of Things Journal*, pp. 1–1, Jul. 2023. 4, IV, IV-C2

[128] J.-H. Lee, D.-H. Lee, E. Sheen, T. Choi, J. Pujara, and J. Kim, “Seq2Seq-SC: End-to-end semantic communication systems with pre-trained language model,” 2022. [Online]. Available: https://arxiv.org/abs/2210.15237 4, 5

[129] F. Liu, W. Tong, Y. Yang, Z. Sun, and C. Guo, “Task-oriented image semantic communication based on rate-distortion theory,” 2022. [Online]. Available: http://arxiv.org/abs/2201.10929 4

[130] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, “Deep source-channel coding for sentence semantic transmission with HARQ,” *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5225–5240, Jun. 2022. 4, IV, IV-C2, IV-C2

[131] X. Luo, Z. Chen, M. Tao, and F. Yang, “Encrypted semantic communication using adversarial training for privacy preserving,” *IEEE Communications Letters*, Apr. 2023. 4, IV, IV-C4

[132] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, “Semantic-preserved communication system for highly efficient speech transmission,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 245–259, Nov. 2022. 4, 5

[133] S. Imran, G. Charan, and A. Alkhateeb, “Environment semantic aided communication: A real world demonstration for beam prediction,” 2023. [Online]. Available: http://arxiv.org/abs/2302.06736 IV-C1, IV

[134] 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 Journal on Selected Areas in Communications*, vol. 41, no. 7, pp. 2025–2040, May 2023. IV-C1, IV

[135] Z. Qin, F. Gao, B. Lin, X. Tao, G. Liu, and C. Pan, “A generalized semantic communication system: From sources to channels,” *IEEE Wireless Communications*, vol. 30, no. 3, pp. 18–26, Jun. 2023. IV, IV-C1

[136] X. Peng, Z. Qin, D. Huang, X. Tao, J. Lu, G. Liu, and C. Pan, “A robust deep learning enabled semantic communication system for text,” in *GLOBECOM 2022-2022 IEEE Global Communications Conference*, Dec. 2022, pp. 2704–2709. IV, IV-C1

[137] J. Dai, S. Wang, K. Yang, K. Tan, X. Qin, Z. Si, K. Niu, and P. Zhang, “Adaptive semantic communications: Overfitting the source and channel for profit,” 2022. [Online]. Available: https://arxiv.org/abs/2211.04339 IV, IV-C1, IV-C1

[138] K. Yang, S. Wang, J. Dai, K. Tan, K. Niu, and P. Zhang, “WITT: A wireless image transmission transformer for semantic communications,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5. IV, IV-C1, IV-C1

[139] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gunduz, “Deep joint source-channel coding for semantic communications,” 2022. [Online]. Available: https://arxiv.org/abs/2211.08747 IV, IV-C1

[140] S. Wang, J. Dai, X. Qin, Z. Si, K. Niu, and P. Zhang, "Improved nonlinear transform source-channel coding to catalyze semantic communications," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–16, Aug. 2023. IV, IV-C1

[141] Y. Bo, Y. Duan, S. Shao, and M. Tao, "Learning based joint coding-modulation for digital semantic communication systems," in *2022 14th International Conference on Wireless Communications and Signal Processing (WCSP)*, Nov. 2022, pp. 1–6. IV, IV-C1

[142] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 142–153, Nov. 2020. IV, IV-C1

[143] 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 Journal of the Communications Society*, vol. 3, pp. 1076–1089, Jul. 2022. IV, IV-C2

[144] D. B. Kurka and D. Gündüz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 178–193, Apr. 2020. IV, IV-C2

[145] M. Yang and H.-S. Kim, "Deep joint source-channel coding for wireless image transmission with adaptive rate control," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 5193–5197. IV, IV-C2

[146] J. Dai, S. Wang, K. Tan, Z. Si, X. Qin, K. Niu, and P. Zhang, "Nonlinear transform source-channel coding for semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 8, pp. 2300–2316, Jun. 2022. IV, IV-C2, IV-C2

