Autoencoder-based Semantic Communication Systems with Relay Channels

Xinlai Luo, Zhiyong Chen, Bin Xia, and Jiangzhou Wang *Fellow, IEEE*

Abstract—In this letter, we propose a semantic communication scheme for wireless relay channels based on Autoencoder, named AESC, which encodes and decodes sentences from the semantic dimension. The Autoencoder module provides anti-noise performance for the system. Meanwhile, a novel semantic forward (SF) mode is designed for the relay node to forward the semantic information at the semantic level, especially for the scenarios that there is no common knowledge shared between the source and destination nodes. Numerical results show that the AESC achieves better stability performance than the traditional communication schemes, and the proposed SF mode provides a significant performance gain compared to the traditional forward protocols.

Index Terms—Semantic Communication, Autoencoder, Relay, Semantic Forward.

I. INTRODUCTION

SEMANTIC communication was considered on the second level of communication problem [1], aiming to convey the semantic information of the transmission symbols accurately, instead of accurately recovering the transmitted information. With the development of artificial intelligence (AI) technology, semantic communication has attracted one's attention again and is considered as one of the future key mobile technologies.

Recently, several semantic communication concepts have been proposed, i.e., semantic-oriented and goal-oriented communication [2], semantic-aware networking based on federated edge intelligence [3]. Besides reconstructing a new paradigm of semantic communication from a mathematical perspective [4], many innovative communication schemes have been developed based on neural networks (NNs) and semantic interpretation modules to replace conventional communication blocks. The work [5] developed a point-to-point communications system whose entire physical layer processing was performed by NNs, but it is difficult to overcome the difference between the actual channel and the channel model used for training. In [6], a conditional generative adversarial net (GAN) was designed to represent channel effects. Moreover, based on a semantic autoencoder (SAE) [7], the network can learn a projection function from feature space to a semantic embedding space in zero-shot learning (ZSL) models. Based on Transformer [8] which is the dominating language model in natural language processing (NLP),

X. Luo, Z. Chen, and B. Xia are with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China (e-mail: {newcomer, zhiyongchen, bxia}@sjtu.edu.cn). J. Wang is with the School of Engineering, University of Kent, Canterbury CT2 7NT, U.K. (e-mail: j.z.wang@kent.ac.uk).

authors developed a deep learning based semantic communication system, named DeepSC [9]. It has better performance in the low signal-to-noise ratio (SNR) regime compared with the traditional communication system. However, the commonly used cross-entropy (CE) loss was flawed. In [10], reinforcement learning (RL) was used to narrow the semantic distance and an RL-based similarity-targeted semantic communication mechanism was established. Furthermore, it is worth noting that most of the existing semantic communication only considered end-to-end (E2E) communications system, and did not consider cooperative communications, e.g., wireless relay channel. Cooperative communication systems have been widely used in the wireless communication system [11], so it is important to study semantic communication for the cooperative communication system.

In this paper, we try to design a novel Autoencoder-based semantic communication system (AESC) to improve the reliability of semantic communication between multiple nodes. Firstly, we introduce a reliable autoencoder in the semantic communications system, including encoder and decoder powered by NNs, to compress semantic information and resist channel noise interference. Secondly, although the semantic communication is based on the common semantic background knowledge (BK) of the transmitter and receiver, in practice, it is difficult for the source node and the destination node to have exactly the same BK. As is well known, the traditional relay node can help establish the communication between a source and a destination. Thus, we propose a new semantic forward (SF) protocol in the relay node to solve the problem. With the SF protocol, the relay node can obtain semantic information from the source node based on the BK between the relay node and the source node, and then forward the semantic information based on the BK between the relay node and the destination node, thereby helping the source node and the destination node to communicate accurately at the semantic level under different background knowledge.

