Journal Pre-proof

Extended context-based semantic communication system for text transmission

Yueling Liu, Shengteng Jiang, Yichi Zhang, Kuo Cao, Li Zhou, Boon-Chong Seet, Haitao Zhao, Jibo Wei

PII: S2352-8648(22)00198-5

DOI: https://doi.org/10.1016/j.dcan.2022.09.023

Reference: DCAN 521

To appear in: *Digital Communications and Networks*

Received Date: 25 July 2022

Revised Date: 13 September 2022

Accepted Date: 28 September 2022

Please cite this article as: Y. Liu, S. Jiang, Y. Zhang, K. Cao, L. Zhou, B.-C. Seet, H. Zhao, J. Wei, Extended context-based semantic communication system for text transmission, *Digital Communications and Networks* (2022), doi: https://doi.org/10.1016/j.dcan.2022.09.023.

Digital Communications and Networks(DCN)

HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/dcan

Extended Context-Based Semantic Communication System for Text Transmission

Yueling Liu[a], Shengteng Jiang*[a], Yichi Zhang[a], Kuo Cao[a], Li Zhou*[a], Boon-Chong Seet[b], Haitao Zhao[a], Jibo Wei[a]

[a]**College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China**
[b]**Department of Electrical and Electronic Engineering, Auckland University of Technology, Auckland, New Zealand**

Abstract

Context information is significant for semantic extraction and recovery of messages in semantic communication. However, context information is not fully utilized in the existing semantic communication systems since relationships between sentences are often ignored. In this paper, we propose an extended context-based semantic communication (ECSC) system for text transmission, in which context information within and between sentences are explored for semantic representation and recovery. At the encoder, self-attention and segment-level relative attention are used to extract context information within and between sentences, respectively. In addition, a gate mechanism is adopted at the encoder to incorporate the context information from different ranges. At the decoder, Transformer-XL is introduced to obtain more semantic information from the historical communication processes for semantic recovery. Simulation results show the effectiveness of our proposed model in improving the semantic accuracy between transmitted and recovered messages under various channel conditions.

KEYWORDS:
semantic communication, extended context, Transformer-XL

1. Introduction

1.1. Background

The purpose of communication is to exchange useful information. Based on [1], communication is divided into three levels: symbol transmission, semantic exchange and the effect of semantic exchange. Due to the limitations of communication technology in 1950s, the classical communication system proposed by Shannon mainly solves the problem of accurate transmission of symbols [2], while ignoring the semantic aspects of transmission symbols. With the rapid development of artificial intelligence and industrial technology, the realization of semantic communication has gradually become possible, which has attracted much attention.

Semantic communication aims to recover the meaning of messages accurately, which helps to eliminate redundant information and save communication resources. Since artificial intelligence can provide machines with more cognitive and reasoning abilities, the semantic communication assisted with artificial intelligence becomes more promising and has sparked lots of discussion in recent years. The authors in [2] proposed a theory of strongly semantic information. Besides, the semantic communication framework incorporating the world model, reasoning process, and background knowledge was developed in [3]. Moreover, the average semantic error was mathematically

*Corresponding author
[1]E-mail address: liuyueling16@nudt.edu.cn.(Y.Liu); jiangshengteng@nudt.edu.cn(S.Jiang); zhangyichi13@nudt.edu.cn(Y.Zhang); caokuo18@nudt.edu.cn(K.Cao); zhouli2035@nudt.edu.cn(L.Zhou); boon-chong.seet@aut.ac.nz(B.Seet); haitaozhao@nudt.edu.cn(H.Zhao); wjbhw@nudt.edu.cn(J.Wei)

modeled based on semantic similarity [4]. The concept and the scope of semantic communication were further explained and broadened in [5].

1.2. Related works

Semantic communication system design can be generally divided into two categories [6]. One scheme is based on the separate encoder and decoder system [7], which brings the semantic into the traditional communication system, but it does not change the structure of the traditional communication model. The other scheme integrates the semantic layer and the physical layer into a whole neural network, where the source coding and channel coding are jointly optimized [8], [9]. Specifically, the authors of [10] designed a deep learning enabled semantic communication system (DeepSC) based on Transformer, which achieved better performance than the source-channel separation coding system in the low signal-to-noise (SNR) region.

