1

Introduction

Wireless communication systems have undergone vigorous advancements from the first generation (1G) to the fifth generation (5G) over the past few decades by developing numerous coding algorithms and channel models to recover accurate sources at the bit level. However, in recent years, the flourishing of artificial intelligence (AI) has revolutionized various industries and incubated multifarious intelligent tasks, which increases the amount of data transmission to the zetta-byte level and requires massive machine connectivity with low transmission latency and energy consumption. In this context, conventional communication systems face severe challenges imposed by ubiquitous AI tasks. Therefore, it is inevitable that a new communication paradigm needs to be developed. Semantic communications have been proposed to address the challenges by extracting semantic information inherent in source data while omitting irrelative redundant information to reduce the transmission data, thereby lowering communication resources and facilitating high semantic fidelity transmission. Nevertheless, the exploration of semantic communications has gone through decades of stagnation since it was first identified because of the inadequacy of mathematical models for semantic information. Inspired by the thriving of AI, deep learning (DL)-enabled semantic communications have been scrutinized as promising solutions to the bottlenecks in conventional communications by leveraging the learning and fitting capabilities of neural networks (NNs) to bypass mathematical models for semantic extraction and representation.

To this end, this book introduces DL-enabled semantic communications, including concepts, applications, and challenges. In particular, it comprehensively covers semantic communications for source reconstruction, task-oriented semantic communications, semantic impairments, semantic knowledge bases (SKBs), and large model-driven semantic communications. The objective of this book is

Deep Learning Enabled Semantic Communications, First Edition. Zhijin Qin et al.

to help readers gain a fundamental appreciation and understanding of the emerging DL-enabled semantic communications and their implications for future intelligent wireless networks.

1.1 Conventional Communications versus Semantic Communications

In this section, we introduce the basic yet important theory of wireless communications and briefly present the history of semantic communications.

1.1.1 Three-level Communications

In 1949, Shannon and Weaver [1] categorized communications into three levels:

- *Level A*: how accurately can the symbols of communication be transmitted? (The technical problem)
- *Level B*: how precisely do the transmitted symbols convey the desired meaning? (The semantic problem)
- *Level C*: how effectively does the received meaning affect conduct in the desired way? (The effectiveness problem)

The first level of communications is to deliver the symbol transmission accurately, that is, syntactic communications, which improves the data rate by expanding the bandwidth resources, increasing the transmission power, and adding transmission antennas. The 1G to 5G wireless communication networks belong to syntactic communications, and the system capacity has been proven to be approaching the Shannon limit by utilizing efficient source coding and channel coding algorithms [2]. The different coding modules in conventional communication systems are designed and optimized separately, which converts the input message into the bit sequence and focuses on performance improvement at the bit or symbol level by taking bit-error rate (BER) or symbol-error rate (SER) as metrics to measure the information loss. However, the bit-oriented transmission framework requires the precise alignment of input and recovery messages. Still, it neglects the underlying meaning behind bits, which augments the possibility of transmitting unnecessary data beyond user requirements, accelerating the consumption of communication resources and increasing transmission latency. Moreover, due to the wide deployment of Internet-of-things (IoT) applications, conventional communications are no longer ideal as they transmit information that could be irrelevant to the downstream intelligent tasks at the receiver.

To this end, the second level of communications, that is, semantic communications, considers the inherent semantics of input information to tackle the technical problems in the bit-oriented paradigm. Semantics take into account the meaning and veracity of source information because they can be both informative and factual [3]. Besides, semantic data can be compressed to a proper size for transmission by using a lossless method [4], which leverages the semantic relationship between different messages, while the traditional lossless source coding represents a signal with the minimum number of binary bits by exploring the dependencies or statistical properties of the input message. Moreover, the semantic information varies for different transmission purposes, which could be in various formats, for example, age of information [5], or more complicated semantic features. Inspired by this, semantic communications deviate from the bit-oriented paradigm by extracting semantics from the input message with minimal ambiguity to facilitate semantic exchange between the transmitter and the receiver, committing to reducing network traffic by transmitting low-dimensional semantic information and optimizing the system performance by minimizing the semantic error instead of the BER or SER. It is worth mentioning that there is currently no consensus on the definition of semantic error in the field of semantic communications. In some works, semantic errors refer to inaccurate semantics that lead to misunderstanding and ambiguity, such as spelling errors in the text.