II. SYSTEM MODEL

Fig. 1 shows an Autoencoder-based semantic communication with one-way relay channels, where the source node transmits the information to the destination node through a relay node based on the semantic communication. The semantic communication consists of two levels: semantic level and transmission level. The semantic level contains the semantic coding layer and the semantic decoding layer for extracting and analyzing the semantic information, respectively. The transmission level guarantees that semantic information can be transmitted accurately in the wireless channel.
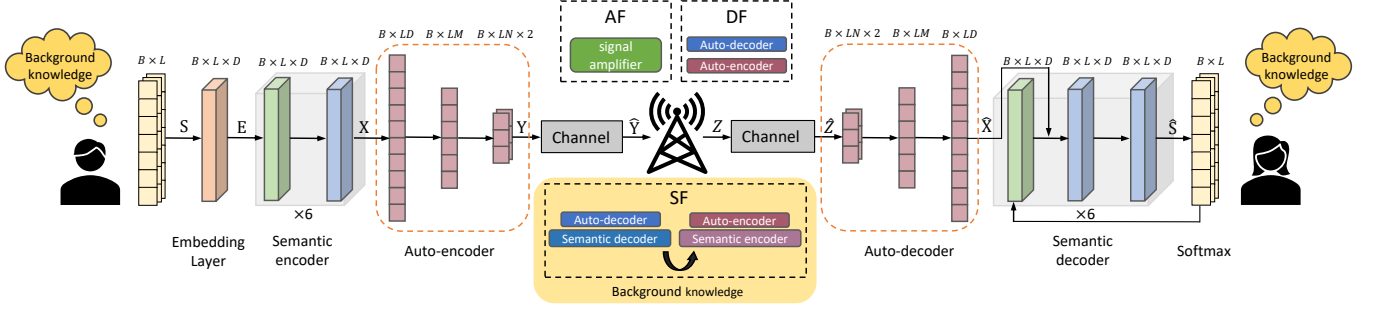
Figure 1: A wireless relay semantic communication model.

A. Autoencoder-based Semantic Communications

1) Encode Layer: The input of the AESC in the source node is a sentence $\mathbf{s} = [w_1, w_2, \cdots, w_L]$, $\mathbf{s} \in \mathcal{K}$, where w_l represents the l-th word in the sentence and $\mathcal{K}$ is the background knowledge. As shown in Fig. 1, the input sentence $\mathbf{s}$ is embedded into a vector $\mathbf{e}$, and each word w in the sentence is mapped to D dimension. Then, we apply semantic encoder which is the encoder of Transformer to encode $\mathbf{e}$ to a semantic vector $\mathbf{x}$. The Transformer encoder consists of two parts, named self-attention sublayer and feed-forward sublayer. Connecting the above two parts alternately six times forms a complete semantic encoder network.

Then the source node passes the semantically encoded vector $\mathbf{x}$ into the Auto-encoder to get compressed vector $\mathbf{y}$. The Auto-encoder has functions similar to traditional channel coding, has anti-noise and encryption capabilities, and can compress the semantic vector to a certain extent.

Accordingly, the receive signal at the relay node is

$$\widehat{\mathbf{y}} = h_1 \mathbf{y} + \mathbf{n}_1, \tag{1}$$

where h_1 represents the Rayleigh fading channel between the source node and the relay node, and $\mathbf{n_1} \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise (AWGN).

Next, the relay node can forward the information to the destination node based on the forward protocol, which will be discussed in the next subsection.

2) Decode Layer: As shown in Fig. 1, the signal received at the destination node is given by

$$\widehat{\mathbf{z}} = h_2 \mathbf{z} + \mathbf{n}_2, \tag{2}$$

where h_2 represents the Rayleigh fading channel between the relay node and the destination node, and $\mathbf{n_2} \sim \mathcal{CN}(0, \sigma_n^2)$ is AWGN.

The destination node gets $\widehat{\mathbf{z}}$ and passes it through the Auto-decoder to recover the semantic vector $\widehat{\mathbf{x}}$ which keeps the semantic information of $\mathbf{x}$. Finally, we input $\widehat{\mathbf{x}}$ into the semantic decoder to recover the transmitted sentences. The semantic decoder is divided into three sublayers, one more layer than the Transformer encoder. Between self-attention sublayer and feed-forward sublayer is an encoder-decoder attention sublayer that can help the decoder focus on the relevant parts of the input sentence. In addition, the semantic decoder has two inputs. One is the semantic vector recovered

by the softmax layer and another is the transmitted sentence according to the framework of Transformer.