Contrasted to the traditional communication system, semantic communication systems utilize the background knowledge shared by the transmitter and the receiver for semantic compression and understanding of messages [11]. The knowledge graph was adopted as a kind of interpretable knowledge base in [12–14], where the semantic information is represented by triples, and the original text is recovered via inference driven by knowledge graphs. The second form of knowledge base is the training corpus and parameters of neural networks [15–17]. The semantic features in images and speech can be extracted and recovered effectively based on the corresponding neural networks [15, 16]. Moreover, the work of [17] investigated task-oriented semantic communication and has illustrated the effectiveness of semantic communication in improving the performance of intelligent tasks. In [18], not only physical noise but also semantic noise such as written errors or expression ambiguity were considered.

Besides, the context information can be seen as special background knowledge which hidden in the transmitted signal. In [19], integrating context awareness into cognitive radio processes was considered to have promising applications. Furthermore, context information has been widely utilized in semantic extraction and recovery in areas such as texts and images [20–23]. In text, each word, even an abbreviation or a scientific term can be assigned a context-dependent representations based on its context, to capture syntactic and semantic properties of the word under diverse linguistic contexts [24]. To capture feature relations in a sentence and among sentences respectively, the work in [21] described the level, type and representation of context where the level of context can be categorized as global and local context. More context means more accumulation of information to exploit. The rapid development of natural language processing technologies (NLP) helps machines learn and infer from semantically-enriched contexts [25]. In machine translation, longer context help resolve ambiguities and inconsistencies. The work in [22] paid attention to document-level context and achieved better translation results. In addition, Transformer-XL was proposed in [23] to capture textual long-term dependency. This motivates us to study reliable semantic communication using longer context to combat channel distortions.

The Transformer network was widely used and performed well in NLP tasks such as language modeling, machine translation and text classification [26, 27]. For multi-class classification, only encoders of the Transformer are usually used to extract semantic information from text while for language modeling, only decoders are usually used to generate text with semantic information [28, 29]. Both encoders and decoders of the Transformer are used in machine translation to transform the source sentence into the target sentence [30]. However, tasks in NLP do not care about signal transmission through the channel. Thus, one challenge of our semantic communication system is to consider the effects of the channel, which includes reducing the semantic loss caused by the channel and recovering semantics of the text under channel distortions. Besides, one basic target of semantic communication is to improve the reliability of the communication system from the semantic level. Although existing semantic communication schemes can reduce errors in semantic interpretation, further improvement in the reliability of semantic communication systems is still another challenge.

1.3. Novel contributions

In this paper, we proposed a robust semantic communication system based on extended-context information to improve the communication reliability at extremely low SNR. Specifically, we investigate exploiting the joint of NLP technology and context information to enhance semantic extraction and recovery abilities of the communication system. To the best of our knowledge, this paper is the first to explore the contextual correlation between sentences at the transmitter and take historical text into the decoding process in semantic communication. The contributions of this paper are summarized as follows:

- A semantic encoding strategy is proposed to fuse the context information within and between sentences, which helps obtain better semantic encoding representations from the extended context and assists the decoding process.

- A semantic decoding strategy based on Transformer-XL is proposed to additionally adopt the historical communication text as the context, which greatly enhances the reasoning ability of the receiver.

- The simulation results show that the proposed model outperforms the baselines in terms of semantic metrics and significantly improves the reliability of communication systems.

The rest of this paper is organized as follows. Section 2 establishes the system model and explains the performance metrics. Section 3 analyzes the proposed semantic encoding and decoding strategies concretely. Section 4 presents simulation results. Finally, conclusions are drawn in Section 5.

2. System model

2.1. Problem description

The effectiveness of contextual correlation within sentences has been verified in semantic encoding and decoding processes [7], [10]. Nevertheless, more context information can provide more background knowledge, especially in scenes of continuous or topic-fixed communication, or in scenes of transmitting the paragraph text or document text. Therefore, we establish a semantic communication model which consists of the semantic encoder, channel encoder, channel, channel decoder, and semantic decoder successively, as shown in Fig. 1. The transmitter contains a semantic encoder and a channel encoder, where the semantic encoder is used to extract semantic features and the channel encoder is applied for generating symbols to cope with the influence of channel distortion. Unlike DeepSC which employs a single sentence as the input to the transmitter, we take paragraph text as the input. When encoding the current sentence, the previous and following texts of the current sentence in the paragraph serve as the extended context to help learn the semantic representation. Specially, the first sentence and the last sentence in the paragraph will only utilize the following and the previous sentences as the extended context, respectively. The receiver is comprised of a channel decoder for symbol detection and a semantic decoder for text estimation. Since the following texts are unknown at the receiver and only the accumulation of decoded text in the historical communication process can be obtained, we take the "historical text" and the currently received symbols as the input to the receiver. Concretely, the "historical text" is the previously decoded sentence saved by the receiver from the historical communication process. Both the transmitter and the receiver are formulated in neural networks. The shared knowledge between the transmitter and the receiver in our model is learned from the large training corpus and will be stored by the parameters of the neural network. For clarity, the context within a sentence and between sentences are named as local context and global context respectively in the following.