Furthermore, the third level of communications involves the effectiveness of semantic transmission, that is, pragmatic communications, which condenses the input message to obtain the semantics only associated with user requests. Inspired by this mechanism, goal/task-oriented semantic communications are boosted to perform specific downstream tasks required by receiver users by delivering task-related semantic features [6]. Conventional communication systems are constrained to achieve source reconstruction in the same modal before performing any modal conversion at the receiver. However, in goal/task-oriented semantic communications, the recovered message is no longer confined to the same modality as the input message, facilitating flexible cross-modal or single-modal to multimodal transmissions, such as speech-to-text and text-to-image/speech, to satisfy different user requests and improve the user quality of experience (QoE).

1.1.2 History of Semantic Communications

Decades ago, the concept of semantic communication was pioneered by engineers and philosophers. As far back as 1925, Dewey [7] asserted that "communication should be regarded as a mechanism for attaining purposes," while Wittgenstein [8] emphasized the centrality of meaning in philosophical discourse. Later, in 1938, Cherry went on to define semantics in terms of signs, which fall into the following categories:

- Syntax: studies the signs and their relations to other signs;
- Semantics: studies the signs and their relation to the world;
- Pragmatics: studies the signs and their relations to users.

Potentially influenced by the aforementioned definition, in 1949, Weaver [1] formulated semantics with the perspective of engineering and outlined three communication levels, that is, *technical-, semantic-, and effectiveness-levels*, as noted earlier. Shannon defined communication as the precise or approximate reconstruction of information from one point to another, falling under the technical level. Consequently, Shannon applied statistical methods to quantify the uncertainty of information while overlooking the semantic and effectiveness levels. Despite Shannon's clear indication, many researchers still tried to adapt the statistical framework to interpret or evaluate semantics. Carnap and Bar-Hillel [3] first attempted to develop a "theory of semantic communication" in 1952 with the concept of logical probability. Since then, explorations into semantic communication have continued steadily.

Up until recently, semantic communication gained more and more attention as a possible solution to enable more efficient ways to exchange information. In 2003, Jeong et al. [9] first employed a knowledge graph in semantic communications. After one year, Floridi et al. [10] outlined the "theory of strongly semantic information." In 2011, Bao et al. [11] proposed a generic model of semantic communications. After a period in the wilderness, the bloom of DL has prepared the way for all progress and development of semantic communications because of the significant progress in semantic extraction and understanding. For example, the remarkable works [12, 13] designed the initial deep joint source-channel coding (JSCC) for text transmission and proposed a single-user joint semantic-channel coding in 2018 and 2021, respectively. After that, the research of semantic communications mainly focuses on the DL-enabled paradigm.

1.2 Introducing Deep Learning to Semantic Communications

We first introduce the concept of DL and present some cutting-edge algorithms relevant to the following chapters. Then, we discuss the motivations and details of applying deep learning in semantic communication systems.

1.2.1 Deep Learning Basics

DL is defined as a subset of machine learning (ML) that falls under the umbrella of AI. DL leverages numerous classifiers working together based on linear regression followed by specific activation functions. The principle behind DL mirrors that of conventional statistical linear regression, albeit with considerable neural nodes in DL instead of just one node in traditional approaches. These neural nodes are connected through complicated topological structures to form a NN. The motivation of the NN derives from the human brain, which allows it to mimic intricate mathematical patterns and learn data representations through hierarchical intelligence.

Inspired by this, DL leads to unprecedented success in various areas, such as computer vision (CV) and natural language processing (NLP). The fundamental DL algorithms are briefly introduced as follows.

1.2.1.1 Fully Connected Neural Network

A fully connected NN is a classic DL model consisting of multiple fully connected layers, also termed dense layers, where each neuron establishes connections with all neurons in the preceding layer. Based on sufficient training data, the fully connected NN adaptively adjusts its weights and/or bias matrices to steer the output toward the predetermined target. An example of the fully connected NN is shown in Figure 1.1. From the figure, the output of each neuron can be denoted as:

$$\text{Output}_j = f_a\left(\sum_{i=1}^{\mathcal{I}} \left(\mathcal{W}_{i,\,j} \times \text{Intput}_i\right) + bias_j\right), \tag{1.1}$$

where $\mathcal{W}_{i,\,j}$ represents the trainable weight corresponding to the i-th neuron in the previous layer to the j-th neuron in the current layer; bia_j and $f_a(\cdot)$ denote the bias of the j-th neuron in the current layer and the adopted activation function, respectively.

The fully connected NN has proved remarkable adaptability and compatibility with multimodal data and intricate cognitive tasks, owing to its robust nonlinear learning and fitting prowess. Nevertheless, it is susceptible to overfitting and demands tremendous computational resources for efficient model training. Consequently, it usually synergizes with other DL algorithms to construct a sophisticated NN architecture.