B. Relay Forward Protocol

The forward protocol of the relay node can be divided into two scenarios according to whether the relay node has the semantic BK.

1) Traditional relay mode: The relay node only processes the signal at the physical level and does not involve semantics. If the amplify-and-forward (AF) mode is used, the signal received at the relay node is directly amplified and forwarded to the destination node, and we thus have $z = \alpha \widehat{\mathbf{y}}$. Here, α is an amplification constant factor that is chosen to satisfy the total power constraint at the relay node. If the relay node uses the decode-and-forward (DF) mode, the relay node can decode and forward the received information at different layers based on its BK. If the relay node does not have the BK shared by the source node and the destination node, it can not further perform semantic decoding; otherwise, the relay node can perform semantic decoding.

2) Semantic relay mode: The traditional relay mode requires the source node and destination node to have the same BK. However, it is difficult for the source node and the destination node to have exactly the same BK. To overcome this problem, we design a novel relay forward mode here, named semantic forward. The key idea of SF is that the relay node can perform the semantic decoding to obtain the semantic information $\mathbf{s}$ from the received signal $\widehat{\mathbf{y}}$ based on the BK between the source node and itself, and then it can recode the sentences $\mathbf{s}$ in a way that the destination node can understand based on another BK between the source node and the relay node. For example, the source node said:" My son is very good at CS", but the BK of the destination defaults that CS is a game called Counter-Strike and the destination node knows the name of the source node's son is Bob. In this case, the advantage of the proposed SF can be reflected at this time. The relay node can know that CS said by the source node means computer science and the destination node knows Bob. Based on the SF protocol, the destination node receives the sentence:"Bob is very good at computer science". We can see from this example that SF not only reduces semantic distortion but also performs semantic compression.

III. Model Training

Different from Fig. 1, we can train each module in the E2E model. At this time, $\widehat{\mathbf{z}} = \widehat{\mathbf{y}}$, the input of the whole network is a sentence $\mathbf{s}$. The encoded semantic vector can be represented by

$$\mathbf{y} = F_\beta(S_\alpha(\mathbf{s})), \qquad (3)$$

where $S_\alpha(\cdot)$ is the semantic encoder network with the parameter set α and $F_\beta(\cdot)$ is the Auto-encoder network with the parameter set β. The decoded sentence can be represented as

$$\widehat{\mathbf{s}} = G_\eta(D_\theta(\widehat{\mathbf{y}})), \qquad (4)$$

where $D_\theta(\cdot)$ is the Auto-decoder network with the parameter set θ and $G_\eta(\cdot)$ is the semantic decoder network with the parameter set η.

A. Training Auto-encoder/decoder

Different from the joint training of source and channel coding, we train the Autoencoder module separately to deal with different channel conditions more flexibly. The training of Auto-encoder and Auto-decoder does not require a true semantic vector as input. The function of this part is to make the distortion of the transmitted vector as small as possible. Therefore, we use randomly generated vectors for training, and the range of random vectors is $[-2, 2]$, similar to the encoded semantic vector. The Auto-encoder has two hidden layers, each layer compresses the semantic vector to a certain extent. In this paper, the first layer is compressed to M dimensions, and the second layer is compressed to 2N dimensions, mapping into the N symbols of the real part plus the imaginary part. Obviously, the Auto-decoder has a symmetrical structure. Because the input and output are two vectors, we use the mean square error (MSE) as the loss function to reduce the distortion. Based on the transmitted vector $\mathbf{x}$ and the distorted vector $\widehat{\mathbf{x}}$, the loss function can be expressed as

$$\mathcal{L}_{MSE} = \frac{\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2}{N}, \qquad (5)$$

where x_i and $\widehat{x}_i$ represent the components of vector $\mathbf{x}$ and $\widehat{\mathbf{x}}$ respectively, N is vector size. Finally, the ADAM algorithm is exploited to optimize the parameters β, θ. The training algorithm is outlined in Algorithm 1.

