Large AI Model-Based Semantic Communications

Feibo Jiang, *Member, IEEE*, Yubo Peng, Li Dong, Kezhi Wang, *Senior Member, IEEE*, Kun Yang, *Fellow, IEEE*, Cunhua Pan, *Senior Member, IEEE*, Xiaohu You, *Fellow, IEEE*

***Abstract*—Semantic communication (SC) is an emerging intelligent paradigm, offering solutions for various future applications like metaverse, mixed-reality, and the Internet of everything. However, in current SC systems, the construction of the knowledge base (KB) faces several issues, including limited knowledge representation, frequent knowledge updates, and insecure knowledge sharing. Fortunately, the development of the large AI model provides new solutions to overcome above issues. Here, we propose a large AI model-based SC framework (LAM-SC) specifically designed for image data, where we first design the segment anything model (SAM)-based KB (SKB) that can split the original image into different semantic segments by universal semantic knowledge. Then, we present an attention-based semantic integration (ASI) to weigh the semantic segments generated by SKB without human participation and integrate them as the semantic-aware image. Additionally, we propose an adaptive semantic compression (ASC) encoding to remove redundant information in semantic features, thereby reducing communication overhead. Finally, through simulations, we demonstrate the effectiveness of the LAM-SC framework and the significance of the large AI model-based KB development in future SC paradigms.**

***Index Terms*—Semantic communication; large AI models; knowledge base.**

I. Introduction

Semantic communication (SC), as a new intelligent paradigm, has recently received much attention. It is expected to contribute to various applications such as metaverse, mixed reality (MR), and the Internet of Everything (IoE) [1]. Unlike traditional communication methods, which focused on ensuring the accuracy of transmitted bits or symbols, SC prioritizes delivering the intended meaning with minimal data. Typically, the SC system comprises the following components:

- *Semantic encoder*: The semantic encoder extracts semantic information from the original data and encodes the these features into semantic features, thus understanding the meanings of data and reducing the scale of the transmitted information from the semantic level.
- *Channel encoder*: To ensure data transmitted on the physical channel, the semantic features should be encoded and modulated by the channel encoder to combat channel impairments and improve the robustness.
- *Channel decoder*: The channel decoder is used to demodulate and decode the received signal and obtain the transmitted semantic features before the original data are recovered.
- *Semantic decoder*: The semantic decoder aims to understand the received semantic features, and infer the semantic information and recover the original data from the semantic level.
- *Knowledge base*: SC is a knowledge based system and the knowledge base (KB) is a universal knowledge model which can help the semantic encoder and decoder to understand and infer the semantic information more effectively.

The above components can be implemented by applying deep neural networks (DNNs) which have superior self-learning and feature extraction capabilities. These DNNs can be trained jointly in tandem to maximize expected faithfulness in semantic representation and minimize communications overhead during transmission, and the whole SC system can achieve the global optimality.

Recently, most AI-powered SC system models, including TOSCN [2], DeepSC-ST [3], and DeepJSCC-V [4], centered around designing an efficient communication model. These models heavily rely on the encoder and decoder of SC to extract and interpret semantics. The primary model architectures that facilitate this process includes encoder-decoder (ED) [5], information bottleneck (IB) [6], knowledge graph (KG) [7], and so on. Although these methods are capable of extracting semantic information from unstructured data sources, they may not fully exploit the potential benefits of utilizing KB in their approach.

A. Composition of an Universal KB in SC Systems

In fact, KB is essential for SC to distinguish itself from conventional communication systems by its capacity to understand and infer semantic information. We can build a universal KB by learning a large amount of world knowledge, which forms the core of the SC system. The universal KB consists of prior and background knowledge that can be understood and recognized by users.

1) *Prior knowledge*: SC defines the structure of semantic representation and the relationships between entities

Feibo Jiang (jiangfb@hunnu.edu.cn) is with Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, China.

Yubo Peng (pengyubo@hunnu.edu.cn) is with School of Information Science and Engineering, Hunan Normal University, Changsha, China.

Li Dong (Dlj2017@hunnu.edu.cn) is with Changsha Social Laboratory of Artificial Intelligence, Hunan University of Technology and Business, Changsha, China.

Kezhi Wang (Kezhi.Wang@brunel.ac.uk) is with the Department of Computer Science, Brunel University London, UK.

Kun Yang (kunyang@essex.ac.uk) is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, U.K., also with Changchun Institute of Technology.

Cunhua Pan (cpan@seu.edu.cn) is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China.

Xiaohu You (xhyu@seu.edu.cn) is with the Frontiers Science Center for Mobile Information Communication and Security, National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, and also with the Purple Mountain Laboratories, Nanjing, China.

through prior knowledge. For instance, semantic information can be represented in triplet form for the image understanding. This means that it is made up of three parts, namely, the objective, attribute, and relationship. The entity typically refers to the nouns in the figure, such as housecat and mouse in Fig. 2. The attribute, on the other hand, is based on the adjectives that describe the entities, for example, "domestic housecat" or "clever mouse". Lastly, the relationship refers to the connections between the entities, like "a housecat catching a mouse" instead of the reverse. In essence, through prior knowledge, machines can effectively communicate with humans through SC based on the same or similar ontology, epistemology and logic. This ensures that the semantic information extracted by the system is fully understood by humans.