Let V represent the set of all the words in the corpus. We define $\boldsymbol{s} = [w_1, \cdots, w_i, \cdots, w_m]$ as the sentence with m items to be transmitted, where $w_i \in V$ is the i^{th} word in the sentence. At the transmitter, let J_{context} denote the set of sentences in the paragraph except the current sentence $\boldsymbol{s}$, $J_{\text{context}}^{\text{previous}}$ denote sentences located before the current sentence in the paragraph and $J_{\text{context}}^{\text{following}}$ denote sentences after the current sentence in the paragraph, thus $J_{\text{context}} = [J_{\text{context}}^{\text{previous}}, J_{\text{context}}^{\text{following}}]$. Let $S_\beta(\cdot)$ represent the semantic encoder with the parameter set β and $C_\alpha(\cdot)$ stand by the channel encoder with the parameter set α. The sentence $\boldsymbol{s}$ and its contextual sentence set J_{context} will be put into semantic encoder $S_\beta(\cdot)$ together. After channel encoding, the final encoding symbol stream at the transmitter can be expressed by

$$\boldsymbol{x} = C_\alpha\left(S_\beta\left(\boldsymbol{s}, J_{\text{context}}\right)\right). \tag{1}$$

Let $\boldsymbol{y}$ represent the sequence of observations at the receiver, which can be formulated as

$$\boldsymbol{y} = h\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where h denotes the channel coefficient and $\boldsymbol{n} \sim CN\left(0, \sigma_{\text{n}}^2\right)$ represents the additive Gaussian noise.

At the receiver, let $C_\delta{}^{-1}(\cdot)$ represent the channel decoder with the parameter set δ and $S_\chi^{-1}(\cdot)$ represent the semantic decoder with the parameter set χ. On the contrary, the received signal will be decoded by passing through the channel decoder $C_\delta{}^{-1}(\cdot)$ and semantic decoder $S_\chi^{-1}(\cdot)$ successively. We define $\hat{\boldsymbol{s}}_{\text{previous}}$ as the previous decoded sentence. After channel decoding, the current sentence and the previous decoded sentence will be put into semantic decoder $S_\chi^{-1}(\cdot)$ together. Therefore, the recovered sentence can be represented as follows:

$$\hat{\boldsymbol{s}} = S_\chi^{-1}\left(C_\delta^{-1}(\boldsymbol{y}), \hat{\boldsymbol{s}}_{\text{previous}}\right). \tag{3}$$

To minimize the difference between $\boldsymbol{s}$ and $\hat{\boldsymbol{s}}$, we adopt cross-entropy (CE) as the loss function, denoted as follows:

$$\begin{aligned}\mathcal{L}_{\text{CE}}(\boldsymbol{s}, \hat{\boldsymbol{s}}; \beta, \alpha, \delta, \chi) = -\sum_{l=1} q\left(w_{\text{i}}\right)\log\left(p\left(w_i\right)\right) + \\ \left(1 - q\left(w_i\right)\right)\log\left(1 - p\left(w_i\right)\right),\end{aligned} \tag{4}$$

where $q\left(w_i\right)$ is the real probability that the word w_{i} appears in transmitted sentence $\boldsymbol{s}$, and $p\left(w_i\right)$ is the predicted probability that w_{i} appears in estimated sentence $\hat{\boldsymbol{s}}$. By minimizing the loss function, the whole system will learn to make full use of context information accurately while ensuring the semantic accuracy of text transmission.