B. Training Semantic Encoder/Decoder

The semantic encoder and decoder training process are illustrated in Algorithm 2. Firstly, we load the pre-trained network shown in Algorithm 1, and fetch minibatch B from BK $\mathcal{K}$ to generate sentences $S \in \Re^{B \times L}$, where B is the batch size and L is the length of sentences. Through the embedding layer, the sentences can be represented as a vector $E \in \Re^{B \times L \times D}$, where D is the embedding dimension of each word. After the transmission of the entire communication model, the reconstructed sentence $\widehat{S} \in \Re^{B \times L}$ will be obtained. The cross-entropy is used as the loss function

---

Algorithm 1: Train Auto-encoder/decoder network.

Input: Channel SNR value and a set of randomly generated vectors X;
Output: Network $F_\beta(\cdot)$, $D_\theta(\cdot)$;
1: Take a batch $\mathbf{x}$ from the set X;
2: $F_\beta(\mathbf{x}) \to \mathbf{y}$;
3: Transmit $\mathbf{y}$ over the Rayleigh fading channel: $\widehat{\mathbf{y}} = h\mathbf{y} + \mathbf{n}$;
4: $D_\theta(\widehat{\mathbf{y}}) \to \widehat{\mathbf{x}}$;
5: Compute loss $\mathcal{L}_{MSE}$ by (5);
6: Gradient descent update β, θ;
7: **return** $F_\beta(\cdot)$, $D_\theta(\cdot)$.

Algorithm 2: Train semantic encoder/decoder network.

Input: Channel SNR value and the backgroud knowledge set $\mathcal{K}$;
Output: Network $S_\alpha(\cdot)$, $G_\eta(\cdot)$;
1: Load the pre-trained network $F_\beta(\cdot)$, $D_\theta(\cdot)$;
2: Take a batch $\mathbf{s}$ from the set $\mathcal{K}$ and embed $\mathbf{s}$;
3: $S_\alpha(\mathbf{s}) \to \mathbf{x}$;
4: $F_\beta(\mathbf{x}) \to \mathbf{y}$;
5: Transmit $\mathbf{y}$ over the Rayleigh fading channel: $\widehat{\mathbf{y}} = h\mathbf{y} + \mathbf{n}$;
6: $D_\theta(\widehat{\mathbf{y}}) \to \widehat{\mathbf{x}}$;
7: $G_\eta(\widehat{\mathbf{x}}) \to \widehat{\mathbf{s}}$;
8: Compute loss $\mathcal{L}_{CE}$ by (6);
9: Gradient descent update α, η;
10: **return** $S_\alpha(\cdot)$, $G_\eta(\cdot)$.

---

to measure the difference between S and $\widehat{S}$, which can be formulated as

$$\mathcal{L}_{CE} = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{l} p(x_{ij})\log(q(x_{ij})), \qquad (6)$$

where $p(x_{ij})$ is the real probability that the j-th word in i-th sample, and $q(x_{ij})$ is the predicted probability that the i-th word in i-th sample, and m represents the number of samples in one batch. The CE can measure the difference between two probability distributions. By reducing the loss value of CE and optimizing parameters α, η by the stochastic gradient descent (SGD), the whole network can be robust to distortion.

IV. Performance Evaluation

A. Performance Metrics

For the semantic communication, the symbol has no value unless it has exploitable semantic information. Thus, we use two evaluation indicators to judge communication quality in this letter. One is the bilingual evaluation understudy (BLEU) score [12] and the other one is semantic vector similarity.

1) BLEU Score: BLEU adopts the k-gram precisions rule, which calculates the similarity of k groups of words between the candidate sentence and the reference sentence.

For example, we use "It is a nice day today" as the candidate sentence, and the reference sentence is "Today is a nice day". If we use 1-gram precision, we can see that the candidate has a total of 6 words, and 5 words all hit the reference, then its 1-gram precision is 5/6. Its 3-gram precision is 2/4.