2) *Background knowledge*: Semantic information is not just about explicit information, but also involve contexts, implicit meaning and common facts. For example, in Fig. 2, the explicit information is a housecat and a mouse, and the background knowledge is "Tom and Jerry". Similarly, SC involves the exchange of background knowledge between the sender and receiver such as user identity, interest preferences, and user environments. This facilitates the semantic encoder in extracting the most relevant information that is of interest to both parties and allows the semantic decoder to accurately recover the intended meaning. Essentially, background knowledge acts as a key enabler for the semantic communication model, facilitating accurate semantic extraction, eliminating redundancy, and ensuring a successful semantic alignment between the sender and receiver.

B. Issues about Current KB Schemes in SC Systems

The current KB schemes in SC are based on the mature deep learning technology, which is a date-driven learning process. However, the complicated and time-consuming learning process will result in various issues.

1) *Limited knowledge representation*: Traditional SC systems, normally using DNNs or KGs as the KB, should learn from the environment by supervised learning. However, the layers and parameters of the KB are limited and the labeled data collected from the environment has high cost. These KBs with restricted parameters and data prevent them from learning abundant semantic knowledge in large data sets and impair their knowledge representation, as well as hinder their ability to comprehensively capture the underlying meaning of human knowledge. For instance, the word "apple" in "Apple Inc" and "apple soda" will be represented as the same features in traditional word embedding model.
2) *Frequent knowledge updates*: Current KB schemes should continuously update their knowledge through training and sharing when the knowledge domain is changed in the environment. In real-world scenarios where there is a massive circulation of data, hence frequent updates are required to maintain the performance of the SC system and these updates normally incur huge energy and resource costs, further reducing the efficiency of the KB.
3) *Insecure knowledge sharing*: In SC systems, current KBs at the source and destination are different because the environments they perceived are different, which should cause semantic errors. Hence, it is essential to share the KBs between users, and ensure that the sender and receiver are semantically aligned, which in turn necessitates the frequent transmission of knowledge models between different users. These knowledge models may include some highly sensitive human-related information, which introduces potential privacy and security risks.

C. Our Contributions

Recently, there has been significant progress in large AI models, which refer to a type of advanced transformer model with billions of parameters. With the continuous improvement of computing power and the increase in data volume, large AI models have made significant progress recently in the fields of natural language processing, image recognition, and speech recognition, etc. It has many advantages, including *accurate knowledge representation*, *rich prior/background knowledge*, and *low-cost knowledge update*, and thus presents a new opportunity to address the aforementioned issues and enhance the SC system. In this paper, we present a large AI model based-SC (LAM-SC) framework specifically designed for image data. Our contributions can be summarized as follows:

1) We apply a large semantic segmentation model-based KB (SKB), focusing on the SC for image data as an example. SKB leverages the accurate knowledge representation to split a raw or unstructured image into different semantic segments or objectives, each of which can be individually selected and encoded by the sender. This allows the sender to focus on specific semantic objectives that are relevant to their communication requirements.
2) We develop an attention-based semantic integration (ASI) mechanism in the SC encoder, which can accurately weight the semantic importance of the segments generated by the SKB. Then, we integrate the most important segments as a new semantic-aware source data. Therefore, the ASI can realize more precise semantic awareness and thus preserve the most critical semantic segments without human intervention.
3) We propose a novel adaptive semantic compression (ASC) encoding in semantic encoder. The ASC can mask a part of transmitted semantic features, and the mask ratio can adjust adaptively according to the content of the transmitted features. Thus, we can ensure that redundant semantic features are further eliminated, leading to a significant reduction in communication overhead.

The remainder of this paper is organized as follows: We begin by presenting various approaches for implementing large AI model-based KBs in SC systems. Following that, we introduce the proposed LAM-SC framework, detailing its key components, such as the SKB, ASI, and ASC methods. We then conduct simulations to demonstrate the advantages of the LAM-SC framework. Finally, we discuss open issues and conclude this paper.

II. Large AI Models-Based KBs in SC systems

A. Advantages of Large AI Models-Based KBs

The large AI model refers to the transformer model that has complicated structure with multi-head attention, enabling it to handle complex AI tasks and generate high-quality outputs. The large AI model can be pretrained on extensive datasets by self-supervised learning with unlabeled date and then the pretrained model can be applied to various tasks by prompt learning or fine tuning. To unlock the potential of large AI models in constructing a more universal KB and thus promoting the development of SC, we summarize the advantages of introducing large AI models in KBs as follows:

- *Accurate knowledge representation*: Current large AI models, like GPT-4.0, CLIP, and T5 [8], have billions of parameters, allowing them to learn complex knowledge representations from the transformer model with multi-head attention mechanism. The multi-head attention mechanism develops a strong understanding of semantics and knowledge structures, hence large AI models can give high-quality semantic representation of input data. For instance, the word "apple" in "Apple Inc" and "apple soda" will be represented as different features in large AI models.
- *Rich prior/background knowledge*: Large AI models are pre-trained on extensive datasets such as ImageNet, UCF101, Audioset, and Wikipedia [9], enabling them to learn from vast amount of information across various domains, and they store rich prior/background knowledge and show remarkable generalization abilities. They can achieve high performance on various tasks, even beyond their pre-trained knowledge domains, eliminating the need for frequent updates of KBs.
- *Low-cost knowledge update*: Large AI models typically come with pre-trained weights and can be prompted using just a few examples or fine-tuned with a small amount of labeled data. Techniques like P-Tuning, LoRA, and prompt-tuning enable low-cost updates [10], mitigating concerns of frequent knowledge updates and insecure knowledge sharing.

