Received 1 December 2023; accepted 15 December 2023. Date of publication 20 December 2023; date of current version 8 January 2024.

Digital Object Identifier 10.1109/OJCOMS.2023.3344882

Cooperative Semantic Communication With On-Demand Semantic Forwarding

BING TANG[1], LIKUN HUANG[2], QIANG LI[1] (Member, IEEE), ASHISH PANDHARIPANDE[3] (Senior Member, IEEE), AND XIAOHU GE[1] (Senior Member, IEEE)

[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

[2]School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan 430205, China

[3]NXP Semiconductors, 5656 AE Eindhoven, The Netherlands

CORRESPONDING AUTHOR: Q. LI (e-mail: qli_patrick@hust.edu.cn)

This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 61971461, and in part by the Hubei Provincial Key Research and Development Program under Grant 2021BAA015 and Grant 2022EHB014.

(Bing Tang and Likun Huang contributed equally to this work.)

ABSTRACT In this paper, a deep learning-based cooperative semantic communication system is proposed on relay channels. In order to enhance the reliability and adaptability of the system to varying channel conditions, an on-demand semantic forwarding framework is established, where the relay attempts to recover and re-transmit the source semantic information, as required by the destination. To be specific, for determining whether a semantic forwarding is needed from the relay, a semantic similarity check (*SSC*) is proposed, based on which the degree of semantic information recovery at the destination can be accurately estimated. On the other hand, in order to effectively merge the semantic information received through different paths, a semantic combining (*SeC*) method is proposed by combining the semantic features abstracted from both the direct link and the relay link. For achieving a desirable performance trade-off between the degree of semantic information recovery and the transmit energy consumption, a new metric of semantic energy efficiency (SEE) is proposed. Simulation results verify the performance gains achieved by the proposed cooperative semantic communication system with on-demand semantic forwarding, as compared to the state-of-the-art schemes employing separate source-channel coding in low-to-medium signal-to-noise ratio (SNR) regimes. Furthermore, as compared to the case with always-forwarding, almost the same performance is achieved by the proposed on-demand forwarding, but with lower energy consumption.

INDEX TERMS Semantic communication, on-demand semantic forwarding, semantic combining, semantic similarity check, semantic energy efficiency.

I. INTRODUCTION

A. MOTIVATION

SHANNON'S information theory [1] serves as the theoretical cornerstone of modern wireless communication systems. With the rapid growth of global mobile data traffic, the channel capacity of communication systems is relentlessly approaching the Shannon limit. Conventionally, the separate source-channel coding (SSCC) paradigm is theoretically optimal, where the end-to-end transmission can be optimized by optimizing the source coding and channel coding separately [2]. However, with the emergence of various new applications, the SSCC may no longer be optimal in dealing with the goal-oriented communications subject to stringent delay, bandwidth, or energy constraints in the 6G era [3].

To meet the surging demands in 6G, semantic communication, which is usually implemented following the paradigm of joint source-channel coding (JSCC) [4], has attracted great research interests [3], [5]. Instead of precisely delivering bit sequence, semantic communication focuses on conveying the relevant and important semantic meaning of the source [6]. Compared to traditional communications

where all source information needs to be delivered intact, semantic communication is able to capture the semantic features based on the task or operation to be performed at the receiver, thus significantly reducing the amount of data transferred while maintaining the original semantics [7].

Owing to the advancement of deep learning techniques [8], numerous deep learning-based JSCC schemes have been proposed for semantic communication [9], [10], [11]. With a joint design of source coding and channel coding, the deep learning based semantic network can be trained, by designating a suitable loss function depending on the specific scenarios [6]. Through this semantic network training, the semantic communication system can not only learn to extract the relevant semantic features for the end-to-end communication task, but also communicate with the highest possible fidelity over various channels conditions [5]. Owing to the aforementioned advantages, semantic communication has been applied for the transmissions of texts, images, and videos [12], [13], [14]. However, a number of issues are still to be tackled before the full-fledged application of semantic communication [15].

Firstly, most existing studies on semantic communication [9], [10], [11] focus on relatively simple point-to-point (P2P) communication scenarios, and it remains a challenge to design efficient semantic communication systems on cooperative multi-point networks [16]. On the other hand, although semantic communication has been applied to fixed relay channels [17], [18], how to effectively improve the system performance by merging the multiple semantic information flows still remains a major challenge. Furthermore, while most existing studies focus solely on measuring the semantic fidelity [9], [10], [11], it remains a challenge to characterize the trade-off between different performance metrics, e.g., semantic fidelity and system energy consumption.

Motivated by the aforementioned technical challenges, the following key issues will be addressed in this paper: 1) How to design a cooperative semantic communication system on multi-point relay channels; 2) How to effectively leverage the semantic information flows in multi-point communication systems; 3) How to systematically evaluate and improve the energy efficiency of semantic communication system.

B. RELATED WORKS

Different semantic communication strategies have been proposed to enhance the reliability of communication systems [19]. The work of [9], [10], [11] studied semantic communication systems in P2P scenarios with text applications. To be specific, a Transformer-based [20] text semantic communication system was designed in [9], through which the semantic gap between the original and received data was minimized by restoring the meaning of the sentence. In [10], the hybrid automatic repeat request mechanism was introduced to further improve transmission reliability and achieve variable-length encoding. In [11], a semantic hybrid automatic repeat request scheme that leverages incremental knowledge was proposed to simultaneously reduce communication costs and semantic errors. It is worth noting that the above studies only considered P2P communication scenarios, which may suffer from severe performance loss when encountered with deep fading channels [21].

In addition to traditional P2P scenarios, semantic communication has also been considered in multi-point scenarios [22], [23]. To be specific, the data correlation between different users was used for object identification in [24], which improved the accuracy without increasing the transmission delay. A goal-oriented multi-user semantic communication system was considered in [25], where reliable transmission was achieved by using multi-modal data between multiple users. In [17], a semantic communication scheme based on auto-encoder was designed for text transmission over a relay channel, which helps forward the source message at the semantic level. An image transmission system based on a relay was proposed in [18], which outperformed the baseline scheme utilizing polar codes and better portable graphics compression. However, it is not straightforward how to effectively combine the signals from different points at the semantic level.

Semantic communication systems have also been studied from the perspective of resource allocation [26], [27], [28]. In order to optimize the channel allocation and the number of transmitted semantic symbols, the metric of semantic spectral efficiency was first defined for text transmissions [26]. In [27], the problem of resource allocation and information extraction was investigated for semantic communication networks, where the communication and computational energy consumption were jointly optimized. In [28], a performance metric called system throughput in message was defined and used as a target for joint optimization of user association and bandwidth allocation in heterogeneous networks. Although new metrics have been proposed to evaluate the performance of semantic communication systems from different aspects, the energy efficiency, which characterizes the performance trade-off between the amount of semantic information recovered and the system energy consumption, remains to be further explored.

C. CONTRIBUTIONS

Motivated by the aforementioned studies, in this paper a multi-point semantic communication system is proposed on a cooperative relay channel. The contributions of this paper can be summarized as follows:

- In order to improve the semantic fidelity while maintaining a reasonably low overhead, a novel on-demand semantic forwarding framework, which aims at improving the reliability and adaptability of the system to diverse channel conditions, is proposed. To be specific, focusing on a text transmission scenario, the relay attempts to recover and re-transmit the source semantic information, only when a negative acknowledgment (NAK) is received indicating that the semantic fidelity of the recovered message at the destination falls below

a predefined threshold. This is able to effectively improve the end-to-end semantic information delivery, while avoiding unnecessary energy waste. To the best of our knowledge, this is the first system of its kind.

- To implement on-demand semantic forwarding, a *SSC* that is similar to cyclic redundancy check (CRC) [29] is proposed. By deploying a quantization module at the source, the source semantic features are extracted and compressed, which are then fed into a detection network that is deployed at the destination. Then the corresponding output is compared with a predefined semantic fidelity threshold, indicating whether the relay transmission is needed. Furthermore, to leverage the signals received from the source and the relay, a novel *SeC* method is proposed for the first time. By constructively merging the semantic features abstracted from both the relay and the direct links, the proposed *SeC* can be viewed as a diversity combination at the semantic level, which provides new insights for improving the fidelity of semantic communication systems.
- Existing evaluation metrics for text transmission systems, such as BLEU [30] and Sentence Similarity (SS) [9], only assess the semantic fidelity of text recovery, while without considering how soon the text message can be delivered or how much energy is consumed in delivering the text message. In order to better evaluate the efficiency of semantic communication systems, a semantic utility function is first defined that characterizes the amount of successfully delivered semantic information per unit time. On this basis, a new metric of SEE is proposed that corresponds to a ratio between the semantic utility and the transmit energy consumption. Based on SEE, a desirable trade-off can be characterized between different performance metrics depending on the specific applications. This, to the best of our knowledge, has not been fully explored in the literature.
- Extensive experiments and performance comparisons are conducted between the proposed cooperative semantic communication and the traditional SSCC techniques on decode-and-forward (DF) relay channels. It is demonstrated that the proposed scheme outperforms the traditional schemes in terms of both the BLEU and SS at low-to-medium signal-to-noise ratio (SNR) regimes. Additionally, as compared to the case with always-forwarding, significant performance gains are achieved by the proposed on-demand semantic forwarding, especially in good channel conditions with high SNR. Furthermore, an inherent performance trade-off between the semantic fidelity and the transmit energy consumption is illustrated.

The rest of this paper is organized as follows. Section II introduces the proposed cooperative semantic communication system on relay channels. Section III describes our proposed *SSC* and *SeC* methods, based on which an energy-efficient on-demand semantic forwarding framework is established. A new performance metric of SEE is proposed in Section IV. Numerical results and analysis are presented in Section V. Finally, Section VI concludes the whole paper.