Firstly, we compute the brevity penalty BP as following

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}, \tag{7}$$

where c is the length of the candidate translation and r is the effective reference corpus length. Secondly, we compute the geometric average of the modified k-gram precisions p_k by using positive weights w_k. Then, we have

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{k=1}^{K} w_k \log p_k\right). \tag{8}$$

The BLEU metric ranges from 0 to 1. It will attain a score of 1 if candidate sentences are identical to reference sentences. In this paper, we use the 2-grams BLEU score.

2) Semantic Vector Similarity: In the task of NLP, the cosine similarity formula is often used to calculate the similarity of two-word vectors. Cosine similarity uses the cosine value of the angle between two vectors in the vector space as a measure of the difference between two individuals. The closer the cosine value is to 1, the closer the angle is to 0 degrees, the more similar the two vectors are.

Given two vectors, A and B, the cosine similarity can be given by

$$\text{similarity} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}. \tag{9}$$

Here A_i and B_i represent the components of vector A and B respectively.

B. Simulation Results

We compare the proposed AESC with the traditional source and channel coding approaches over Rayleigh fading channels. We also evaluate the performance of the proposed SF mode when the source node and the destination node have different BK, which is set to different chat histories in the simulation.

Fig. 2 shows the BLEU score vs. SNR for the different schemes with the same number of transmitted symbols. For comparison, we consider a benchmark scheme, i.e., Huffman for source coding, uses Reed-Solomon (RS) (7, 5) for channel coding and binary phase shift keying (BPSK) for the modulation. We also compare with the scheme without Autoencoder. Firstly, it can be seen from the figure that the proposed AESC has a higher BLEU score than that of the scheme without Autoenoder because the Autoencoder can resist the interference of channel noise. The traditional approach performs a little better than the proposed AESC method when the SNR is above 6 dB but when SNR is less than 0 dB, the performance of the traditional approach is far less than the proposed AESC method. Secondly, when the source node and the destination node have the same BK, the
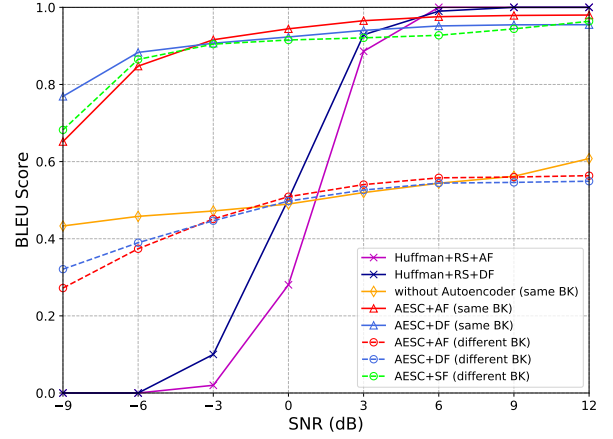

Figure 2: BLEU score versus SNR for the same total number of transmitted symbols in the semantic communication with relay channels for different schemes.

proposed AESC method with the DF mode performs better than that of the proposed AESC method with the AF mode at low SNR. When the SNR is above 0 dB, the proposed AESC method with the AF mode performs better than the proposed AESC method with the DF mode. Thirdly, when the source node and the destination node have different BKs, the proposed SF mode has a higher BLEU score than the AF and DF modes for any SNR. The performance of the SF mode is similar to that of the DF mode when the source node and the destination node have the same BK.

Fig. 3 demonstrates the performances of the proposed AESC with the AF mode and DF mode in the semantic communication with relay channels when the source node and the destination node have the same BK. We fix the SNR from source to relay and change the SNR from relay to destination. The DF mode can maintain relatively good performance in a poor SNR region. When both SNRs are -10 dB, the BLEU score of the DF mode is also close to 0.8, which means that most of the semantics can be transmitted. It can also be seen that the semantic vector at the destination node with the AF mode is closer to the real transmission semantic vector in high SNR. The reason is that the distortion caused by Auto-encoder and Auto-decoder is greater than the impact of the channel in high SNR. However, when the SNR is low, the distortion of semantic vectors caused by DF mode is significantly lower than that of the AF mode.