B. Design Suggestions of Large AI Models in SC Systems

In this subsection, we suggest several design schemes catering for different types of SC systems (i.e. text, image, audio, etc.), allowing for streamlined integration of large AI models into KB creation, as shown in Fig. 1,

1) GPT-Based KBs: Towards the text-based SC system, the KB should be capable of comprehending the text's content and identifying various subjects, their attributes, and relationships. Recently, large language models have emerged, such as ChatGPT [11], which can serve as a semantic knowledge base for text data. ChatGPT is an AI assistant developed by OpenAI based on the GPT-3.5 model, which can accurately understand the content of the text and provide correct responses to a wide range of questions. By using ChatGPT as the KB for text data, it can extract the key content from the input text based on the user's requirements. In the receiver, the received text data recovered by the SC decoder can be fed to ChatGPT to eliminate semantic noise. Additionally, the received text can be reorganized according to the receiving user's preference, such as applying different languages.

2) SAM-Based KBs: For the image-based SC system, the KB should be capable of segmenting various objectives in an image and recognizing their respective categories and interrelationships. One promising AI model that can be applied here is the Segment Anything Model (SAM) which is introduced by Meta AI [12]. SAM is a groundbreaking segmentation system that can generalize zero-shot to unfamiliar images and objectives without any additional training. Therefore, SAM can be considered as the perfect KB for images. For real systems, the sender can use SAM to segment the input image and select the most important and meaningful segments to the SC encoder. On the receiver side, the SC decoder outputs the recovered image data, which then removes any semantic noise or interference by SAM. The segments of interest can then be identified and extracted effectively.

3) WavLM-Based KBs: To enable the SC system for audio, the KB should be capable of performing a variety of audio tasks, including automatic speech recognition, speaker identification, and speech separation. This ensures that the raw audio data can be analyzed and semantic information can be extracted effectively. WavLM [13], as a large-scale audio model proposed by MSRA, can be one potential solution to this application. Trained on 94,000 hours of unsupervised English data, WavLM is highly effective across a range of speech recognition tasks and non-content recognition speech tasks. By using WavLM as the KB, the sender can first separate and recognize the audio data from different speakers, discarding unimportant information such as background noise. The remaining audio data is then integrated and encoded by the SC encoder. In the receiver side, the SC decoder can be used to recover the audio, followed by speech denoising and recognition by WavLM according to the user's requirements.

Among the three major SC systems based on large models, the WavLM-based SC system is well-suited for real-time interactions and instant communication, enabling quick and efficient information exchange. In contrast, the GPT-based SC system excels at conveying thoughts and ideas clearly through textual summaries, making it easy to store, retrieve, and analyze text information. The SAM-based SC system focuses on transmitting visual information via images, capturing intricate details, spatial organization, and colour, as well as accurately representing expressions, emotions, and non-verbal cues for a more intuitive communication experience. At present, there is relatively little research on image-based SC systems, hence we conduct further research on SAM-based SC systems.

III. Architecture of LAM-SC Framework

Introducing the large AI models into SC systems is a promising solution to realize more precise semantic awareness