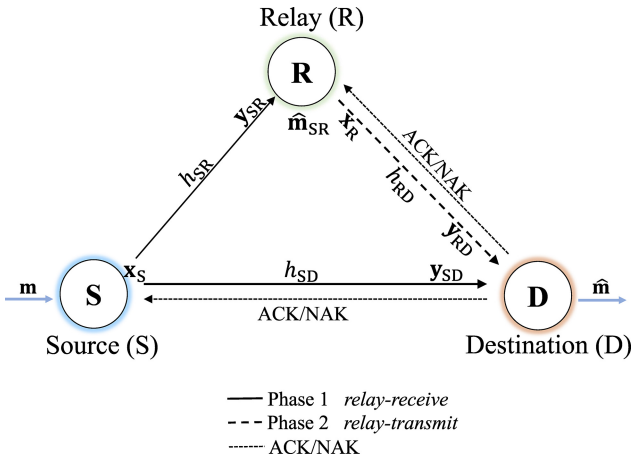


FIGURE 1. An illustration of the proposed cooperative semantic communication system on relay channels.

II. SYSTEM MODEL

A. SYSTEM DESCRIPTION

In order to enhance the reliability of communication systems in wireless deep fading channels, semantic communication is implemented on a cooperative relay channel. As shown in Fig. 1, the proposed cooperative semantic communication system consists of a source node S, a relay node R and a destination node D. Then for delivering a message from S to D, the entire communication process is divided into two successive phases [31], namely the *relay-receive* phase and the *relay-transmit* phase.

During the *relay-receive* phase, the semantic information contained in the source message is encoded and broadcast from S to R and D under power constraints. Upon the reception of the source signal, D attempts to recover the original source message. Specifically, semantic detection is performed on the signals transmitted through the S-D link at node D. If the semantic fidelity of the recovered message is beyond a predefined threshold, D sends back an acknowledgment (ACK) and S proceeds to transmitting a new message [32], [33]. Otherwise, if the semantic fidelity of the recovered message falls below the threshold, then D sends back a NAK, and the system enters the *relay-transmit* phase [34]. During this phase, the relay R attempts to extract the semantic information from the received source message and forwards it to D with best effort. Although perfect decoding cannot be guaranteed at the relay R, the message forwarded by R can be combined with the message received from S. This results in a recovered message of high semantic fidelity, by using the proposed *SeC* method that will be discussed in Section III-B.

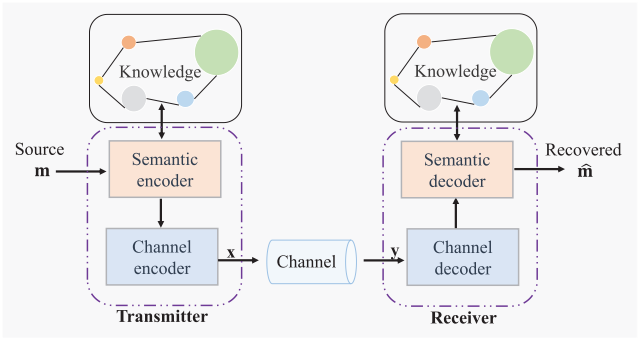


FIGURE 2. The framework of general semantic communication system.

B. CHANNEL MODEL

For ease of exposition, the channel coefficients of the links S-D, S-R, R-D are denoted as h_{SD}, h_{SR} and h_{RD}, respectively. Without loss of generality, a class of general channel models are considered, where both the effects of large-scale fading, i.e., path-loss attenuation, and small-scale fading, i.e., Rayleigh fading, are taken into consideration [35]. Thus, the channel coefficient is modeled as

$$h_{ij} = \sqrt{\Omega_{ij}}\hat{h_{ij}}, \tag{1}$$

where $\Omega_{ij} = (d_{ij}/d_{ref})^{-\upsilon_{ij}}$ represents the large-scale path loss, d_{ij} represents the physical distance between two nodes in the network, d_{ref} represents the reference distance, and υ_{ij} is the path loss exponent, for $i \in \{\mathrm{S}, \mathrm{R}\}$ and $j \in \{\mathrm{R}, \mathrm{D}\}$. For small-scale fading, the channel coefficient $\hat{h_{ij}}$ is modeled by an independent circularly symmetric complex Gaussian random variable where $\hat{h_{ij}} \sim \mathscr{CN}(0, 1)$. The transmit SNR of the link i-j is defined as $\gamma_{ij} = 10\lg(\frac{P_i}{n_{ij}})$, where P_i represents the transmit power of i and n_{ij} represents the complex additive white Gaussian noise (AWGN) with zero mean and variance σ_{ij}^2.

C. PROPOSED COOPERATIVE SEMANTIC COMMUNICATION

As shown in Fig. 2, a typical semantic communication model consists of a semantic encoder and decoder, as well as a channel encoder and decoder. With a joint design of semantic coding and channel coding, the transmitter/receiver pair can be trained with any desired loss function depending on the specific conditions [6], thus learning to not only extract the relevant semantic features for the end-to-end communication task, but also communicate with the highest possible fidelity over various channels conditions [5]. Firstly, the semantic encoder performs feature extraction on the message $\mathbf{m}$ to be sent by the source S. Then, the channel encoder maps the input signal to a symbol stream $\mathbf{x}$, which is transmitted over the physical channel with noise. The received symbol stream $\mathbf{y}$ is decoded and reconstructed as $\hat{\mathbf{m}}$, which contains the original semantic information in $\mathbf{m}$. Taking text transmission as an example, the source S sends a message $\mathbf{m}$ each time, where $\mathbf{m} = [w_1, w_2, \ldots, w_L]$. Then w_L represents the L-th word in the sentence, and L is the length of the sentence.

Generally, the semantic communication process can be divided into the following steps. Firstly, the transmitter extracts and encodes the semantic information from the source message $\mathbf{m}$. The resulting semantic symbol stream can be expressed as

$$\mathbf{x} = C_\varphi(S_\varepsilon(\mathbf{m})), \tag{2}$$

where $\mathbf{x} \in \mathbb{C}^{L\times K}$, K is the number of symbols required to represent each word, and $L \times K$ denotes the total length of the encoded message symbols. $S_\varepsilon(\cdot)$ denotes the semantic encoder network with parameter set ε, and $C_\varphi(\cdot)$ denotes the channel encoder with parameter set φ. For convenience, we use $f_i(\cdot;\alpha_i) = C_\varphi(S_\varepsilon(\cdot))$ to represent the entire JSCC process at transmitter i, where α_i represents the parameter set of the encoder. With the following power constraint [36] on the transmitted symbol stream

$$\frac{1}{L}\sum_{l=1}^{L}|x_l|^2 \le P_{\mathrm{S}}, \tag{3}$$

the received signal can be expressed as

$$\mathbf{y} = h\mathbf{x} + n, \tag{4}$$

where $\mathbf{y} \in \mathbb{C}^{L\times K}$. The receiver attempts to recover the original source message, resulting in an estimated message

$$\hat{\mathbf{m}} = S_\chi^{-1}\left(C_\delta^{-1}(\mathbf{y})\right), \tag{5}$$

where $C_\delta^{-1}(\cdot)$ denotes the channel decoder with parameter set δ, and $S_\chi^{-1}(\cdot)$ denotes the semantic decoder network with parameter set χ. Similarly, we let $f_{ij}^{-1}(\cdot;\theta_{ij}) = S_\chi^{-1}(C_\delta^{-1}(\cdot))$ denote the joint source-channel decoding process at receiver j, where θ_{ij} represents the parameter set of the decoder.

As shown in Fig. 1, when semantic communication is implemented on cooperative relay channels, S first encodes the source text message $\mathbf{m}$ into a semantically meaningful symbol stream during the *relay-receive* phase, i.e., $\mathbf{x}_{\mathrm{S}} = f_{\mathrm{S}}(\mathbf{m};\alpha_{\mathrm{S}})$. After the transmissions through wireless fading channels, the signals received at D and R can be respectively expressed as

$$\begin{aligned}\mathbf{y}_{\mathrm{SD}} &= h_{\mathrm{SD}}\mathbf{x}_{\mathrm{S}} + n_{\mathrm{SD}}\\ \mathbf{y}_{\mathrm{SR}} &= h_{\mathrm{SR}}\mathbf{x}_{\mathrm{S}} + n_{\mathrm{SR}}\end{aligned}. \tag{6}$$

Upon receiving $\mathbf{y}_{\mathrm{SD}}$, D attempts to recover the original source message $\mathbf{m}$, resulting in an estimated message of $\hat{\mathbf{m}}_{\mathrm{SD}} = f_{\mathrm{SD}}^{-1}(\mathbf{y}_{\mathrm{SD}};\theta_{\mathrm{SD}})$. Then *SSC* is performed to evaluate whether the semantic fidelity of the recovered message $\hat{\mathbf{m}}_{\mathrm{SD}}$ is above a predefined threshold. The details of *SSC* will be elaborated in Section III. As shown in Fig. 3, we have the following two cases depending on the result of *SSC*:

1. When the SNR is high or channel fading is mild, the semantic meaning of the source text message can be delivered through the S-D link with high fidelity, thus passing the *SSC*. Then D will send back an ACK signal [33], and the system enters the next *relay-receive* phase where a new message is transmitted by S.