[147] S. Wang, J. Dai, Z. Liang, K. Niu, Z. Si, C. Dong, X. Qin, and P. Zhang, "Wireless deep video semantic transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 214–229, Nov. 2022. IV, IV-C2, IV-C2

[148] B. Zhang, Z. Qin, and G. Y. Li, "Semantic communications with variable-length coding for extended reality," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–14, Aug. 2023. IV, IV-C2

[149] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, and R. Q. Hu, "Cognitive semantic communication systems driven by knowledge graph," in *ICC 2022-IEEE International Conference on Communications*, May 2022, pp. 4860–4865. IV, IV-C3, IV-C3

[150] B. Wang, R. Li, J. Zhu, Z. Zhao, and H. Zhang, "Knowledge enhanced semantic communication receiver," *IEEE Communications Letters*, vol. 27, no. 7, pp. 1794–1798, May 2023. IV, IV-C3

[151] S. Seo, J. Park, S.-W. Ko, J. Choi, M. Bennis, and S.-L. Kim, "Toward semantic communication protocols: A probabilistic logic perspective," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2670–2686, Jun. 2023. IV, IV-C3

[152] J. Choi, S. W. Loke, and J. Park, "A unified approach to semantic information and communication based on probabilistic logic," *IEEE Access*, vol. 10, pp. 129 806–129 822, Dec. 2022. IV, IV-C3

[153] D. Wheeler, E. E. Tripp, and B. Natarajan, "Semantic communication with conceptual spaces," *IEEE Communications Letters*, Dec. 2022. IV, IV-C3

[154] Z. Li, X. Liu, G. Nan, J. Zhou, X. Lyu, Q. Cui, and X. Tao, "Boosting physical layer black-box attacks with semantic adversaries in semantic communications," 2023. [Online]. Available: https://arxiv.org/abs/2303.16523 IV, IV-C4

[155] G. Nan, Z. Li, J. Zhai, Q. Cui, G. Chen, X. Du, X. Zhang, X. Tao, Z. Han, and T. Q. S. Quek, "Physical-layer adversarial robustness for deep learning-based semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2592–2608, Jun. 2023. IV, IV-C4

[156] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, "Vulnerabilities of deep learning-driven semantic communications to backdoor (Trojan) attacks," in *2023 57th Annual Conference on Information Sciences and Systems (CISS)*, Apr. 2023, pp. 1–6. IV, IV-C4

[157] T.-Y. Tung and D. Gunduz, "Deep joint source-channel and encryption coding: Secure semantic communications," 2022. [Online]. Available: https://arxiv.org/abs/2208.09245 IV, IV-C4

[158] T. Han, J. Tang, Q. Yang, Y. Duan, Z. Zhang, and Z. Shi, "Generative model based highly efficient semantic communication approach for image transmission," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5. IV, IV-C4

[159] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, M. Guizani, and D. I. Kim, "Rethinking wireless communication security in semantic internet of things," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 36–43, Jun. 2023. IV, IV-C4

[160] Z. Yang, M. Chen, G. Li, Y. Yang, and Z. Zhang, "Secure semantic communications: Fundamentals and challenges," 2023. [Online]. Available: https://arxiv.org/abs/2301.01421 IV-C4

[161] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, Apr. 2021. 5

[162] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, "Heterogeneous semantic and bit communications: A semi-NOMA scheme," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 155–169, Nov. 2022. 5, VI-C

[163] W. Li, H. Liang, C. Dong, X. Xu, P. Zhang, and K. Liu, "Non-orthogonal multiple access enhanced multi-user semantic communication," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, Aug. 2023. 5

[164] H. Hu, X. Zhu, F. Zhou, W. Wu, R. Q. Hu, and H. Zhu, "One-to-many semantic communication systems: Design, implementation, performance evaluation," *IEEE Communications Letters*, vol. 26, no. 12, pp. 2959–2963, Sep. 2022. 5

[165] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, May 2019. 5

[166] D. B. Kurka and D. Gündüz, "Successive refinement of images with deep joint source-channel coding," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2019, pp. 1–5. 5

[167] M. Ding, J. Li, M. Ma, and X. Fan, "SNR-adaptive deep joint source-channel coding for wireless image transmission," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, pp. 1555–1559. 5