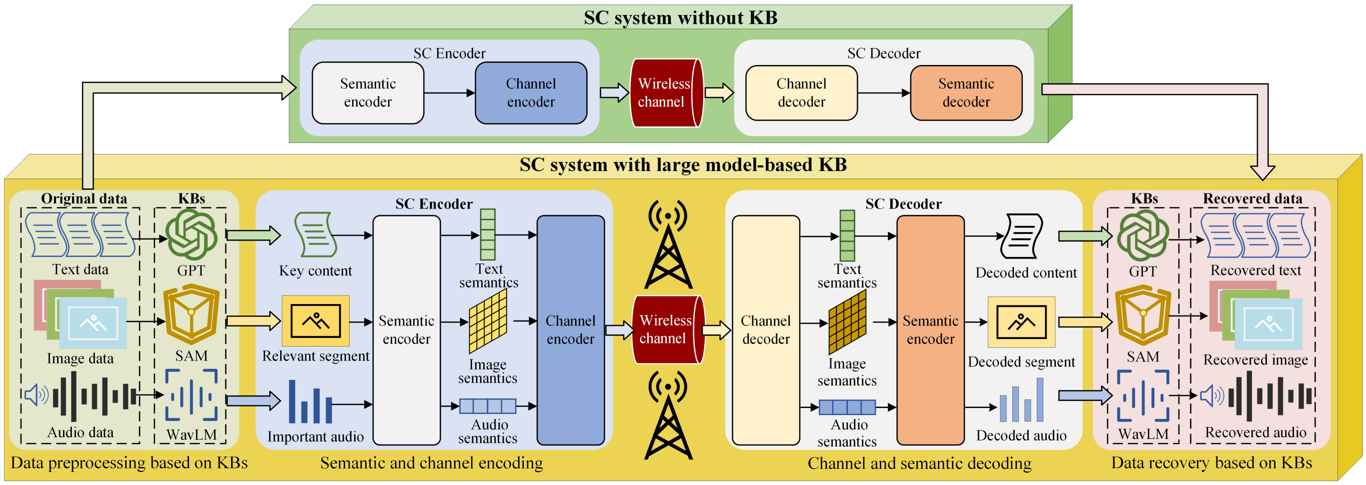

Fig. 1: Implementation of large AI models-based KBs in different SC models.

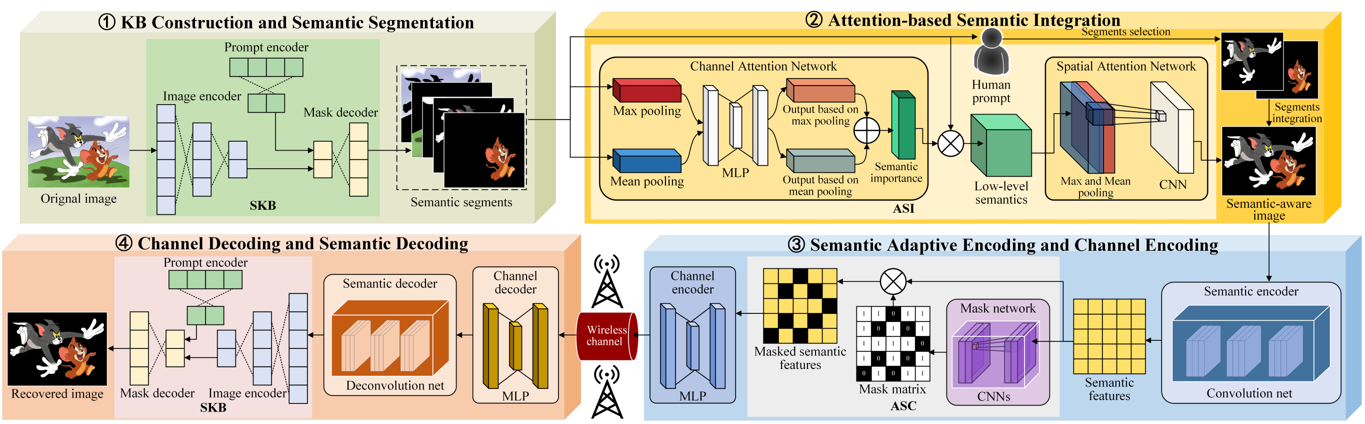

Fig. 2: The illustration of the proposed LAM-SC framework.

and universal KB in the image-based SC. In this section, we present the LAM-SC framework based on image data that incorporates the SAM model in SC systems, the workflow of the LAM-SC framework is shown in Fig. 2.

A. Introduction to LAM-SC Framework

1) KB Construction and Semantic Segmentation: To achieve semantic segmentation for any original image without KB training, the SKB can be designed to facilitate the recognition and segmentation of every semantic objective in an input image. This process involves analyzing the visual information conveyed by the image to identify each individual objective. As a result, multiple segments are generated, each containing only one semantic objective.

2) Attention-Based Semantic Integration: The ASI can be used to simulate the human perception to select the semantic segments that are most worthy of concern by channel attention and spatial attention. Additionally, we also give a human prompt way to select the interested semantic segments directly. As a result, the selected segments can then be merged as a new semantic-aware image.

3) Semantic Adaptive Encoding and Channel Encoding: The semantic-aware image is encoded into semantic features by the semantic encoder. Here, the semantic encoder is built based on convolutional neural networks (CNNs) that have excellent extraction capabilities of image features. Moreover, the ASC can be applied to adaptively mask the unimportant features of the semantic information according to its content. Then, the channel encoder that builds based on the multilayer perceptron (MLP) can be used to perform signal encoding and modulation for the physical channel.

4) Channel Decoding and Semantic Decoding: In these modules, when the transmitted signals reach the receiver through the physical channel, the channel decoder performs signal demodulation and decoding, then the semantic features can be obtained. Here, the channel decoder adopts the MLP architecture. Next, the semantic decoder that consists of deconvolution layers decodes the semantic features and thus recovers the image data. Then, the SKB can be employed again on the recovered source image to identify and segment objectives accurately, aiming to evaluate and confirm the integrity of the interested semantics.

B. SKB

To achieve precise image semantic segmentation for any input images without specific training, we employ SAM as the KB in our proposed LAM-SC framework, namely SKB, as depicted in Fig. 2, which handles the image semantic segmentation process. SAM is a revolutionary segmentation system that is trained on the largest and most comprehensive dataset, Segment Anything 1-Billion (SA-1B), which contains over 1 billion masks across 11 million licensed and privacy-conscious images [12]. This breakthrough system can successfully generalize zero-shot segmentation for previously unseen images or objectives without requiring additional knowledge and training.