1.2.1.2 Convolutional Neural Network

The inception of the convolutional neural network (CNN) can be attributed to LeCun et al. [14], who initially proposed it as an end-to-end (E2E) image analysis framework tailored for handwritten digit recognition. Over the intervening

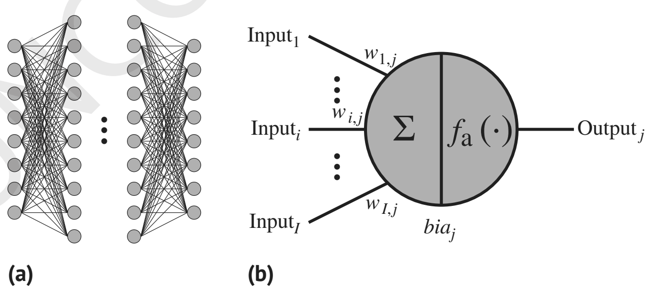


Figure 1.1 (a) The fully connected NN architecture; (b) one neuron processing unit.

decades, CNN and its variants have shown significant superiority in handling visual data, surmounting substantial bottlenecks in CV that exceed human limits. The central concept of CNN is the comprehensive processing and fusion of feature maps, enabling the inference of nonlinear correlations between input data and established ground truths.

The typical CNN architecture for the image classification task is shown in Figure 1.2. From the figure, the CNN architecture consists of the upstream feature extractor and the downstream classifier. The backbone of the feature extractor includes multiple convolutional layers and pooling layers. In particular, the convolutional layer meticulously orchestrates the output of neurons connected to local regions of the input image by utilizing the property of translation invariance and determines the output of each neuron by calculating scalar products between neural weights within the convolutional kernel and the corresponding input region. Subsequently, the pooling layer adeptly executes the spatial downsampling on the output of the convolutional layer to diminish the number of parameters during the activation process. Note that the depth of extracted features depends on the number of kernels, and the trainable weights within these kernels are updated during the training phase. Moreover, the classifier contains tandem fully connected layers, transforming the extracted features into class predictions and returning a probability vector. As a result, the accurate classification label for the input image can be inferred by indexing the position within the predefined label vector that corresponds to the highest probability.

CNNs exhibit lower complexity without sacrificing performance compared to fully connected NNs and dramatically mitigate the risk of overfitting through sharing parameters across the network. In addition, CNN is designed based on parallel computing architecture, providing great superiority for acceleration when implemented on hardware. CNNs are considered a pivotal DL algorithm and have been

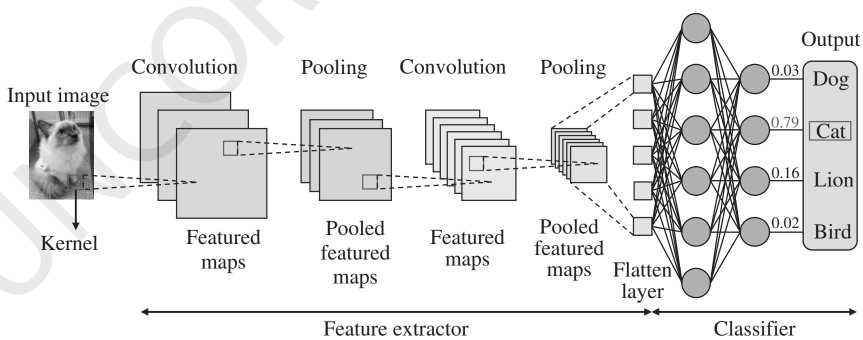


Figure 1.2 The typical CNN architecture for image classification.

extended from image processing to multimodal, for example, text and speech. Furthermore, the compatibility and universality of CNNs have been verified across diverse domains.

1.2.1.3 Recurrent Neural Network

The Hopfield network [15] developed the precursor of the recurrent neural network (RNN), showcasing the feasibility of recurrent structures within NNs. RNNs are bespoke to detect patterns in sequential data by maintaining an internal state, that is, memory, that captures information about previous inputs, allowing it to demonstrate dynamic temporal behavior and perform intelligent tasks, for example, time series prediction, machine translation, and speech recognition. RNNs are structured with recurrent connections, facilitating the effective extraction of features and modeling of dependencies inherent in sequential data. The basic RNN architecture is shown in Figure 1.3. From the figure, the input from the previous layer at the current time step and the hidden state of the current layer at the previous time step are both fed into the hidden state of the current layer at the current time step to produce the output.

However, the phenomenon of memory decay amplifies with the continuous expansion of temporal sequences, constraining the capacity to understand long-term dependencies and precipitating the predicament of gradient vanishing/-exploding. Toward this end, variants of the standard RNN, for example, long short-term memory (LSTM) [16] and gated recurrent unit (GRU) [17], were proposed to contend with the challenge of prolonged temporal interdependencies by incorporating specialized mechanisms to better process and preserve information over time.