[168] C. Dong, H. Liang, X. Xu, S. Han, B. Wang, and P. Zhang, "Semantic communication system based on semantic slice models propagation," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 202–213, Nov. 2022. 5

[169] J. Dai, P. Zhang, K. Niu, S. Wang, Z. Si, and X. Qin, "Communication beyond transmitting bits: Semantics-guided source and channel coding," *IEEE Wireless Communications*, pp. 1–8, Aug. 2022. 5

[170] 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 Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 170–185, Nov. 2022. 5

[171] W. Zhang, K. Bai, S. Zeadally, H. Zhang, H. Shao, H. Ma, and V. Leung, "DeepMA: End-to-end deep multiple access for wireless image transmission in semantic communication," 2023. [Online]. Available: https://arxiv.org/abs/2303.11543 5

[172] S. Ma, W. Qiao, Y. Wu, H. Li, G. Shi, D. Gao, Y. Shi, S. Li, and N. Al-Dhahir, "Features disentangled semantic broadcast communication networks," 2023. [Online]. Available: https://arxiv.org/abs/2303.01892 5

[173] S. Yang, H. Pan, T.-T. Chan, and Z. Wang, "Semantic communication-empowered physical-layer network coding," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2023, pp. 1–6. 5

[174] Q. Fu, H. Xie, Z. Qin, G. Slabaugh, and X. Tao, "Vector quantized semantic communication system," *IEEE Wireless Communications Letters*, vol. 12, no. 6, pp. 982–986, Mar. 2023. 5

[175] Z. Weng, Z. Qin, and G. Y. Li, "Semantic communications for speech signals," in *ICC 2021 - IEEE International Conference on Communications*, Jun. 2021, pp. 1–6. 5

[176] Y. Tang, N. Zhou, Q. Yu, D. Wu, C. Hou, G. Tao, and M. Chen, "Intelligent fabric enabled 6G semantic communication system for in-cabin scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 1153–1162, Jun. 2022. 5

[177] 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 Transactions on Wireless Communications*, pp. 1–1, Feb. 2023. 5

[178] N. Chinchor, "MUC-4 evaluation metrics," in *Proceedings of the 4th Conference on Message Understanding*, Jun. 1992, pp. 22–29. V-A1

[179] Y. Qiu, S. Wu, Y. Wang, J. Jiao, N. Zhang, and Q. Zhang, "On scheduling policy for multiprocess cyber–physical system with edge computing," *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 18 559–18 572, Mar. 2022. V-A2

[180] K. Huang, W. Liu, Y. Li, B. Vucetic, and A. Savkin, "Optimal downlink–uplink scheduling of wireless networked control for industrial IoT," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1756–1772, Oct. 2019. V-A2

[181] Y. E. Sagduyu, S. Ulukus, and A. Yener, "Age of information in deep learning-driven task-oriented communications," in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2023, pp. 1–6. V-B1, V

[182] S. Xie, S. Ma, M. Ding, Y. Shi, M. Tang, and Y. Wu, "Robust information bottleneck for task-oriented communication with digital modulation," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, Jun. 2023. V-B1, V

[183] C. Liu, C. Guo, Y. Yang, and N. Jiang, "Adaptable semantic compression and resource allocation for task-oriented communications," 2022. [Online]. Available: http://arxiv.org/abs/2204.08910 V-B1, V

[184] M. Wang, J. Li, M. Ma, and X. Fan, "SNN-SC: A spiking semantic communication framework for classification," 2023. [Online]. Available: http://arxiv.org/abs/2210.06836 V-B1, V-B1, V

[185] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications against semantic noise," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, Sep. 2022, pp. 1–6. V-B1, V

[186] ——, "Robust semantic communications with masked VQ-VAE enabled codebook," *IEEE Transactions on Wireless Communications*, pp. 1–1, Apr. 2023. V-B1, V

[187] J. Shao, X. Zhang, and J. Zhang, "Task-oriented communication for edge video analytics," 2022. [Online]. Available: http://arxiv.org/abs/2211.14049 V-B1, V