SKB utilizes an efficient transformer-based architecture, designed for both natural language processing and image recognition tasks [12]. The system comprises a visual transformer-based image encoder for feature extraction, prompt encoder for user engagement, and a mask decoder for segmentation and confidence score generation. In this research, we leverage SKB to automatically achieve objective separation, producing multiple semantic segments for further analysis and processing. In summary, the SKB possesses sufficient prior/background knowledge and powerful semantic representation to accurately perform semantic segmentation on original image data, which ensures the universality of the KB in SC systems.

C. ASI

The attention mechanism mimics human vision, focusing on crucial details while ignoring irrelevant content. The ASI introduces attention mechanism to identify and weight significant objectives in images, which consists of two parts:

1) Channel Attention Network: Using the channel attention network, we can extract low-level semantics from semantic segments. Each segment is treated as a channel, and global and mean pooling operations are performed. The results are then input into an MLP network for assessing the channel significance. The MLP outputs are combined to determine semantic importance, which is then multiplied by the semantic segments to obtain low-level semantics.

2) Spatial Attention Network: Each low-level semantic represents a single segment, inadequately capturing the whole image semantics. To address this, we use the spatial attention network to merge low-level semantics for a high-level semantic representation. Specifically, we separately perform global and mean pooling on low-level semantics and concatenate results along the channel dimension, then we apply a CNN to integrate all low-level semantics into a high-level semantic-aware image.

In summary, the ASI is capable of intuitively recognizing and retaining essential objectives in original image that are typically of greater interest to humans, even without any human involvement.

D. ASC

We propose the ASC method, which adaptively masks the transmitted semantic features from semantic level, effectively reducing redundant data and significantly decreasing communication overhead. As illustrated in Fig. 2, we utilize a learnable mask network to generate the mask matrix, thereby eliminating unimportant data from the encoded semantic features. During transmission, the encoded semantic features are fed into the mask network, which outputs a corresponding mask matrix with values of either 0 or 1. The semantic features are then multiplied by the mask matrix, causing a portion of the unimportant features to be set to 0 and then obtaining the masked semantic features.

To sum up, by applying the ASC to the semantic transmission process, essential semantic features can be maintained while superfluous semantic features are excluded, leading to a substantial reduction in communication overhead.

IV. TRAINING OF LAM-SC FRAMEWORK

In this section, we show how to train the proposed LAM-SC framework, as illustrated in Fig. 3. Remarkably, SAM in SKB is a pretrained large AI model, and it does not require training.

1) ASI Training Based on Human Experience: As mentioned before, the aim of the proposed ASI is to mimic human perception in identifying interested objectives in original images, and then producing semantic-aware images that correspond with human preferences. To accomplish this, we record human interested semantics as experiences, which forms the foundational training for the attention networks, encompassing both channel and spatial attention networks. In this experience base, semantic segments can be served as input samples for attention networks, while the semantic-aware images created through human prompt can be seen as associated labels. By supervised learning on the experience database, attention networks can effectively adapt to human behavior and make decisions that closely resemble human perception.

2) Crossover-Based SC Encoder and Decoder Training: The SC encoder consists of semantic and channel encoders, while the SC decoder comprises channel and semantic decoders. Firstly, to jointly train the channel encoder and decoder, we can use the mutual information as the objective function, which eliminates noise or fading effects during transmission and prevents signal distortion [14]. Then, for training semantic encoder and decoder, we can apply semantic alignment, which introduces the difference between the original and recovered images as the optimization function to guide the learning. We then implement a crossed training strategy involving both the channel encoder/decoder and semantic encoder/decoder models. To be more specific, we first train the channel model, then freeze its parameters, and next train the semantic model. Next, we freeze the semantic model parameters and train the channel model again. This process can be repeated until the entire SC model achieves convergence [14].

3) ASC Training: For generating a mask array that accurately reflects the importance of semantic features, we propose a joint training approach for the mask network and the SC model (i.e., channel/semantic encoder/decoder), where the parameters of the SC model and the attention networks are frozen. The training process includes the following steps: initially, both raw and masked semantic features are transmitted. Then, the two sets of semantics are decoded independently.