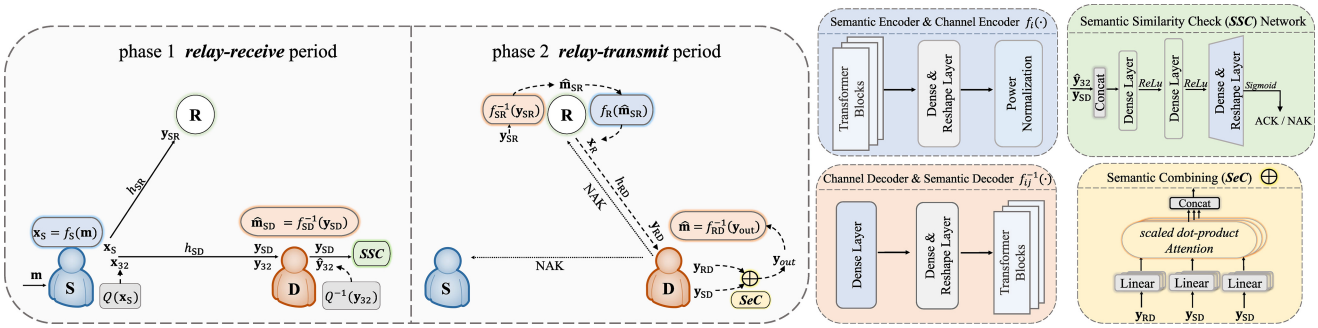


FIGURE 3. The architecture of the proposed cooperative semantic communication system with on-demand semantic forwarding.

2. At low SNR or severe channel fading causes difficulty in correctly recovering the source message $\mathbf{m}$ at D, the *SSC* fails, and in response, D will send a NAK signal [33] to both S and R. The system then enters the *relay-transmit* phase, where R attempts to first obtain an estimate $\widehat{\mathbf{m}}_{\mathrm{SR}} = f_{\mathrm{SR}}^{-1}(\mathbf{y}_{\mathrm{SR}}; \theta_{\mathrm{SR}})$, then extracts the semantic information from $\widehat{\mathbf{m}}_{\mathrm{SR}}$ as $\mathbf{x}_{\mathrm{RD}} = f_{\mathrm{R}}(\widehat{\mathbf{m}}_{\mathrm{SR}}; \alpha_{\mathrm{R}})$, which is then forwarded over the relay channel to D. The corresponding received signal at D can thus be expressed as

$$\mathbf{y}_{\mathrm{RD}} = h_{\mathrm{RD}}\mathbf{x}_{\mathrm{RD}} + n_{\mathrm{RD}}. \tag{7}$$

Together with (6) and (7), *SeC* is performed to recover the source message $\mathbf{m}$ by jointly utilizing $\mathbf{y}_{\mathrm{SD}}$ and $\mathbf{y}_{\mathrm{RD}}$. The details of *SeC* will be elaborated in Section III.

III. PROPOSED ON-DEMAND SEMANTIC FORWARDING WITH SSC AND SeC

In contrast to a simplistic scenario where the relay R always attempts to recover and forward the source semantic information to the destination D [37], an on-demand semantic forwarding framework is proposed where the relay R attempts to recover and forward the source semantic information only when required by the destination D. This mechanism is similar to the hybrid automatic re-transmission request protocol [38], where semantic forwarding is triggered upon receiving a NAK and performance gains can be obtained by exploiting both the S-D link and the R-D link. In this way, channel utilization is improved while effectively reducing the system energy consumption [39].

In order to facilitate the implementation of the proposed on-demand semantic forwarding, an efficient *SSC* method is proposed to ascertain the necessity of triggering the semantic forwarding from the relay R under varying channel conditions. On the other hand, in order to effectively integrate the semantic information received through different paths, a novel *SeC* method is proposed to enhance the fidelity of semantic restoration. Next, an explanation of these two methods will be provided.

A. A SSC METHOD

Although traditional CRC can achieve error detection at the bit level, it is unable to assess the semantic similarity of information, e.g., traditional CRC treats cases of synonym substitution and reordering of words as errors. In view of this, a *SSC* method is proposed to enable on-demand relaying at the semantic level, which leverages a quantization module and a semantic detection network deployed at S and D, respectively.

To determine the semantic correctness of the recovered message at D, the recovered message is compared with the original source message by mapping them into the semantic feature space. To be specific, the source semantic features are extracted and compressed by the quantization module into a 32-bit semantic information, which is delivered together with the source message. Then at the receiver side, the compressed semantic features are fed into the semantic detection network for similarity check, whose output is compared with a predefined threshold, indicating whether the relay re-transmission is needed.

Ideally, assuming the destination D possesses prior knowledge of $\mathbf{m}$, we employ a pre-trained language model at D to perform *SSC*. Firstly, D attempts to decode the received signal $\mathbf{y}_{\mathrm{SD}}$, given as

$$\hat{\mathbf{m}}_{\mathrm{SD}} = f_{\mathrm{SD}}^{-1}(\mathbf{y}_{\mathrm{SD}}; \theta_{\mathrm{SD}}), \tag{8}$$

where $f_{\mathrm{SD}}^{-1}(\cdot; \theta_{\mathrm{SD}})$ denotes the channel decoder represented by dense layer and the semantic decoder represented by Transformer [20] blocks, as shown in Fig. 2. Since the pre-trained BERT [40] model is capable of transforming input sentences into semantic vectors, the SS between two sentences can be quantified by using the cosine similarity between their respective semantic vectors, which is defined as

$$S_S(\mathbf{m}, \hat{\mathbf{m}}_{\mathrm{SD}}) = \frac{BERT(\mathbf{m})BERT(\hat{\mathbf{m}}_{\mathrm{SD}})^T}{|BERT(\mathbf{m})||BERT(\hat{\mathbf{m}}_{\mathrm{SD}})|}, \tag{9}$$

where $S_S(\mathbf{m}, \hat{\mathbf{m}}_{\mathrm{SD}}) \in [0,1]$, 1 means the highest similarity and 0 means there is no similarity between $\mathbf{m}$ and $\hat{\mathbf{m}}_{\mathrm{SD}}$. Although perfect decoding cannot be guaranteed, the semantic fidelity of the recovered message can be evaluated by the cosine distance in the semantic space. For ease of exposition, we let ϕ_{th} denote the predefined SS threshold. To be specific,

$$S_S(\mathbf{m}, \hat{\mathbf{m}}_{\mathrm{SD}}) \geq \phi_{\mathrm{th}} \tag{10}$$

indicates that the recovered sentence $\hat{\mathbf{m}}_{\mathrm{SD}}$ is very similar to the original sentence $\mathbf{m}$ in the semantic level. In other words, the recovered sentence $\hat{\mathbf{m}}_{\mathrm{SD}}$ passes the *SSC*, and then D sends back a ACK to both S and R. Otherwise if

$$S_S\left(\mathbf{m}, \hat{\mathbf{m}}_{\mathrm{SD}}\right) < \phi_{\mathrm{th}}, \tag{11}$$

it means the recovered sentence $\hat{\mathbf{m}}_{\mathrm{SD}}$ fails to pass the *SSC*, and then D sends back a NAK to trigger the semantic forwarding by the relay.

Practically, the original message $\mathbf{m}$ cannot be known *a priori* at D. To tackle this issue, we draw inspiration from [10] and propose a quantization module at S and a semantic detection network at D, which bears resemblance to the traditional CRC.

1) QUANTIZATION MODULE AT S

Specifically, at node S, the original message $\mathbf{m}$ is initially encoded and transformed into a 32-bit information by using a quantization module, which can be expressed as

$$\mathbf{x}_{32} = Q(f_{\mathrm{S}}(\mathbf{m}; \alpha_{\mathrm{S}})), \tag{12}$$

where the encoded features $\mathbf{x}_{32}$ are passed through a quantization module $Q(\cdot)$, to compress the source semantic features to 32 bits. The purpose is to retain the semantic characteristics of the original message while reducing the transmission cost. The quantization module includes weight quantization and activation function quantization. Specifically, to quantize the weights, the maximum and minimum values of the weights are first calculated, denoted by $\max(\mathrm{W}^{(N)})$ and $\min(\mathrm{W}^{(N)})$. Then the following formula is applied

$$\tilde{\mathrm{W}}_{o,p}^{(N)} = \mathrm{round}\left(q_w\left(\mathrm{W}_{o,p}^{(N)} - \min\left(\mathrm{W}^{(N)}\right)\right)\right), \tag{13}$$

where $\mathrm{W}_{o,p}^{(N)}$ represents the weight connecting the o-th neuron in layer $N+1$ to the p-th neuron in layer N, and q_w is the scaling factor that maps the dynamic range of floating-point numbers to M-bit integers, which is given by

$$q_w = \frac{2^M - 1}{\max\left(\mathrm{W}^{(N)}\right) - \min\left(\mathrm{W}^{(N)}\right)}. \tag{14}$$

To estimate the gradient of the quantized weights during backpropagation, the Straight-Through Estimator [41] is employed. After quantizing the weights by using the above equation, the activations are also quantized with the same method. Since the output of the activation function may have outliers, exponential moving average [42] is used to eliminate the impact of outliers, as described in [43]. For ease of exposition, the entire quantization process is denoted as $Q(\cdot)$, with the corresponding reverse de-quantization process denoted by $Q^{-1}(\cdot)$.

Then, $\mathbf{x}_{32}$ is delivered together with $\mathbf{x}_{\mathrm{S}}$ from S to D. As expressed in (4), the corresponding received signal is denoted by $\mathbf{y}_{32}$, which is dequantized as

$$\hat{\mathbf{y}}_{32} = Q^{-1}(\mathbf{y}_{32}). \tag{15}$$

2) SEMANTIC DETECTION NETWORK AT D

To facilitate *SSC*, a semantic detection network $F_{\mathrm{sim}}(\cdot)$ is proposed at D to detect the received $\mathbf{y}_{\mathrm{SD}}$ at the semantic level, which can be expressed as

$$S_{32}\left(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\mathrm{SD}}\right) = F_{\mathrm{sim}}\left(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\mathrm{SD}}; \gamma_{S_{32}}\right), \tag{16}$$

where $F_{\mathrm{sim}}(\cdot; \gamma_{S_{32}})$ denotes the semantic detection network consisting of three fully connected layers with parameter set $\gamma_{S_{32}}$, as shown in Fig. 3. The size of the hidden layer is two times that of the input, by utilizing the ReLU activation function. Subsequently, the output is passed through a Sigmoid function to produce a single value between 0-1. That is, by feeding the received signal $\mathbf{y}_{\mathrm{SD}}$ and the quantized $\hat{\mathbf{y}}_{32}$ that retains the original semantic information into the detection network, *SSC* is accomplished and the output of the network is denoted by $S_{32}(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\mathrm{SD}})$, where $S_{32}(\cdot) \in [0,1]$.