The structure diagrams of the LSTM unit and the GRU unit are shown in Figure 1.4. From the figure, the LSTM unit introduces an innovative module, the memory cell, to control the information flow and maintain the ability to retain long-term memory. Specifically, the memory cell consists of the forget gate,

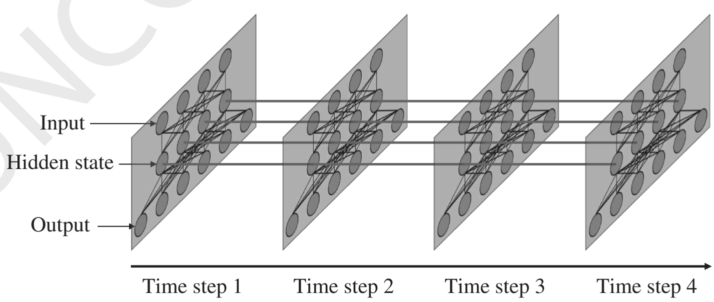
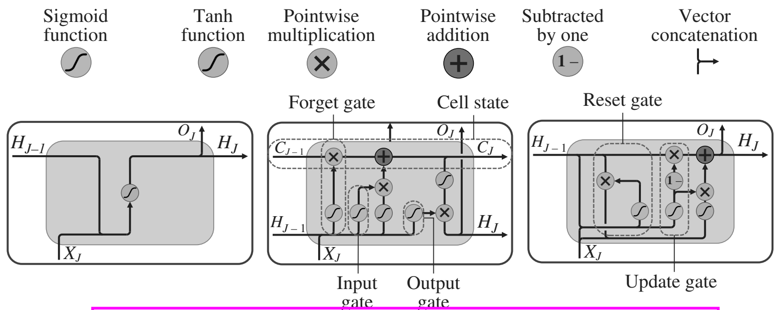


Figure 1.3 The basic RNN architecture.



(a) RNN unit (b) LSTM unit (c) GRU unit

Figure 1.4 The structure diagrams of (a) RNN unit, (b) LSTM unit, (c) GRU unit.

the input gate, and the output gate, taking responsibility for adjudicating which information should be forgotten, managing the input of new information, and supervising the generation of the hidden state, respectively. Inspired by the LSTM unit, the GRU unit modifies the gate-controlled architecture and simplifies the data flow process by omitting the memory cell, thereby alleviating the computing complexity. Moreover, the reset gate and the update gate are devised to enable the reset of information from the previous time step and regulate the degree of integration of new information.

RNNs and RNN variants, for example, bidirectional RNN (BRNN) [18], have demonstrated superior performance compared to CNNs in intelligent tasks related to time series analysis. Nonetheless, the limited attention of RNNs restricts the comprehension of vast multilingual information, while the training complexity and instability hinder the deployment of RNNs in large language models.

1.2.1.4 Transformer

The transformer model [19] was originally carried out to implement machine translation tasks, relying solely on the novel fully connected layer-enabled attention mechanism and completely dispensing with RNNs and CNNs. The basic transformer architecture is shown in Figure 1.5. From the figure, the transformer skeleton comprises multiple encoder blocks and decoder blocks. The positional encoding is injected into the embedded tokens to record the relative or absolute positions of tokens within the sequence. At the encoder, the multi-headed self-attention (MHSA) module is invoked to facilitate comprehensive token representation learning while simultaneously identifying critical tokens. Besides, it leverages a softmax activation to calculate the self-attention based on three projections of tokens, that is, *query*, *key*, and *value*, denoted as

$$\text{Attention}\left(\mathcal{Q},\ \mathcal{K},\ \mathcal{V}\right) = \text{softmax}\left(\frac{\mathcal{Q}\mathcal{K}^T}{\sqrt{\mathrm{d}_k}}\right)\mathcal{V}, \tag{1.2}$$