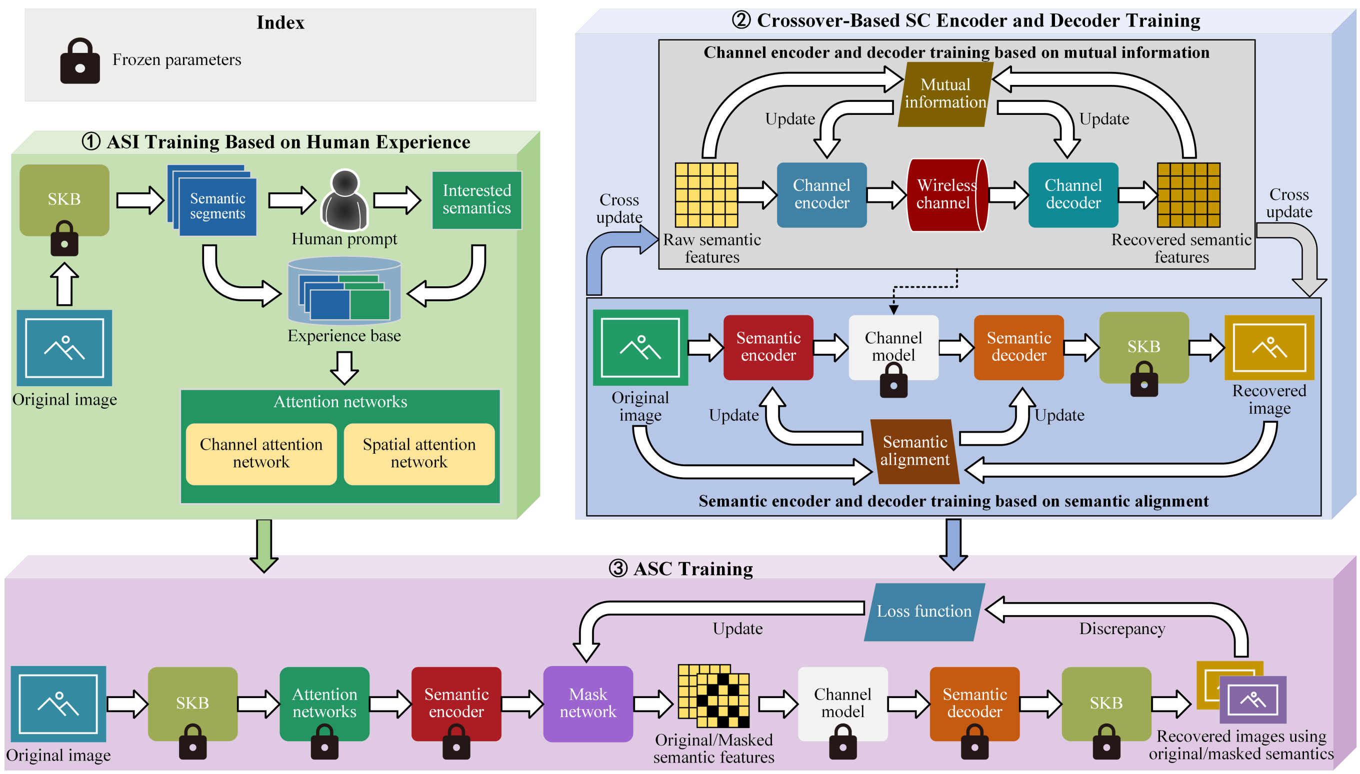

Fig. 3: The training process of the proposed LAM-SC framework.

Next, the difference between the recovered images using raw and masked semantics is employed as the loss function for the mask network, enabling it to learn how to produce an optimal mask array that can minimize this difference.

V. Simulation Results

To showcase the effectiveness of the proposed framework, we conduct the case study comparing the LAM-SC and traditional SC approaches, using the VOC2012 dataset that consists of 17,125 RGB images [15]. It is essential to note that the SKB is not trained on the VOC2012 dataset during our simulations. Regarding the LAM-SC model architecture, the following components are included:

1) *Attention network*: The channel network components comprise a max pooling layer, a mean pooling layer, and an MLP layer. The spatial network components include a mean pooling layer, a max pooling layer, and a convolutional layer.
2) *Semantic encoder*: It comprises two blocks, each with a convolutional layer and a pooling layer.
3) *Mask network*: It consists of two convolutional layers, with each layer being followed by a Rectified Linear Unit (ReLU) activation function.
4) *Channel model*: The design of the channel model, encompassing channel encoding and decoding as well as wireless channel configuration, adopts similar settings to those presented in [14].
5) *Semantic decoder*: It is made up of two blocks, each with a deconvolution layer and an upsampling layer.

The conventional SC model is used as the benchmark, which only comprises the semantic encoder and decoder, and the channel encoder and decoder. Furthermore, the traditional SC model does not utilize the SKB for segmenting raw images prior to transmission. We evaluate performance using three key metrics: loss value, peak signal-to-noise ratio (PSNR), and structural similarity (SSIM).

The simulation outcomes are presented in Fig. 4, where Fig. 4(a) illustrates the decline in the loss value as the number of epochs increases, indicating superior convergence results for the LAM-SC compared to the traditional SC scheme under the same SNR conditions. Fig. 4(b) reveals that images transmitted using LAM-SC attain higher PSNR values, implying that LAM-SC effectively minimizes image distortion during transmission. Likewise, Fig. 4(c) demonstrates that LAM-SC maintains the structural consistency of transmitted images, thus achieving higher SSIM values.

In short, compared with traditional communication methods that require transmitting precise original images, SC focuses on transmitting only the extracted semantic features from the images, resulting in a notable reduction of data size. Building upon classical SC, LAM-SC leverages the SKB and ASI to select key semantic objectives of the original image to encapsulate the image's semantics, which can further decrease the communication overhead without compromising the accuracy of the semantic representation. In our simulations, the original image in Fig. 2 needs 49,152 bits for transmitting, the semantic features transmitted in the traditional SC are 21,632 bits, and the semantic features transmitted in LAM-SC only require 8,960 bits.