In Fig. 4, we evaluate the performance of different relay forward protocol via. the placement of the relay node. In the simulation, we normalize the distance between the source node and the destination node to 1, and we use $d \in (0, 1)$ to denote the distance between the source node and the relay node, while $(1 - d)$ denotes the distance between the relay node and the destination node. The results reveal that in the symmetrical situation of the transmit powers $P_1 = P_2 = 5$ dB, the optimal placement of the relay node in each mode is $d = 0.5$. When $P_1 = 5$ dB and $P_2 = 10$ dB, the optimal relay location in each mode is $d \approx 0.4$, which means that the relay should be deployed closer to source node. We can
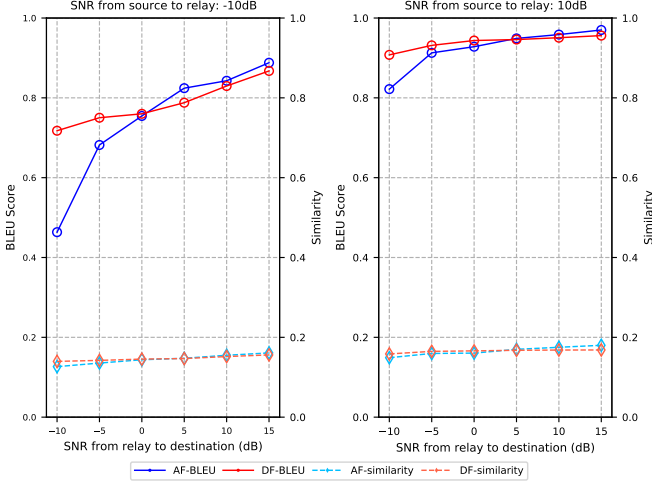
Figure 3: BLEU score and sentence similarity versus SNR from relay to destination for the proposed AESC with the AF mode and the DF mode.
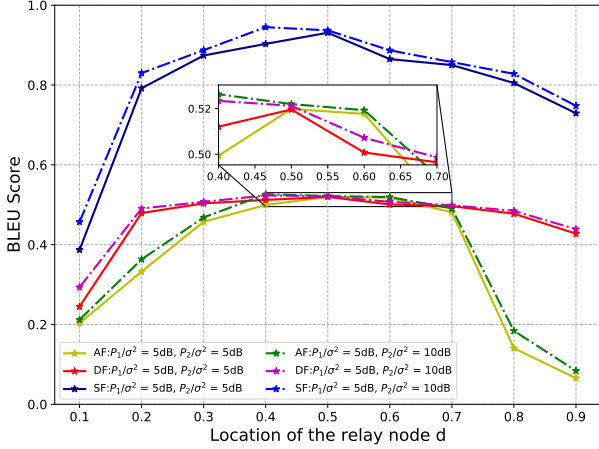

Figure 4: BLEU score versus the relay location d in the AF mode, the DF mode, and the SF mode. Here, we use $\sigma^2 = 1$.

also observe from Fig. 4 that the SF mode and the DF mode can more stably cope with the change of relay node position than the AF mode.

V. CONCLUSION

This letter proposed a semantic communication approach based on Autoencoder for the wireless relay channel with a novel semantic forward protocol. The proposed Autoencoder-based approach uses Autoencoder to extract and compress semantic information and reconstructs its semantic features. For the proposed SF protocol, the relay node can cooperatively use the background knowledge of the source node and the destination node to forward semantic information at the semantic level, to solve the problem of direct semantic communication under different background knowledge between the source node and the destination node. Simulation results demonstrated the effectiveness and advancement of the proposed approach.

REFERENCES

[1] C. E. Shannon, W. Weaver, R. E. Blahut, and B. Hajek, "The mathematical theory of communications." 1949.
[2] E. Calvanese Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.
[3] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *arXiv preprint arXiv:2012.15405*, 2020.
[4] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *2011 IEEE Network Science Workshop*, 2011, pp. 110–117.
[5] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2018.
[6] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional gans as unknown channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3133–3143, 2020.
[7] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
[9] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.
[10] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, and H. Zhang, "Rethinking modern communication from semantic coding to semantic communication," *arXiv preprint arXiv:2110.08496*, 2021.
[11] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, 2004.
[12] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.