Similarly, with a predefined semantic detection threshold ϕ_{32}, $S_{32}(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\mathrm{SD}}) \geq \phi_{32}$ indicates that the quantized and compressed semantic features are well retained at the destination and the recovered sentence passes the *SSC*. Otherwise if $S_{32}(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\mathrm{SD}}) < \phi_{32}$, the *SSC* fails and the semantic forwarding is triggered.

B. A SeC METHOD

In traditional multi-point communication scenarios, the maximum ratio combining (MRC) [44] is commonly employed to combine the signal from different paths, aiming to enhance the decoding performance at the receiving end. Inspired by MRC, the proposed *SeC* method can be viewed as a diversity combining method at the semantic feature level, which provides new insights for improving the performance of multi-point semantic communication systems. Different from MRC that combines signals at the physical layer, the proposed *SeC* is able to abstract the semantic features from both the S-D link and the R-D link at the semantic layer, which are then constructively integrated to recover a better version of the source message.

Upon entering the *relay-transmit* phase, the relay R attempts to first decode $\mathbf{y}_{\mathrm{SR}}$, and subsequently re-encodes the decoded message $\hat{\mathbf{m}}_{\mathrm{SR}}$ into a symbol stream $\mathbf{x}_{\mathrm{RD}}$ to be transmitted, given as

$$\begin{aligned} \hat{\mathbf{m}}_{\mathrm{SR}} &= f_{\mathrm{SR}}^{-1}(\mathbf{y}_{\mathrm{SR}}; \theta_{\mathrm{SR}}) \\ \mathbf{x}_{\mathrm{RD}} &= f_{\mathrm{R}}\left(\hat{\mathbf{m}}_{\mathrm{SR}}; \alpha_{\mathrm{R}}\right) \end{aligned}. \tag{17}$$

Since the channel conditions of the S-R and R-D links are generally better than that of the S-D link, the corresponding received signal $\mathbf{y}_{\mathrm{RD}}$ retains a greater amount of semantic information in general as compared to that received from the direct link. Even so, those signals acquired from the direct link also contain certain aspects of the original semantics that may supplement the entire semantic information recovery. Thus, there is a pressing need for a method that can merge the semantic information from both links at the semantic level, thereby achieving improved semantic recovery performance at D.

To better combine the semantic information received from both the relay link and the direct link, we employ an attention mechanism [45], which combines the abstracted semantic features as

$$\mathbf{y}_{\text{out}} = \text{Attn}(\mathbf{y}_{\text{RD}}, \mathbf{y}_{\text{SD}}, \mathbf{y}_{\text{SD}}; \vartheta_{\text{SeC}}), \tag{18}$$

where the attention $\text{Attn}(\cdot; \vartheta_{\text{SeC}})$ is constructed through operations such as dot product and linear mapping with parameter set ϑ_{SeC}, as detailed in Fig. 2. As a basic building block that is widely adopted in neural network models, the attention mechanism is a highly effective method for combining the representations of relevant vectors, which can be expressed as

$$\text{Attn}(q, k, v) = \text{softmax}\left(\frac{qk^T}{\sqrt{d_k}}\right)v, \tag{19}$$

where q, k, and v denote the query, key, and value vectors respectively, and d_k represents the dimension of the key vectors. That is, for each query q, the similarity between q and k is computed, and then the softmax function is applied to convert the similarities into weights. Then, these weights are used to perform a weighted sum of the values v, thus obtaining the final attention output. The fusion and representation of different features can be achieved by using the self-attention mechanism by connecting different positions in a single sequence via dot product and linear mapping, based on which a representation of the sequence can be computed subsequently [46]. The whole process of obtaining the signal $\mathbf{y}_{\text{out}}$, which combines semantic information from different links, is denoted as *SeC*. With $\mathbf{y}_{\text{out}}$, an estimate of $\mathbf{m}$ can be decoded as

$$\hat{\mathbf{m}} = f_{\text{RD}}^{-1}(\mathbf{y}_{\text{out}}; \theta_{\text{SeC}}). \tag{20}$$

C. OPTIMIZATION OBJECTIVES AND TRAINING PROCESS

With the employment of deep neural networks, an end-to-end design is performed throughout the entire system, as shown in Fig. 3. The detailed training process is divided into the following steps:

1) TRAINING SEMANTIC ENCODER AND DECODER

This step involves training semantic encoders $f_i(\cdot)$ and semantic decoders $f_{ij}^{-1}(\cdot)$ belonging to different links, where $i \in \{\text{S}, \text{R}\}$, $j \in \{\text{R}, \text{D}\}$, and $i \neq j$. The encoder and decoder on any link are trained end-to-end in a similar way. Taking the S-D link as an example, the semantic encoder $f_{\text{S}}(\cdot; \alpha_{\text{S}})$ at S and the semantic decoder $f_{\text{SD}}^{-1}(\cdot; \theta_{\text{SD}})$ at D are trained simultaneously. The optimization objective of this process can be expressed as

$$\begin{aligned}&\mathscr{L}_{CE}\left(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}; \alpha_{\text{S}}, \theta_{\text{SD}}\right)\\ &= -\sum_{l=1}^{L} q(w_l)\log(p(w_l)) + (1 - q(w_l))\log(1 - p(w_l)),\end{aligned} \tag{21}$$

where $\mathscr{L}_{CE}(\cdot)$ represents the cross-entropy loss function [12]. $q(w_l)$ is the actual probability that the word w_l appears in message $\mathbf{m}$, and $p(w_l)$ is the predicted probability that w_l appears in the recovered message $\hat{\mathbf{m}}_{\text{SD}}$. The model is trained by updating α_{S} and θ_{SD}, with the objective of minimizing the cross-entropy loss function $\mathscr{L}_{CE}(\cdot)$. This training approach is applied to other encoders and decoders on different links similarly.

2) TRAINING *SeC* NETWORK

After training the semantic encoders and decoders for each link, the *SeC* network is further trained at D to effectively integrate the semantic information from different links. Utilizing the previously trained semantic encoder and decoder, the combined semantic feature $\mathbf{y}_{\text{out}}$ is acquired by the *SeC* network $\text{Attn}(\cdot; \vartheta_{\text{SeC}})$ as given in (18). Subsequently, according to (20), the semantic decoding network $f_{\text{RD}}^{-1}(\cdot; \theta_{\text{SeC}})$ deployed at node D generates an estimated $\hat{\mathbf{m}}$. The parameters ϑ_{SeC} and θ_{SeC} of the corresponding networks are updated by minimizing the error between the estimated $\hat{\mathbf{m}}$ and the original $\mathbf{m}$ through training, which can be expressed as

$$\begin{aligned}&\mathscr{L}_{CE}\left(\mathbf{m}, \hat{\mathbf{m}}; \vartheta_{\text{SeC}}, \theta_{\text{SeC}}\right)\\ &= -\sum_{l=1}^{L} q(w_l)\log(p(w_l)) + (1 - q(w_l))\log(1 - p(w_l)).\end{aligned} \tag{22}$$

It is worth noting that, during the training of $\text{Attn}(\cdot)$, the relevant parameters of the previously trained semantic encoding and decoding network are fixed.

3) TRAINING SEMANTIC DETECTION NETWORK IN *SSC*

In order to discern the semantic relevance between different pairs of sentences without knowing $\mathbf{m}$ *a priori*, the results obtained through the pre-trained model in (9) are used as labels to guide the subsequent training of the semantic detection network $F_{\text{sim}}(\cdot; \gamma_{S_{32}})$ in (16), as expressed below

$$label\left(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}\right) = \begin{cases} 1, & S_S\left(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}\right) \geq \phi_{\text{th}} \\ 0, & S_S\left(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}\right) < \phi_{\text{th}} \end{cases}, \tag{23}$$

where $S_S(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}) \geq \phi_{\text{th}}$ indicates that the SS between the recovered sentence $\hat{\mathbf{m}}_{\text{SD}}$ and the original sentence $\mathbf{m}$ is very high, therefore the label is set to 1. Otherwise when $S_S(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}) < \phi_{\text{th}}$, the SS between $\mathbf{m}$ and $\hat{\mathbf{m}}_{\text{SD}}$ is low, then the label is set to 0. The semantic detection network is trained by minimizing the cross-entropy loss $\mathscr{L}_{CE}(\cdot)$, which can be expressed as

$$\arg\min_{\gamma_{S_{32}}} \mathscr{L}_{CE}\left(label(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}}), S_{32}(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\text{SD}})\right). \tag{24}$$

The training process aims to align the output distributions of $S_{32}(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\text{SD}})$ and $S_S(\mathbf{m}, \hat{\mathbf{m}}_{\text{SD}})$. Then the trained $F_{\text{sim}}(\cdot)$ is able to determine the semantic fidelity of the recovered message, i.e., $S_{32}(\hat{\mathbf{y}}_{32}, \mathbf{y}_{\text{SD}})$ as given in (16), by using the 32-bit encoded information $\hat{\mathbf{y}}_{32}$ and the received signal $\mathbf{y}_{\text{SD}}$.

IV. PROPOSED SEMANTIC ENERGY EFFICIENCY METRIC

Traditionally, symbol error rate or bit error rate is commonly used as a performance metric to characterize the reliability of communication systems. For text transmission, however, traditional performance metrics are incapable of reflecting the system performance in the semantic level. Although BLEU score [30] and SS [9] have been proposed, these performance metrics solely measure the degree of semantic information recovery of the text sentences. In order to better characterize the efficiency of semantic communication systems, next we propose a new metric of SEE, based on which new insights can be gained by trading off between the achieved semantic fidelity and the energy consumption of the system.