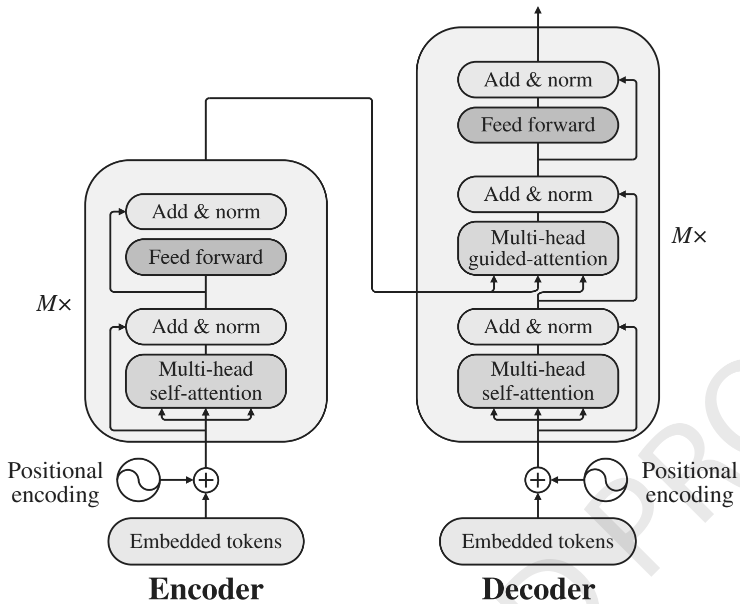


Figure 1.5 The basic transformer architecture.

where $\mathcal{Q}$, $\mathcal{K}$, and $\mathcal{V}$ represent projections *query*, *key*, and *value*, respectively. The scaling factor, $\sqrt{\mathrm{d}_k}$, is the dimension of the fully connected layer in the *key* projection.

The feed-forward network (FFN) reduces the intricacy of the MHSA output and extracts local patterns within the input sequence by introducing nonlinear transformation, which supports varied attention allocation and independent processing of each position. In FNN, the rectified linear units (ReLU) activation function is adopted between two fully connected layers. Therefore, the processing of the MHSA output, $\mathcal{O}_{\mathrm{MHSA}}$, can be expressed as

$$\mathrm{FFN}\,(\mathcal{O}_{\mathrm{MHSA}}) = \max\,(0,\ \mathcal{O}_{\mathrm{MHSA}}\mathcal{W}_1 + \boldsymbol{bias}_1)\,\mathcal{W}_2 + \boldsymbol{bias}_2, \tag{1.3}$$

where $\mathcal{W}_1$ and $\mathcal{W}_2$ are the weight matrices of the first and second fully connected layers, and $bias_1$ and $bias_2$ are the corresponding bias vectors, respectively.

Similarly, the decoder block involves three components, including the MHSA module, the multi-head guided-attention module, and the FNN. Specifically, the embedded tokens at the decoder are fed to the MHSA module after the positional encoding and produce the output as *query* position for the multi-head guided-attention module. In addition, the FNN output from the encoder is passed to the decoder and adopted as the input positions, *key* and *value*, for the multi-head guided-attention module. The FNN further refines the learned token representation and attains the transformer output. It is noteworthy that the layer normalization followed by the residual connection is employed for all modules across the transformer encoder and decoder.

The transformer architecture outperforms RNNs and CNNs on multifarious intelligent tasks, characterized by enhanced parallelizability and notably diminished training time. In the past few years, the meteoric emergence of numerous foundation models in the transformer realm, such as universal transformer [20] and vision transformer [21], was witnessed, contributing to the advancement of the NLP field and propelling the advent of the much-touted large language models, such as bidirectional encoder representations from transformers (BERT) [22] and generative pre-trained transformer 4 (GPT-4) [23].

1.2.1.5 Generative Adversarial Network

The generative adversarial network (GAN) model is an emerging technique that offers a solution for acquiring deep representations without the requisite of enormous labeled training data. It is achieved by backpropagating gradients through a pair of competing NNs to estimate the high-dimensional data distribution. The GAN framework was first proposed to generate realistic images [24], and the model structure is shown in Figure 1.6. From the figure, the GAN model includes two major parts: the generator and the discriminator. The generator has no access to real images but functions as a forger to synthesize plausible data from the latent space input. The discriminator plays the role of an expert in distinguishing between the real labels from the training set and the fake labels generated by the generator, providing a simple ground truth of the discriminative result. The ultimate goal of the generator is to create forgeries indistinguishable from the real data to deceive the discriminator. In contrast, the discriminator continuously sharpens its observation to detect fake samples, driving both entities to evolve and reinforce competitive capabilities to combat each other.

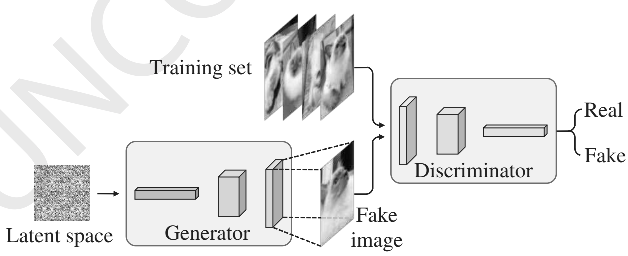


Figure 1.6 The GAN model structure for image generation.