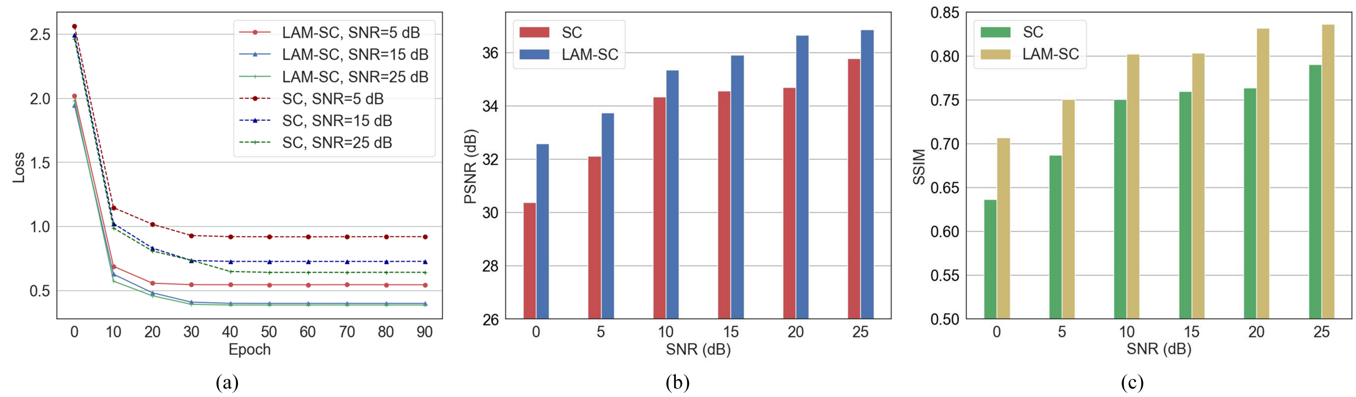

Fig. 4: The simulation results of the proposed LAM-SC framework. (a) Loss versus epoch. (b) PSNR versus SNR. (c) SSIM versus SNR.

VI. Open Issues

Currently, there is little research on the large AI model-empowered SC, and there are several open issues and challenges to be dealt with in the near future, which are summarized as follows.

A. High Latency in Real-Time Applications

Large AI models with millions or billions of parameters require substantial runtime, resulting in significant latency during training, updating, and decision-making processes. Furthermore, bandwidth limitations in communication systems can lead to bottlenecks when transferring considerable amounts of data from large AI models in SC systems. Reducing latency is crucial for real-time applications such as metaverse and XR, where immediate responses are essential. Prospective solutions may include the collaborative design of SC encoders/decoders and the knowledge base from an efficient perspective.

B. High Energy Consumption for Edge Devices

The implementation of large AI models in SC systems requires a significantly higher level of energy compared to traditional methods. This increased energy consumption raises environmental concerns and presents accessibility challenges for mobile and IoT devices. Consequently, striking a balance between computational demands and energy constraints is of critical importance. To tackle these issues, research and development initiatives should prioritize model optimization, resource scheduling, data compression, and lightweight learning methods.

C. Explainability and Transparency

Interpreting the decisions of semantic encoder made by large AI models during SC can be difficult. The large AI models often lack interpretability, making it challenging to understand the semantic analysis process. This can pose difficulties in identifying potential biases or errors in SC systems. It is essential to develop methods to provide explanations and increase transparency so that users can understand the reasoning behind semantic model's responses.

D. Privacy and Security

Large AI models can capture sensitive information during training or infer sensitive details from the data they process. Integrating these models into communication systems raises concerns about privacy and security. Ethical considerations regarding issues like consent and responsible use become even more critical. Proper safeguards and mechanisms need to be implemented to protect user data and prevent unauthorized access to the models or the information they handle.

VII. Conclusion

In this paper, we introduce the importance and composition of KBs, and then we discuss the issues about current KB schemes in SC systems. To address these issues, we recommend introducing large AI models building KBs, and we explore several large AI model-based schemes to realize KBs in different SC systems. Then, we propose a LAM-SC framework focused on image data, in which the large AI model SAM is applied to the KB for high-quality semantic segmentation, and ASI is presented to integrate segment semantics as a new semantic-aware source. Additionally, ASC is proposed to reduce communication overhead in semantics transmission. Finally, we conduct simulations to demonstrate the effectiveness of the proposed LAM-SC framework.

References

[1] 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, 2022.

[2] H. Zhang, H. Wang, Y. Li, K. Long, and A. Nallanathan, "Drl-driven dynamic resource allocation for task-oriented semantic communication," *IEEE Transactions on Communications*, pp. 1–1, 2023.

[3] 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, 2023.

[4] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, and V. C. M. Leung, "Predictive and adaptive deep coding for wireless image transmission in semantic communication," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[5] 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, 2022.

[6] 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*, 2023.