It is worth noting that the proposed semantic encoding/decoding relies on deep neural network training, which usually consumes a considerable power that may not be feasible or affordable on wireless mobile devices [43]. Thus, an edge computing scenario is assumed where the energy-consuming semantic encoding/decoding operations can be offloaded to the network edge in the proximity, which is a widely adopted assumption in existing studies [47], [48], [49]. That is, for fair comparisons between the proposed semantic communication and traditional communications, only the transmit energy consumption is considered for both of them [26].

A. SEMANTIC UTILITY

The transmission rate of bit information is measurable, but the definition or characterization for the transmission rate of semantic information is not straightforward. Next, we define a metric of semantic utility, which corresponds to the average amount of semantic information that is successfully recovered at the destination per unit time.

Taking text transmission as an example, previous works [26], [50] have often employed SS as a metric to characterize the fidelity of semantic information in the transmitted messages. As defined in (9), SS depends mainly on the average number of semantic symbols K employed per word, i.e., the adopted semantic encoding/decoding scheme, and the SNR of the transmit signal γ. In other words, the SS between two sentences $\mathbf{m}$ and $\hat{\mathbf{m}}$ can be expressed as a function of K and γ, i.e., $S_S(\mathbf{m}, \hat{\mathbf{m}}) = \xi(K, \gamma)$.

For ease of exposition, we define I, measured in *semantic unit (Sut)*, as the expected amount of semantic information contained in the transmitted sentence [26]. Since K denotes the average number of semantic symbols used per word and L denotes the expected number of words in the transmitted sentence, $K \cdot L$ corresponds to the number of semantic symbols contained in a sentence, and $\frac{I}{K \cdot L}$ corresponds to the average amount of semantic information contained in each semantic symbol within a sentence. Then given a symbol rate R_s $(Symbols/s)$, the corresponding semantic utility can be expressed as

$$U(K, \gamma) = R_s \cdot \frac{I}{K \cdot L} \xi(K, \gamma). \tag{25}$$

In order to characterize how soon the source message can be delivered to the destination, we let B denote the number of sentences transmitted per unit time. Then the symbol rate R_s can be rewritten as

$$R_s = K \cdot L \cdot B. \tag{26}$$

For a given K, it can be observed that $\xi(K, \gamma)$ is monotonically non-decreasing with γ [50], and taking values within the range of $\xi(K, \gamma) \in [\xi_{\min}, \xi_{\max}]$, where $\xi_{\min}$ and $\xi_{\max}$ denote the minimum and maximum values and $0 < \xi_{\min} < \xi_{\max} < 1$. To be specific, at high SNR with almost perfect semantic information recovery, the SS $\xi(K, \gamma)$ is upper bounded by the maximum that is usually smaller than 1, e.g., $\xi_{\max}$=0.9. Conversely, at low SNR with almost no semantic information recovered, $\xi(K, \gamma)$ is lower bounded by the minimum that is usually greater than 0, e.g., $\xi_{\min}$=0.2. In other words, the original SS $\xi(K, \gamma)$ defined in (25) is not sensitive enough to characterize the varying semantic utility with γ.

Hence, in order to better reflect the fine-grained variations in semantic utility, a linear mapping approach is employed to scale $\xi(K, \gamma)$ for a given K. The specific procedure is given as follows:

$$\widetilde{\xi}(K, \gamma) = \frac{\xi(K, \gamma) - \xi_{\min}}{\xi_{\max} - \xi_{\min}}. \tag{27}$$

Through this linear mapping, we have $\widetilde{\xi}(K, \gamma) \in [0, 1]$, which means that the semantic recovery capability of the system can be better differentiated. Then the semantic utility can be further written as

$$U(K, \gamma) = R_s \cdot \frac{I}{K \cdot L} \widetilde{\xi}(K, \gamma) = I \cdot B \cdot \widetilde{\xi}(K, \gamma). \tag{28}$$

Subject to a transmit power constraint, a higher semantic utility $U(K, \gamma)$ indicates that more semantic information can be reliably delivered and recovered at the destination per unit time.

B. SEMANTIC ENERGY EFFICIENCY

To better evaluate the relationship between the achieved semantic utility and the transmit energy consumption of the system, we define SEE as

$$\eta_{\text{SEE}} = \frac{U(K, \gamma)}{E}, \tag{29}$$

where E denotes the transmit power consumed by the considered semantic communication system. SEE indicates the amount of semantic utility that is obtained per unit energy consumption, and is measured in *Suts/Joule*. Depending on the specific relay forwarding mechanism, the SEE will be analyzed in detail in the following.

1) BASELINE CASE WITH ALWAYS FORWARDING

For this case, a message is delivered through S-R-D link over two phases, then the corresponding transmit energy consumption is given as

$$E = \frac{1}{2}(P_{\mathrm{S}} + P_{\mathrm{R}}), \tag{30}$$

where P_{S} and P_{R} denote the power consumed by S and R, respectively. The corresponding SEE can thus be expressed as

$$\eta_{\mathrm{SEE}} = \frac{U}{\frac{1}{2}(P_{\mathrm{S}} + P_{\mathrm{R}})}. \tag{31}$$

2) PROPOSED ON-DEMAND SEMANTIC FORWARDING

For ease of exposition, we define p_f as the probability that the on-demand semantic forwarding is triggered. Then with probability p_f, the transmit energy consumption is given as

$$E = \frac{1}{2}(P_{\mathrm{S}} + P_{\mathrm{R}}), \tag{32}$$

and with probability $(1 - p_f)$, the transmit energy consumption is given as

$$E = \frac{1}{2}P_{\mathrm{S}}. \tag{33}$$

Taking into account of both the above cases, the overall SEE can be expressed as

$$\eta_{\mathrm{SEE}} = (1 - p_f)\frac{U_{\mathrm{SD}}}{\frac{1}{2}P_{\mathrm{S}}} + p_f \frac{U_{\mathrm{SeC}}}{\frac{1}{2}(P_{\mathrm{S}} + P_{\mathrm{R}})}, \tag{34}$$

where U_{SD} denotes the semantic utility gained from the direct link S-D only, and U_{SeC} denotes the semantic utility gained by using the proposed *SeC* of both the S-D and R-D links.

3) EQUIVALENT SEE UNDER TRADITIONAL SOLUTIONS

For fair comparisons between the proposed cooperative semantic communication that corresponds to a JSCC and the traditional SSCC strategies, the equivalent SEE achieved by traditional communication systems will be analyzed next.

For text transmissions, the bit rate R_{B} under traditional communication system is defined as

$$R_{\mathrm{B}} = \mu \cdot L \cdot B, \tag{35}$$

where μ represents the average number of bits used per word, which is determined by the encoding scheme. Then the equivalent semantic utility function can be derived as

$$U_{\mathrm{B}} = R_{\mathrm{B}} \cdot \frac{I}{\mu \cdot L} \cdot \xi_{\mathrm{B}} = I \cdot B \cdot \xi_{\mathrm{B}}, \tag{36}$$

where $\mu \cdot L$ corresponds to the total number of bits used to encode a sentence, $\frac{I}{\mu \cdot L}$ corresponds to the amount of semantic information contained per bit, and ξ_{B} corresponds to the SS between the successfully decoded message and the original message by using the conventional scheme. It is worth noting that for conventional communication where classic channel coding is adopted, e.g., Low Density Parity Check Code (LDPC) [51], either a message is successfully decoded when all errors are corrected, i.e., $\xi_{\mathrm{B}} = 1$, or no message can be recovered when channel coding fails to correct all errors, i.e., $\xi_{\mathrm{B}} = 0$. For ease of exposition, we define p_{SD} as the probability that the decoded bit sequence passes the CRC at D. Then with probability p_{SD}, the transmit energy consumption is given as

$$E = \frac{1}{2}P_{\mathrm{S}}. \tag{37}$$

Otherwise with probability $(1 - p_{\mathrm{SD}})$, the relay attempts to decode and forward the received bit sequence to D, for which we define p_{MRC} as the probability that the decoded bit sequence by using MRC passes the CRC at D. Then the corresponding transmit energy consumption is given as

$$E = \frac{1}{2}(P_{\mathrm{S}} + P_{\mathrm{R}}). \tag{38}$$

Thus, we can obtain the overall SEE achieved by the conventional SSCC schemes as

$$\eta_{\mathrm{SEE}} = p_{\mathrm{SD}} \cdot \frac{I \cdot B}{\frac{1}{2}P_S} + (1 - p_{\mathrm{SD}}) \cdot p_{\mathrm{MRC}} \cdot \frac{I \cdot B}{\frac{1}{2}(P_S + P_R)}. \tag{39}$$

V. NUMERICAL RESULTS

To comprehensively evaluate the performance of the proposed cooperative semantic communication system with on-demand semantic forwarding, we conducted the following verifications in the simulations:

- Based on the same topology of the cooperative relay network S-R-D, the proposed semantic communication, which corresponds to a JSCC, was compared with the state-of-the-art SSCC. The performance of end-to-end semantic fidelity was evaluated by using BLEU and SS metrics, where significant performance improvements were achieved by the proposed scheme in low-to-medium SNR regimes. The generality of the conclusions was also demonstrated by evaluating the impact of different channel conditions.
- The proposed on-demand semantic forwarding was compared with the case with always forwarding. While almost the same semantic fidelity was achieved by these two schemes in terms of BLEU or SS, obvious performance gains in terms of SEE was achieved by the proposed on-demand semantic forwarding as compared to the case with always forwarding. Furthermore, with an equivalent transformation, the SEE achieved by the traditional SSCC schemes with CRC and MRC was also illustrated, which verified the effectiveness of the proposed *SSC* and *SeC* methods.
- An inherent performance trade-off between the semantic fidelity and SEE is demonstrated with varying values of transmit power. Although the semantic fidelity is in general monotonically non-decreasing with the transmit power, the performance improvement decays gradually due to the limitations of the network. This leads to a

TABLE 1. The settings of the simulations.