[188] Q. Pan, H. Tong, J. Lv, T. Luo, Z. Zhang, C. Yin, and J. Li, "Image segmentation semantic communication over Internet of vehicles," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2023, pp. 1–6. V-B1, V

[189] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "Wireless image retrieval at the edge," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 89–100, Nov. 2020. V-B1, V

[190] W. F. Lo, N. Mital, H. Wu, and D. Gündüz, "Collaborative semantic communication for edge inference," *IEEE Wireless Communications Letters*, pp. 1–1, Mar. 2023. V-B1, V

[191] S. Wan, Q. Yang, Z. Shi, Z. Yang, and Z. Zhang, "Cooperative task-oriented communication for multi-modal data with transmission control," 2023. [Online]. Available: http://arxiv.org/abs/2302.02608 V-B2, V

[192] A. Mostaani, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Task-oriented data compression for multi-agent communications over bit-budgeted channels," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1867–1886, Oct. 2022. V-B2, V

[193] R. Li, C. Huang, X. Qin, S. Jiang, N. Ma, and S. Cui, "Coexistence between task-and data-oriented communications: A whittle's index guided multi-agent reinforcement learning approach," *IEEE Internet of Things Journal*, pp. 1–1, Jul. 2023. V, V-B2

[194] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, and D. I. Kim, "AI-generated incentive mechanism and full-duplex semantic communications for information sharing," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, Jun. 2023. V, V-B3

[195] J. Kang, H. Du, Z. Li, Z. Xiong, S. Ma, D. Niyato, and Y. Li, "Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 186–201, Nov. 2022. V, V-B3

[196] W. Xu, Y. Zhang, F. Wang, Z. Qin, C. Liu, and P. Zhang, "Semantic communication for Internet of vehicles: A multi-user cooperative approach," 2022. [Online]. Available: http://arxiv.org/abs/2212.03037 V, V-B3

[197] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented multi-user semantic communications for VQA," *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 553–557, Dec. 2021. V, V-B3, VI-B

[198] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2584–2597, Jul. 2022. V, V-B3

[199] H. Xie, Z. Qin, and G. Y. Li, "Semantic communication with memory," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2658–2669, Jun. 2023. V, V-B3

[200] C. Liu, C. Guo, S. Wang, Y. Li, and D. Hu, "Task-oriented semantic communication based on semantic triplets," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2023, pp. 1–6. V, V-B3

[201] W. Maass, "Network of spiking neurons: the third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, Dec. 1997. V-B1

[202] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, X. Tao, and G. Y. Li, "A unified multi-task semantic communication system for multimodal data," 2022. [Online]. Available: http://arxiv.org/abs/2209.07689 V-B1

[203] G. Zhang, Q. Hu, Z. Qin, Y. Cai, and G. Yu, "A unified multi-task semantic communication system with domain adaptation," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Dec. 2022, pp. 3971–3976. V-B1

[204] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *2011 IEEE Network Science Workshop*, Jun. 2011, pp. 110–117. VI-A

[205] Y. Bar-Hillel and R. Carnap, "Semantic information," *The British Journal for the Philosophy of Science*, vol. 4, no. 14, pp. 147–157, Aug. 1953. VI-A

[206] A. Li, S. Wu, S. Meng, and Q. Zhang, "Towards goal-oriented semantic communications: New metrics, open challenges, and future research directions," 2023. [Online]. Available: http://arxiv.org/abs/2304.00848 VI-A

[207] D. Wen, P. Liu, G. Zhu, Y. Shi, J. Xu, Y. C. Eldar, and S. Cui, "Task-oriented sensing, computation, and communication integration for multi-device edge AI," *IEEE Transactions on Wireless Communications*, pp. 1–1, Aug. 2023. VI-C

[208] P. Liu, G. Zhu, S. Wang, W. Jiang, W. Luo, H. V. Poor, and S. Cui, "Toward ambient intelligence: Federated edge learning with task-oriented sensing, computation, and communication integration," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 158–172, Dec. 2022. VI-C

[209] X. Mu and Y. Liu, "Exploiting semantic communication for non-orthogonal multiple access," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2563–2576, Jun. 2023. VI-C