[7] A. Li, X. Wei, D. Wu, and L. Zhou, "Cross-modal semantic communications," *IEEE Wireless Communications*, vol. 29, no. 6, pp. 144–151, 2022.

[8] Y.-L. Sung, J. Cho, and M. Bansal, "Vl-adapter: Parameter-efficient transfer learning for vision-and-language tasks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5227–5237.

[9] L. Ericsson, H. Gouk, C. C. Loy, and T. M. Hospedales, "Self-supervised representation learning: Introduction, advances, and challenges," *IEEE Signal Processing Magazine*, vol. 39, no. 3, pp. 42–62, 2022.

[10] J. R. Chowdhury, Y. Zhuang, and S. Wang, "Novelty controlled paraphrase generation with retrieval augmented conditional prompt tuning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 10, 2022, pp. 10 535–10 544.

[11] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, "A brief overview of chatgpt: The history, status quo and potential future development," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1122–1136, 2023.

[12] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[13] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao *et al.*, "Wavlm: Large-scale self-supervised pre-training for full stack speech processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1505–1518, 2022.

[14] 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, 2021.

[15] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, pp. 98–136, 2015.

BIOGRAPHIES

Feibo Jiang received his B.S. and M.S. degrees in School of Physics and Electronics from Hunan Normal University, China, in 2004 and 2007, respectively. He received his Ph.D. degree in School of Geosciences and Info-physics from the Central South University, China, in 2014. He is currently an associate professor at the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, China. His research interests include artificial intelligence, fuzzy computation, Internet of Things, and mobile edge computing.

Yubo Peng received the B.S. degree from Hunan Normal University, Changsha, China, in 2019, where he is currently pursuing the master's degree with the College of Information Science and Engineering. His main research interests include federated learning and semantic communication.

Li Dong received the B.S. and M.S. degrees in School of Physics and Electronics from Hunan Normal University, China, in 2004 and 2007, respectively. She received her Ph.D. degree in School of Geosciences and Info-physics from the Central South University, China, in 2018. She is currently an associate professor at Hunan University of Technology and Business, China. Her research interests include machine learning, Internet of Things, and mobile edge computing.

Kezhi Wang received the Ph.D. degree in Engineering from the University of Warwick, U.K. He was with the University of Essex and Northumbria University, U.K. Currently, he is a Senior Lecturer with the Department of Computer Science, Brunel University London, U.K. His research interests include wireless communications, mobile edge computing, and machine learning.

Kun Yang received his PhD from the Department of Electronic Electrical Engineering of University College London (UCL), UK. He is currently a Chair Professor in the School of Computer Science Electronic Engineering, University of Essex, leading the Network Convergence Laboratory (NCL), UK. He is also an affiliated professor at UESTC, China. Before joining in the University of Essex at 2003, he worked at UCL on several European Union (EU) research projects for several years. His main research interests include wireless networks and communications, IoT networking, data and energy integrated networks and mobile computing. He manages research projects funded by various sources such as UK EPSRC, EU FP7/H2020 and industries. He has published 400+ papers and filed 30 patents. He serves on the editorial boards of both IEEE (e.g., IEEE TNSE, IEEE ComMag, IEEE WCL) and non-IEEE journals (e.g., Deputy EiC of IET Smart Cities). He was an IEEE ComSoc Distinguished Lecturer (2020-2021). He is a Member of Academia Europaea (MAE), a Fellow of IEEE, a Fellow of IET and a Distinguished Member of ACM.

Cunhua Pan received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. He was a Research Associate with the University of Kent, Canterbury, U.K., from 2015 to 2016. He held a postdoctoral position with the Queen Mary University of London, London, U.K., from 2016 and 2019, and was a Lecturer from 2019 to 2021. He has been a Full Professor with Southeast University since 2021. He has published over 120 IEEE journal papers. His research interests mainly include reconfigurable intelligent surfaces (RIS), intelligent reflection surface, ultrareliable low-latency communication, machine learning, UAV, Internet of Things, and mobile-edge computing.

Xiaohu You received the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively. Since 1990, he has been with the National Mobile Communications Research Laboratory, Southeast University, where he is currently the Director and a Professor. From 1999 to 2002, he was a Principal Expert of the C3G Project, responsible for organizing China 3G Mobile Communications Research and Development Activities. From 2001 to 2006, he was a Principal Expert of the China National 863 Beyond 3G FuTURE Project. Since 2013, he has been a Principal Investigator of the China National 863 5G Project. He has contributed over 200 IEEE journal articles and two books in the areas of adaptive signal processing and neural networks, and their applications to communication systems. His research interests include mobile communication systems, and signal processing and its applications.

Dr. You was selected as an IEEE Fellow for his contributions to the development of mobile communications in China in 2011. He was a recipient of the National 1st Class Invention Prize in 2011. He has served as the General Chair of the IEEE Wireless Communications and Networking Conference 2013, the IEEE Vehicular Technology Conference 2016, and the IEEE International Conference on Communications 2019. He is the Secretary-General of the FuTURE Forum and the Vice-

Chair of the China IMT-2020 Promotion Group and the China National Mega Project on New Generation Mobile Network.