Parameters	Value
Transmit power, $P_S = P_R = P$	30 dBm [26]
Noise power, $\sigma_{SD}^2 = \sigma_{SR}^2 = \sigma_{RD}^2 = \sigma^2$	22 dBm [26]
Average number of symbols per word, K	32 symbols/word [26]
Path loss exponent of S-R and R-D link, $\upsilon_{SR}, \upsilon_{RD}$	2.2 [52]
Path loss exponent of S-D link, υ_{SD}	2.6 [52]
Number of sentences transmitted per unit of time, B	256 sentences/s
Quantization order, M	5
Distance between S and D, d_{SD}	160 m
Distance between S and R, R and D, d_{SR}, d_{RD}	120 m
Reference distance, d_{ref}	100 m
Sentence similarity threshold, ϕ_{th}	0.85
Semantic dectection threshold, ϕ_{32}	0.5

degradation in SEE when an unnecessarily high transmit power is adopted. Thus, the transmit power needs to be prudentially designed for achieving a desirable performance between the semantic fidelity and SEE, depending on the specific applications.

A. SIMULATION SETTINGS

In the simulations, the experiments are conducted on the European Parliament dataset [53]. The dataset was preprocessed to create a set of sentences ranging from 4 to 30 words in length for training and testing purposes. In the experiment, only the English text was selected for training, totaling 70,906 sentences. The number of epochs for training is set to 300, i.e., 70,906 sentences are used repeatedly for 300 rounds of training. The mentioned SNR values in the simulation results all refer to the transmit SNR.

For the proposed semantic communication, similar network configurations are adopted as in [12], where the semantic encoder is set to be a stack of 4 Transformer blocks, each with 8 heads. Two fully connected layers are used to extract and map the semantic features into semantic symbols. The decoder at the receiving end is configured with the same structure as the transmitter. For ease of reference, the settings of system parameters are listed in Table 1, unless otherwise specified. For the traditional SSCC schemes, Huffman [54] coding is used for source coding, Turbo [55] and LDPC [51] coding are selected as channel coding schemes. Due to the varying lengths of transmitted sentences, the interleaver indices for Turbo coding are set to be adaptive, and the decoder is configured to perform 10 iterations. LDPC adopts the BG1 code and the maximum number of decoding iterations is set to 25. In the simulation experiments, we assume that the S-D link experiences relatively high path loss, and the S-R-D link has favorable channel conditions. The 32-QAM modulation is used.

B. SIMULATION RESULTS

1) EVALUATION OF SEMANTIC FIDELITY

The semantic recovery capability of the proposed cooperative semantic communication and the traditional cooperative communication under the same number of transmitted sentences is evaluated, as shown in Fig. 4 and Fig. 5 by using BLEU and SS, respectively. Different transmit SNR are achieved by keeping the transmit power constant, i.e., $P = 30$ dBm, while varying the noise power [9]. For comparison purposes, the following schemes are also simulated. Semantic S-R-D refers to the method where messages are transmitted solely through the relay link S-R-D without the presence of *SeC*, Semantic S-D refers to the scheme where messages are transmitted exclusively through the direct link S-D, Always-Forwarding refers to the scheme where R always attempts to recover and forward messages to D where *SeC* is used.

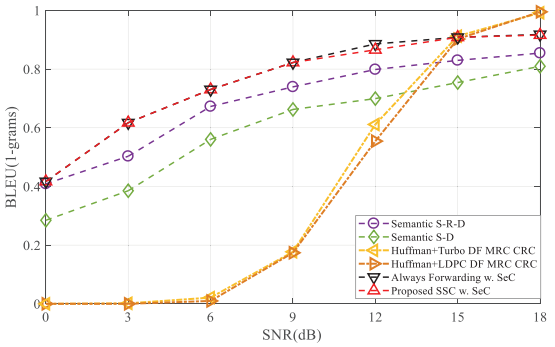


FIGURE 4. The BLEU under different transmission schemes.

From the simulation results, it can be observed that the proposed on-demand semantic forwarding, which combines *SSC* and *SeC*, achieves significant performance improvements in terms of both BLEU and SS, as compared to schemes that solely rely on semantic recovery from S-R-D link or S-D link. In comparison to Always-Forwarding, almost identical semantic information recovery is achieved by the proposed scheme. On the other hand, the traditional SSCC strategies using Huffman source coding and Turbo and LDPC channel coding are also simulated. For fair comparisons, MRC is used to merge the content from multiple links, and CRC is employed to detect the decoded sentences at D. It is observed that the proposed scheme exhibits significant performance advantages at low-to-medium SNR, while the traditional scheme demonstrates stronger semantic recovery capability at high SNR.

2) EVALUATION OF SEE

Subsequently, to thoroughly showcase the superior energy efficiency of the proposed on-demand relaying scheme, an exhaustive performance analysis was conducted to compare various communication systems in terms of the proposed metric of SEE, as shown in Fig. 6. It is observed that our proposed quantized module exhibits only marginal performance degradation as compared to the baseline model, indicating that the receiver is capable of reliably performing *SSC* by using the proposed $F_{\text{sim}}(\cdot; \gamma_{S_{32}})$ semantic detection network without prior knowledge of $\mathbf{m}$.

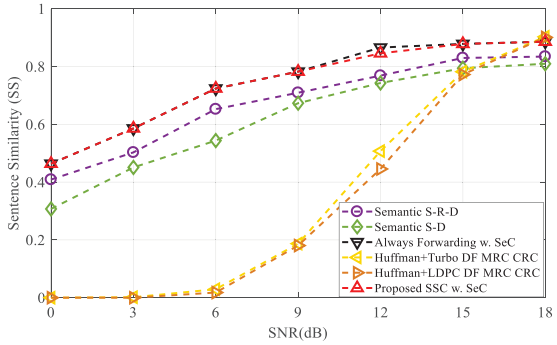


FIGURE 5. The SS under different transmission schemes.

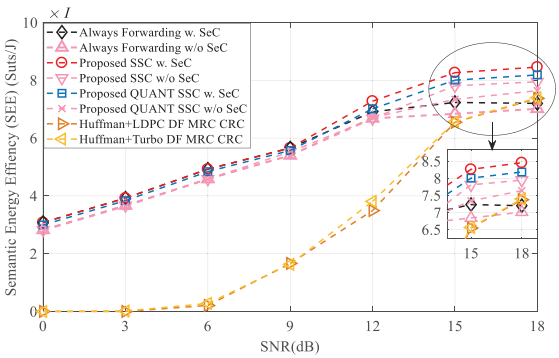


FIGURE 6. The SEE under different transmission schemes.

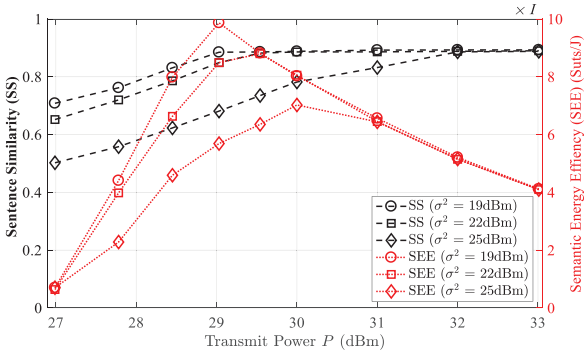


FIGURE 7. The SS and SEE achieved by the proposed cooperative semantic communication system with varying values of P with σ^2 =19, 22, 25dBm.

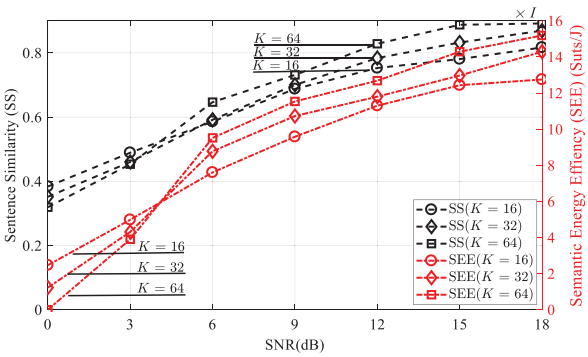


FIGURE 8. The SS and SEE of the cooperative semantic network under different K.

A notable observation is that the scheme utilizing *SSC* without *SeC* achieves higher SEE compared to the Always-Forwarding scheme at high SNR. This can be attributed to the fact that, under high SNR conditions, the message can be recovered with high SS through the direct link S-D only. By leveraging *SSC*, the number of relay-assisted transmissions can be reduced, thereby enhancing the SEE. On the other hand, it can be observed that our semantic on-demand forwarding scheme achieves better results in terms of SEE as compared to the traditional SSCC strategies. Particularly, in low SNR scenarios, the traditional schemes often fail in CRC detection, resulting in the inability to recover the transmitted sentences at the receiver side, leading to a SEE of zero.

3) PERFORMANCE TRADE-OFF BETWEEN SEE AND SEMANTIC FIDELITY

Figure 7 illustrates the performance of the proposed on-demand semantic forwarding in terms of both SS and SEE under different P. It can be observed that as P increases, the proposed scheme exhibits a rapid improvement on SS. However, when P increases to a certain level, the semantic recovery capability approaches the performance upper limit of the semantic network, which is constrained by the network structure and parameter settings, thus reaching a plateau. In terms of the SEE, when P is low, the model exhibits poor semantic recovery capability, resulting in poor SEE performance. As P increases, the SEE of the model also improves accordingly. When P increases to a level that enables the model to achieve a competent degree of semantic recovery, the semantic utility tends to stabilize. During that period, the SEE decreases with the increase in P. This indicates that the transmit power needs to be properly selected for reaching a balance between different performance metrics like SS and SEE.