The generator and discriminator are typically implemented using multiple NNs, such as CNNs and RNNs, for various generation tasks, adopting an alternate training manner to ensure iterative weight updates. The loss function to train the discriminator can be denoted as

$$\mathcal{L}_{\mathrm{D}} = \frac{1}{2}\log \mathfrak{T}_{\mathrm{D}}(\mathcal{S}) + \frac{1}{2}\log(1 - \mathfrak{T}_{\mathrm{D}}(\mathfrak{T}_{\mathrm{G}}(\boldsymbol{Z}))), \tag{1.4}$$

where $\mathfrak{T}_{\mathrm{D}}(\cdot)$ and $\mathfrak{T}_{\mathrm{G}}(\cdot)$ are the discriminator and the generator, respectively. $\mathcal{S}$ is the real data, and the $\boldsymbol{Z}$ is the latent space. Note that the NN parameters of the generators remain fixed throughout the discriminator's entire training process. Besides, the loss function can be varied according to different tasks.

Similarly, the loss function to train the generator can be expressed as

$$\mathcal{L}_{\mathrm{G}} = \log(1 - \mathfrak{T}_{\mathrm{D}}(\mathfrak{T}_{\mathrm{G}}(\boldsymbol{Z}))), \tag{1.5}$$

where the loss convergence is stabilized until the discriminator can no longer detect counterfeits of the generator from the real data. The NN parameters of the discriminator are fixed when training the generator.

GANs have experienced swift development over the past decade, and many variants have been developed, including conditional GANs [25] that control the generation by introducing additional information and deep convolutional GANs [26] that tackle the inherent training problems. However, further endeavors are necessary to tackle many challenges in GANs, such as the mode collapse, the training instability, and the ethical considerations regarding the generated content.

1.2.2 Deep Learning-enabled Semantic Communications

The conventional bit-oriented transmission protocol has evolved from syntax communications and undergone thriving developments from the 1G to the 5G wireless communication networks. Nonetheless, attention to semantic information has been neglected in conventional communications and ubiquitous intelligent tasks in the 5G and beyond communication era have aggravated the burden of data transmission to satisfy personalized and multifarious user requests. In this context, the semantic communication problem has been listed as one of the promising techniques for future intelligent communications to alleviate the scarcity of bandwidth resources and the explosion of data traffic [27–29]. Therefore, it is imperative to reconceive the framework of semantic communications and establish an innovative architecture capable of catering to escalating transmission requirements and serving on-demand mobile applications.

Thanks to the success of AI in recent years, a DL-enabled semantic communication paradigm has been proposed and has shown its great potential to tackle the technical challenges across many aspects of conventional communications, for example, physical-layer communications [30, 31] and wireless resource allocation [32, 33]. Such success is attributed to the learning and fitting capabilities of

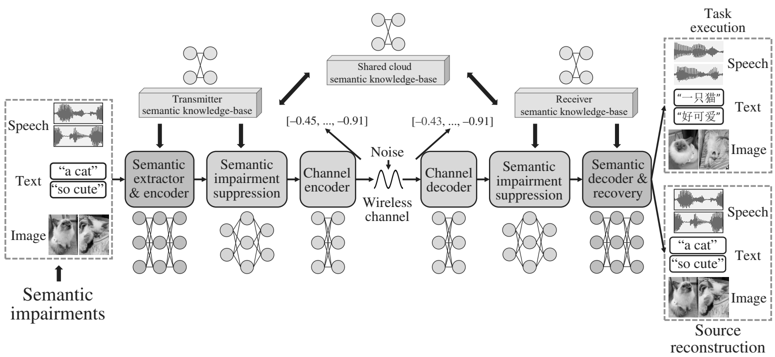


Figure 1.7 The structure of the DL-enabled semantic communication framework.

sophisticated NNs, breaking the constraint of a mathematical model to represent and exchange semantic information. A general DL-enabled semantic communication system is shown in Figure 1.7. From the figure, the semantic encoder takes responsibility for extracting the semantic information from the input source, and the semantic impairment suppression mechanism commits to perceiving and alleviating semantic impairments. The transmitter and receiver SKBs are invoked to better facilitate semantic representation and semantic recovery, respectively. The DL-enabled semantic communication paradigm requires the least semantic ambiguity between the source and the recovered messages. It maps the source message into low-dimensional semantic features, reducing the volume of transmission data and mitigating the network traffic without performance degradation. According to the cutting-edge DL-enabled semantic communications, the transmission goal could be categorized into two types: source data reconstruction and intelligent task execution. It is noteworthy that the DL-based JSCC also adopts NNs to replace conventional coding and decoding modules. The principal differences are clarified as follows to clarify the distinction between DL-based JSCC and DL-enabled semantic communications.

- The DL-based JSCC is a semantic-agnostic solution that jointly optimizes the source compression and error correction by training an integrated NN to improve the E2E system performance. Meanwhile, DL-enabled semantic communications utilize NNs to learn critical semantic information from the source and facilitate semantic transmission to convey the meaning of the transmitter to the receiver.
- The JSCC architecture only considers the physical noise, whereas semantic communications take into account both the physical noise and the specific semantic impairments. Consequently, DL-enabled semantic communications incorporate the semantic impairment suppression module to mitigate the effects of semantic impairments.

- JSCC has no SKBs. DL-enabled semantic communications synergize with SKBs to ensure high semantic fidelity transmission.

SKBs are memorable knowledge network models that are intricately structured and offer relevant semantic descriptions for data. As shown in Figure 1.7, SKBs are divided into the transmitter SKB, the shared cloud SKB, and the receiver SKB. In semantic communications, the semantic encoder produces the compressed codestream based on the learning and inference algorithm, where the transmitter SKB acts as a booster to accelerate the learning process and standardize the optimal inference path by providing multi-level semantic knowledge vectors. The description of propagation environments is shared between the transmitter and the receiver based on the shared cloud SKB to facilitate high semantic fidelity transmission. At the receiver, the resilient decoding mechanism is adopted to calibrate corrupted semantic features by sequentially incorporating received semantic knowledge vectors from the highest level to the lowest level.

DL-enabled semantic communications for source reconstruction have proven superior in tackling the performance bottlenecks of conventional communications. To achieve precise source reconstruction, the global semantic information is extracted for transmission. To serve intelligent downstream tasks, only task-related semantic features are extracted, while other irrelative features are ignored to minimize the transmitted data. Besides, the produced data is not required to exhibit precise alignment with the source data at the binary level, and it typically diverges from the source. More details of DL-enabled semantic communications and task-oriented semantic communications are introduced in Chapters 3 and 4, respectively.

1.3 Semantic Communications for Further Networks

The previous 5G has rapidly changed the lifestyle of humans and given birth to connected intelligence, which transforms society and industry into data-centric and automated. In the era of 6G and beyond, it is anticipated that the trend will be further enhanced. The fusion of digital and physical realities, combined with intelligently interconnected systems, is fueling the upcoming technological innovation. The necessary technological advancements for 6G and beyond are expected to be more intelligent, more efficient, and more secure. The semantic communication paradigm is expected to play a crucial role in the 6G and beyond, as it can provide a more efficient and intelligent way to transmit information. The semantic communication paradigm can be used to reduce the amount of data transmitted, improve the accuracy of the transmitted information, and enable more intelligent communication between devices. In Section 1.3, we discuss the potential applications of semantic communications in the 6G and beyond.

Semantic communications can support a wide range of 6G use cases as following table,[1] for example, multi-sense services, extended reality (XR), human-to-machine communications, and so forth. What these use cases have in common is intensive computation, accurate generation, and reliable communication. Due to the paradigm of semantic communications, the integration between computation, communication, and generation can effectively support these use cases. Take the XR as an example, virtual reality is generated by human interaction and requires cooperation between the edge server and the user's device. In this case, the semantic communications can extract the accurate semantics behind the human language, then transmit the meanings to the edge, and finally receive the generated data, for example, three-dimensional (3D) scenarios, from the edge. Semantic communications can reduce the data transmission and improve the accuracy of generated data. Besides, semantic communications can also support multi-sensory services, for example, multi-sensory holographic teleportation, multi-sensory haptic communications for virtual and augmented reality, and multi-sensory affective computing. These services require the accurate transmission of the human senses, for example, vision, hearing, and touch. Semantic communications can extract the accurate semantics behind the human senses, then transmit the meanings to the edge, and finally receive the generated data, for example, holographic teleportation, haptic communications, and affective computing, from the edge. Semantic communications can reduce data transmission and improve the user experience. In addition, semantic communications can also support human-centric AI, brain-to-computer interactions, and wireless robotics. These services require the accurate transmission of the human's intentions, for example, the commands, the thoughts, and the actions. Semantic communications can extract the accurate semantics behind the human's intentions, then transmit the meanings to the edge, and finally receive the generated data, for example, the AI support, the brain–computer interactions, and the robotics, from the edge. Semantic communications can improve the accuracy of interactions.

Use Case	Service
Multi-sense Services	URLLCCC, **Semantic**
Multi-sensory and mobile immersive XR	URLLCCC, **Semantic**
Multi-sensory holographic teleportation	URLLCCC, **Semantic**

1 Machine Type Communications supporting Distributed Intelligence (MMTCCxDI); Geographically Mobile Broadband (GeMBB); Ultra Reliable Low Latency Communications and Computing (URLLCCC).