Fig. 8 illustrates the performance of our proposed on-demand semantic forwarding when each word is represented by a different number K of semantic symbols. The fixed transmit power is configured as $P = 27$ dBm and different SNR is achieved varying σ^2. It can be observed that at low SNR, a smaller value of K, e.g., $K = 16$, yields better SEE performance for the model. This can be attributed to the fact that under the same transmit power, a smaller value of K results in a higher average power for each semantic symbol. As a result, the model can better capture the semantic information of sentences, leading to improved semantic utility and consequently higher SEE. As SNR increases, a larger value of K, e.g., $K = 64$, gradually gains an advantage in terms of SEE performance. This is reasonable in that a larger value of K signifies a stronger capability of the semantic model in extracting semantic information. Thus, depending on the channel conditions, the value of K needs to be properly selected for trading off between the symbol power and the capability of the semantic model.

Fig. 9 illustrates the performance of our proposed on-demand semantic forwarding under different transmission

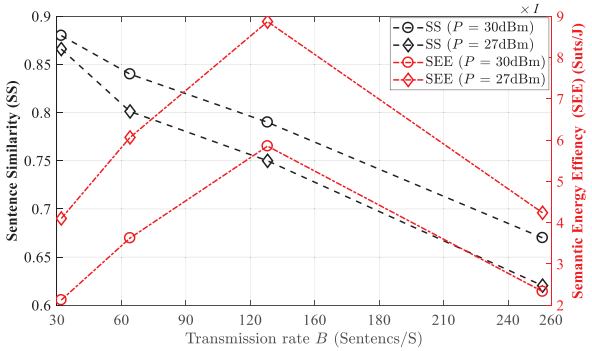


FIGURE 9. The SS and SEE of the cooperative semantic network under different transmit rate B.

rates B, where $B \in \{32, 64, 128, 256\}$, $P \in \{27, 30\}$ and $K = 32$. For a given P, different values of B represent varying numbers of semantic symbols transmitted per unit time. Taking the SS metric as an example, the SS decreases monotonically with an increase in B. This is reasonable in that. As B increase, the energy consumed in transmitting a symbol is decreased, which degrades the capability of semantic recovery. As for the SEE aspect, SEE initially increases and then decreases with an increase in B. This behavior is attributed to our defined semantic utility function as given in (28), which characterizes the trade-off between the transmission rate and the semantic recovery capability of the model. To be specific, when B is very small, the achieved semantic utility is significantly increased as B increases, thus leading to an improved SEE. However, when an unnecessarily large B is adopted, the semantic utility is severely limited by the poor SS, thus in return degrading the SEE. The above observations indicate that the transmit rate B needs to be properly selected for achieving a desirable performance trade-off between SS and SEE.

4) EVALUATION OF THE IMPACT OF CHANNEL CONDITIONS

In order to evaluate the system performance under different channel conditions, the impact of path loss exponent υ_{SD} and distance d_{SD} between nodes is also simulated in Fig. 10. For comparison purposes, the traditional SSCC schemes using Huffman coding and Turbo coding are also simulated. To evaluate the impact of d_{SD}, we let $d_{\mathrm{SD}} \in \{160, 180, 200\}$ and $d_{\mathrm{SR}} = d_{\mathrm{RD}} \in \{120, 140, 160\}$. To evaluate the impact of υ_{SD}, we let $\upsilon_{\mathrm{SD}} \in \{2.6, 3, 3.4\}$ and $\upsilon_{\mathrm{SR}} = \upsilon_{\mathrm{RD}} \in \{2.2, 2.6, 3\}$.

It is observed from Fig. 10 (a) that with a higher path loss exponent υ_{SD}, since the path loss attenuation becomes more severe, the overall SS is degraded accordingly. However, the proposed semantic forwarding performs better in low-to-medium SNR regimes as compared to the traditional communication schemes. Similar phenomena can be observed from Fig. 10 (b), where the overall SS is degraded with a longer distance between S and D, whereas performance gains are achieved by the proposed semantic forwarding in low-to-medium SNR regimes.

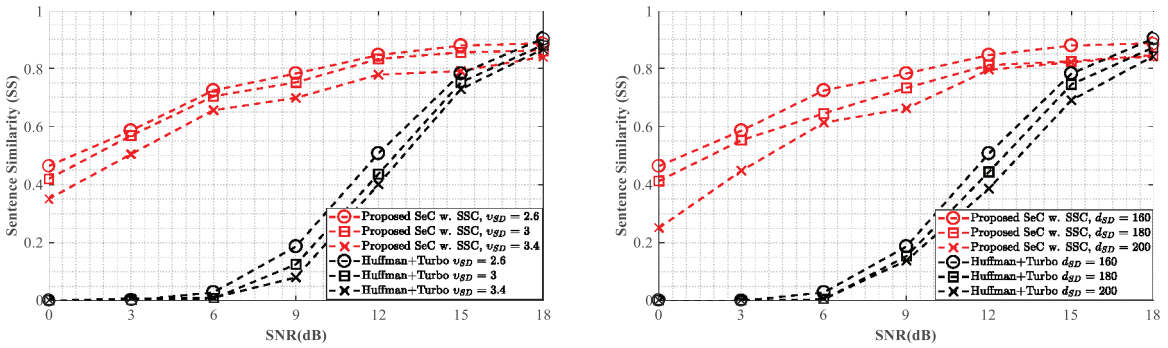
Algorithm 1: Training Algorithm of the Proposed Cooperative Semantic Communication System With *SSC* and *SeC*

Step 1: Train semantic networks on different links.
Initialization: Set the path loss and network parameters for different links.
Input: The transmitted message $\mathbf{m}$.
Output: $f_i(\cdot;\alpha_i)$, $f_{ij}^{-1}(\cdot;\theta_{ij})$ where $i \in \{\mathrm{S}, \mathrm{R}\}$ and $j \in \{\mathrm{R}, \mathrm{D}\}$, and $i \neq j$.
1) Transmitter: $\mathbf{x}_i = f_i(\mathbf{m})$, transmit $\mathbf{x}_i$ over the physical channel as expressed in (4).
2) $\mathbf{y}_{ij} = h\mathbf{x}_i + n$.
3) Receiver: $\hat{\mathbf{m}}_{ij} = f_{ij}^{-1}(\mathbf{y}_{ij})$.
4) Update α_{S}, α_{D}, θ_{SD}, θ_{SR}, θ_{RD} by loss function (21).
Step 2: Train *SeC* network.
Initialization: Load the parameters of the trained network in Step 1.
Input: $\mathbf{y}_{\mathrm{SD}}, \mathbf{y}_{\mathrm{RD}}$ obtained by the trained network.
Output: $\mathrm{Attn}(\cdot;\vartheta_{\mathrm{SeC}})$, $f_{\mathrm{RD}}^{-1}(\cdot;\theta_{\mathrm{SeC}})$.
1) $\mathbf{y_{out}} = \mathbf{Attn}(\mathbf{y_{RD}}, \mathbf{y_{SD}}, \mathbf{y_{SD}}; \vartheta_{\mathbf{SeC}})$.
2) $\hat{\mathbf{m}} = f_{\mathrm{RD}}^{-1}(\mathbf{y}_{\mathrm{out}};\theta_{\mathrm{SeC}})$.
3) Update ϑ_{SeC}, θ_{SeC} by loss function (22).
Step 3: Train semantic detection network in *SSC*.
Initialization: Load the trained parameters of Step 1 and Step 2. Set the order M of the quantization module.
Input: The label obtained from (23) and the output from (16).
Output: $F_{\mathrm{sim}}(\cdot;\gamma_{S_{32}})$.
1) Update $\gamma_{S_{32}}$ by loss function (24).

C. COMPUTATIONAL COMPLEXITY ANALYSIS

Next, we conduct an analysis on the computational complexity of the proposed method, where L is the sentence length, and K is the number of symbols required to represent each word. To be specific, the implementation complexity of Algorithm 1 mainly includes the complexity of the semantic encoding network, the complexity of the *SeC* network and the complexity of the semantic detection network, respectively.

The semantic encoding network primarily employs Transformer networks [20], while the *SeC* network mainly utilizes self-attention mechanisms, both of which have a computational complexity of $O(L^2 \cdot K)$. Additionally, with the prior knowledge on $\mathbf{m}$, the semantic detection network deployed in the proposed scheme primarily consists of multi-layer perception, thus having a computational complexity of $O(L \cdot K)$. Hence, the complexity of the semantic detection network deployed at D can be neglected compared to the complexity of $O(L^2 \cdot K)$. On the other hand, in the absence of the prior knowledge on $\mathbf{m}$, a quantization module is employed to quantize the weights and activation functions of the network. Although this quantization module introduces an additional 32-bit data transfer volume during data transmission, it effectively reduces the size and parameters of the network. Thus, significant performance improvements can be



(a) The effect of path loss exponent υ_{SD} on SS.

(b) The effect of distance d_{SD} on SS.

FIGURE 10. The effect of channel conditions.

achieved by the proposed on-demand semantic forwarding in terms of both SS and SEE, while at the cost of affordable computational complexity.