Use Case	**Service**
Multi-sensory haptic communications for virtual and augmented reality	URLLCCC, **Semantic**
Multi-sensory affective computing	URLLCCC, **Semantic**
Holographic communication & telepresence	URLLCCC, **Semantic**
Remote XR education	URLLCCC, **Semantic**
Consumption of digital experiences over physical products	URLLCCC, **Semantic**
Human-centric AI support	URLLCCC, **Semantic**
Brain-to-computer interactions	URLLCCC, **Semantic**
Wireless robotics	MMTCCxDI, URLLCCC, **Semantic**
Sustainable connectivity & AI support	MMTCCxDI, GeMBB, URLLCCC, **Semantic**
Society 5.0	MMTCCxDI, GeMBB, URLLCCC, **Semantic**
High precision manufacturing, remote monitoring & control	MMTCCxDI, GeMBB, URLLCCC, **Semantic**
Intelligent disaster prediction	MMTCCxDI, GeMBB, URLLCCC, **Semantic**
Smart digital-twin environments	MMTCCxDI, GeMBB, URLLCCC, **Semantic**
Bidirectional intelligence intertwining (natural and/or artificial)	**Semantic**

References

1 C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign, IL, USA: Univ. Illinois Press, 1949.
2 C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, 1996.
3 R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information," in *Res. Lab. Electron.,* Massachusetts Inst. Technol., Cambridge, MA, USA, RLE Tech. Rep. 247, 1952.
4 P. Basu, J. Bao, M. Dean, and J. Hendler, "Preserving quality of information by using semantic relationships," *Pervasive Mob. Comput.*, vol. 11, pp. 188–202, 2014.
5 A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: an enabler of semantics-empowered communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2621–2635, 2023.

6 M. Kalfa, M. Gok, A. Atalik, et al., "Towards goal-oriented semantic signal processing: applications and future challenges," *Digit. Signal Process.*, vol. 119, pp. 103–134, 2021.

7 J. Dewey, *Experience and Nature*. Courier Corporation, 1958, vol. 471.

8 L. Wittgenstein, *Philosophical Investigations*. John Wiley & Sons, 2010.

9 M. Jeong, B. Kim, and G. G. Lee, "Semantic-oriented error correction for spoken query processing," in *Proc. IEEE Autom. Speech Recognit. Underst. Workshop (ASRU Workshop)*. IEEE, 2003, pp. 156–161.

10 L. Floridi, "Outline of a theory of strongly semantic information," *Minds Mach.*, vol. 14, no. 2, pp. 197–221, 2004.

11 J. Bao, P. Basu, M. Dean, et al., "Towards a theory of semantic communication," in *Proc. IEEE Netw. Sci. Workshop*, West Point, NY, USA, 2011, pp. 110–117.

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

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

14 Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

15 J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 79, no. 8, pp. 2554–2558, 1982.

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

17 K. Cho, B. van Merriënboer, C. Gulcehre, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empir. Methods Nat. Lang. Process. (EMNLP),* Doha, Qatar, 2014, pp. 1724–1734.

18 M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

19 A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Proces. Syst. (NeurIPS)*, vol. 30, Long Beach, CA, USA, 2017.

20 M. Dehghani, S. Gouws, O. Vinyals, et al., "Universal transformers," in *Proc. Int. Conf. Learn. Represent. (ICLR),* New Orleans, LA, USA, 2023.

21 A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR),* Vienna, Austria, 2021.

22 J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. N. Am. Chapter Assoc. Comput. Linguistics: hum. Lang. Technol. (NAACL HLT)*, vol. 1, Minneapolis, MN, USA, 2019, pp. 4171–4186.

23 J. Achiam, S. Adler, S. Agarwal, et al., "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

24 I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Proces. Syst. (NeurIPS)*, vol. 27, Montreal, QC, Canada, 2014.

25 M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

26 A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.064 34*, 2015.

27 W. Tong and G. Y. Li, "Nine challenges in artificial intelligence and wireless communications for 6G," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 140–145, 2022.

28 Y. Shi, Y. Zhou, D. Wen, et al., "Task-oriented communications for 6G: vision, principles, and technologies," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 78–85, 2023.

29 Z. Qin, X. Tao, J. Lu, et al., "Semantic communications: principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

30 H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, 2018.

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

32 H. Sun, X. Chen, Q. Shi, et al., "Learning to optimize: training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, 2018.

33 Z. Ji, Z. Qin, and X. Tao, "Meta federated reinforcement learning for distributed resource allocation," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7865–7876, 2023.