VI. CONCLUSION

In this work, a multi-point semantic communication system was proposed on a cooperative relay network. To enhance system reliability and adaptability to diverse channel conditions, an on-demand semantic forwarding framework was proposed to avoid unnecessary energy waste. Within this framework, an efficient *SSC* was proposed for accurately detecting the degree of semantic information recovery at the destination, and a novel *SeC* was proposed to jointly abstract and integrate the semantic features from the direct and relay links. In order to characterize the efficiency of semantic communication systems, a new metric of SEE was first proposed, which measured the amount of semantic utility obtained per unit energy consumption. Simulation results validated the performance improvement achieved by the proposed cooperative semantic communication with on-demand semantic forwarding, as compared to the state-of-the-art SSCC schemes and different baseline schemes. Furthermore, an inherent performance trade-off was demonstrated between the semantic fidelity and SEE. This indicated that the key system parameters like the transmit power, the number of symbols used per word, and the transmission rate, need to be properly selected for achieving a balance between different performance metrics. Based on the findings of this paper, further explorations on the semantic communication in multi-antenna systems and the integration of communication and computing will be delegated to our future work.

REFERENCES

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign IL, USA: Univ. Illinois Press, 1949.

[2] T. M. Cover, *Elements of Information Theory*. New York, NY, USA: Wiley, 1999.

[3] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," 2021, *arXiv:2201.01389*.

[4] F. Hekland, "A review of joint source-channel coding," Dept. Electron. Telecommun., Norwegian Univ. Sci. Technol., Trondheim, Norway, Rep., Feb. 2004, p. 12.

[5] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. D. Gündüz, "Deep joint source-channel coding for semantic communications," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 42–48, Nov. 2023.

[6] D. Gündüz et al., "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, Jan. 2023.

[7] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Commun. Mag.*, vol. 59, no. 8, pp. 44–50, Aug. 2021.

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

[9] 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.

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

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

[12] B. Tang, Q. Li, L. Huang, and Y. Yin, "Text semantic communication systems with sentence-level semantic fidelity," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2023, pp. 1–6.

[13] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 230–244, Jan. 2023.

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

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

[16] S. Ma, W. Liang, B. Zhang, and D. Wang, "An investigation on intelligent relay assisted semantic communication networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2023, pp. 1–6.

[17] X. Luo, B. Yin, Z. Chen, B. Xia, and J. Wang, "Autoencoder-based semantic communication systems with relay channels," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2022, pp. 711–716.

[18] C. Bian, Y. Shao, H. Wu, and D. Gunduz, "Deep joint source-channel coding over cooperative relay networks," 2022, *arXiv:2211.06705*.

[19] S. Iyer et al., "A survey on semantic communications for intelligent wireless networks," *Wireless Pers. Commun.*, vol. 129, no. 1, pp. 569–611, 2023.

[20] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.

[21] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[22] Z. Lin, H. Liu, and Y.-J. A. Zhang, "Relay-assisted cooperative federated learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7148–7164, Sep. 2022.

[23] Y. Lu, P. Cheng, Z. Chen, Y. Li, W. H. Mow, and B. Vucetic, "Deep autoencoder learning for relay-assisted cooperative communication systems," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5471–5488, Sep. 2020.
[24] Y. Zhang, W. Xu, H. Gao, and F. Wang, "Multi-user semantic communications for cooperative object identification," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2022, pp. 157–162.
[25] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented multi-user semantic communications for VQA," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 553–557, Mar. 2022.
[26] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1394–1398, Jul. 2022.
[27] Z. Yang, M. Chen, Z. Zhang, and C. Huang, "Energy efficient semantic communication over wireless networks with rate splitting," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1484–1495, May 2023.
[28] L. Xia, Y. Sun, X. Li, G. Feng, and M. A. Imran, "Wireless resource management in intelligent semantic communication networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2022, pp. 1–6.
[29] G. Castagnoli, S. Brauer, and M. Herrmann, "Optimization of cyclic redundancy-check codes with 24 and 32 parity bits," *IEEE Trans. Commun.*, vol. 41, no. 6, pp. 883–892, Jun. 1993.
[30] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 311–318.
[31] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.
[32] K. Eswaran, M. Gastpar, and K. Ramchandran, "Bits through ARQs: Spectrum sharing with a primary packet system," in *Proc. IEEE Int. Symp. Inf. Theory*, 2007, pp. 2171–2175.
[33] Q. Li, S. H. Ting, A. Pandharipande, and M. Motani, "Cooperate-and-access spectrum sharing with ARQ-based primary systems," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2861–2871, Oct. 2012.
[34] G. Yu, Z. Zhang, and P. Qiu, "Cooperative ARQ in wireless networks: Protocols description and performance analysis," in *Proc. IEEE Int. Conf. Commun.*, vol. 8, 2006, pp. 3608–3614.
[35] Z. Ren, G. Wang, Q. Chen, and H. Li, "Modelling and simulation of rayleigh fading, path loss, and shadowing fading for wireless mobile networks," *Simulat. Model. Pract. Theory*, vol. 19, no. 2, pp. 626–637, 2011.
[36] H.-M. Wang, M. Luo, X.-G. Xia, and Q. Yin, "Joint cooperative beamforming and jamming to secure AF relay systems with individual power constraint and no eavesdropper's CSI," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 39–42, Jan. 2013.
[37] H. A. Suraweera, G. K. Karagiannidis, Y. Li, H. K. Garg, A. Nallanathan, and B. Vucetic, "Amplify-and-forward relay transmission with end-to-end antenna selection," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2010, pp. 1–6.
[38] H. A. Ngo and L. Hanzo, "Hybrid automatic-repeat-request systems for cooperative wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 25–45, 1st Quart., 2014.
[39] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
[40] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," 2019, *arXiv:1908.10084*.
[41] Y. Bengio, N. Léonard, and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation," 2013, *arXiv:1308.3432*.
[42] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.
[43] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, Jan. 2021.
[44] M. Z. Win and J. H. Winters, "Analysis of hybrid selection/maximal-ratio combining in rayleigh fading," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, 1999, pp. 6–10.
[45] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[46] K. Han et al., "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.
[47] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "QoE-aware resource allocation for semantic communication networks," in *Proc. IEEE Global Commun. Conf.*, 2022, pp. 3272–3277.
[48] H. Zhang, H. Wang, Y. Li, K. Long, and A. Nallanathan, "DRL-driven dynamic resource allocation for task-oriented semantic communication," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 3992–4004, Jul. 2023.
[49] H. Zhang, H. Wang, Y. Li, K. Long, and V. C. Leung, "Toward intelligent resource allocation on task-oriented semantic communication," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 70–77, Jun. 2023.
[50] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, "Heterogeneous semantic and bit communications: A semi-NOMA scheme," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 155–169, Jan. 2023.
[51] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.
[52] A. M. Al-Samman, T. A. Rahman, M. H. Azmi, N. R. Zulkefly, and A. M. Mataria, "Path loss model for outdoor environment at 17 GHz mm-Wave band," in *Proc. IEEE 12th Int. Colloq. Signal Process. Its Appl. (CSPA)*, 2016, pp. 179–182.
[53] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proc. Mach. Transl. Summit X Papers*, 2005, pp. 79–86.
[54] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.
[55] C. Heegard and S. Wicker, *Turbo Coding*. Boston, MA, USA: Kluwer, 1999.

BING TANG received the B.Eng. degree in electronic and information engineering from the Wuhan Institute of Technology, Wuhan, China, in 2021. He is currently pursuing the master's degree with the School of Electronic Information and Communications, Huazhong University of Science and Technology. His research interests are semantic communication and joint source-channel coding.

LIKUN HUANG received the master's degree in electronic information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2010, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2014. She is currently a Lecturer with the School of Electrical and Information Engineering, Wuhan Institute of Technology, Hubei, China. Her current research interests include semantic communication, person re-identification, gas leakage infrared video detection, video captioning, image/video generation, and lifelong learning.

QIANG LI (Member, IEEE) received the B.Eng. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2007, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2011, where he was a Research Fellow from 2011 to 2013. Since 2013, he has been an Associate Professor with the Huazhong University of Science and Technology, Wuhan, China, where he has been a Full Professor since 2020. He was a Visiting Scholar with the University of Sheffield, Sheffield, U.K., from March 2015 to June 2015. His current research interests include future broadband wireless networks, cooperative communications, simultaneous wireless/lightwave information and power transfer, fog computing, and edge caching.

ASHISH PANDHARIPANDE (Senior Member, IEEE) received the M.S. degrees in electrical and computer engineering, and mathematics, and the Ph.D. degree in electrical and computer engineering from the University of Iowa, Iowa City, in 2000, 2001, and 2002, respectively.

Subsequently, he was a Postdoctoral Researcher with the University of Florida, a Senior Researcher with the Samsung Advanced Institute of Technology, a Senior Scientist with Philips Research, and a lead Research and Development Engineer with Signify. He has held visiting positions with AT&T Laboratories, NJ, USA, and the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India. He is currently the Innovation Director with NXP Semiconductors, Eindhoven, The Netherlands. His research interests are in sensing, networking and controls, data analytics, and their applications in domains like autonomous mobility, smart lighting systems, energy monitoring and control, and cognitive wireless systems. He has around 200 international conference and journal publications and more than 100 patent grants/applications in these fields. He is currently a Senior Editor of IEEE SIGNAL PROCESSING LETTERS, a Topical Area Editor for IEEE SENSORS JOURNAL, and an Associate Editor for IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS.

XIAOHU GE (Senior Member, IEEE) received the Ph.D. degree in communication and information engineering from the Huazhong University of Science and Technology (HUST) in 2003.

He is currently a Full Professor with the School of Electronic Information and Communications, HUST. He is an Adjunct Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. He has been working with HUST since November 2005. Prior to that, he worked as a Researcher with Ajou University, South Korea, and the Politecnico Di Torino, Italy, from January 2004 to October 2005. His research interests are in the area of mobile communications, traffic modeling in wireless networks, green communications, and interference modeling in wireless communications. He has published more than 200 papers in refereed journals and conference proceedings and has been granted about 25 patents in China. He received the Best Paper Awards from IEEE Globecom 2010. He served as the General Chair for the 2015 IEEE International Conference on Green Computing and Communications. He serves as an Associate Editor for IEEE WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE ACCESS.