2.2. Performance metrics

Traditional performance metrics such as bit-error rate are no longer suitable for semantic communication. Here we adopt Bilingual evaluation understudy (BLEU) score [31], metric for evaluation of translation with explicit ordering (METEOR) score [32] and

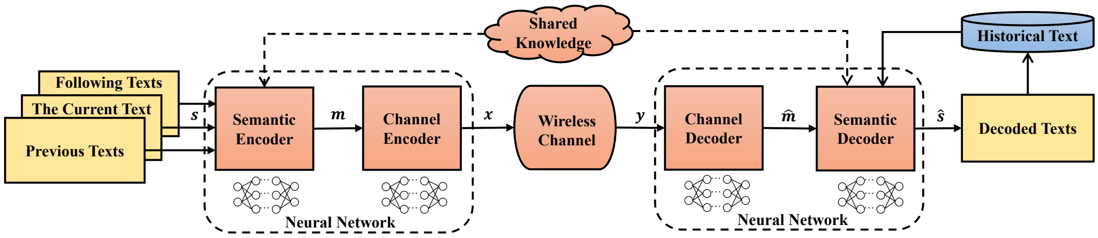


Fig. 1. The proposed semantic communication system model.

similarity score based on BERT [28] to calculate semantic similarity between transmitted and recovered messages, which have been widely used in semantic communication systems [7], [10].

The BLEU evaluates the similarity by counting the number of the same n-grams between transmitted and received texts, where n-gram means n consecutive words in the text. The BLEU-N gram score is obtained as follows:

$$\mathrm{BLEU} = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right), \tag{5}$$

where BP denotes the penalty factor, w_n is the weight of n-gram, and p_n is the n-gram precision score.

METEOR evaluates the performance of sentences based on the harmonic average of precision and recall. The METEOR score can be expressed as

$$F_{\mathrm{mean}} = \frac{PR}{\partial P + (1-\partial)R}, \tag{6}$$

$$METEOR = (1 - Pen)F_{\mathrm{mean}}, \tag{7}$$

where P represents precision, R denotes recall, ∂ is the hyperparameter according to WordNet, F_{mean} is the harmonic mean combining the precision and recall, and Pen is the penalty coefficient.

The similarity score based on BERT evaluates the semantic similarity between sentences by comparing the degree of similarity between vectors. The calculation formula is given as follows:

$$\mathrm{sim}_{\mathrm{v_{Bert}}}(s_1, s_2) = \frac{\mathbf{v}_{\mathrm{BERT}}(s_1)\,\mathbf{v}_{\mathrm{BERT}}(s_2)^T}{\|\mathbf{v}_{\mathrm{BERT}}(s_1)\|\|\mathbf{v}_{\mathrm{BERT}}(s_2)\|}, \tag{8}$$

where $\mathbf{v}_{\mathrm{BERT}}(s_1)$ and $\mathbf{v}_{\mathrm{BERT}}(s_2)$ are word vectors of sentence s_1 and s_2 respectively.

All metrics introduced above have a value range between 0-1. A value of 1 means the highest semantic score, while a value of 0 denotes that the recovered text and the transmitted text have no semantic similarity.

3. Methodology

3.1. *Overall framework*

The detailed architecture of the proposed ECSC is shown in Fig.2. The input contains a minibatch of the current sentences $\boldsymbol{S}$ and the set of contextual sentences $\boldsymbol{J}_{\mathrm{context}}$. Our semantic encoder consists of an embedding module, several local encoding modules, a concatenation operation, a global encoding module and a gated linear layer. The function of each module is shown as follows:

- The embedding module not only utilizes word embedding and position embedding as the typical Transformer does, but also utilizes the segment embedding.
- The local encoding module is the typical Transformer encoder [26] which is shown in Fig.3 (a) to extract context information within a sentence.
- The concatenation operation is utilized to concatenate the hidden states of all sentences in the paragraph text.
- The global encoding module is added on the local encoding modules to extract inter-sentential context information. The detailed structure of the global encoding module is shown in Fig.3 (b).
- The gated linear layer is utilized to fuse the context information extracted by local encoding modules and the global encoding module.

Besides, our channel encoder and channel decoder both contain several linear layers. In the proposed model, the channel is an untrainable layer. Our semantic decoder consists of several extended context decoding modules, a linear layer and a softmax layer. The function of each module is shown as follows:

- The extended context decoding module utilizes the Transformer-XL layer and the Cross-Attention layer to fuse the information from the historical communication text and the current received sentence. The details of extended decoding module are drawn in Fig.3 (c).
- The linear layer transforms the dimension size of the semantic feature vector to the dimension size of the word dictionary. The softmax layer outputs the probability distribution over the word dictionary.

The training process of the whole system is given in Algorithm 1.

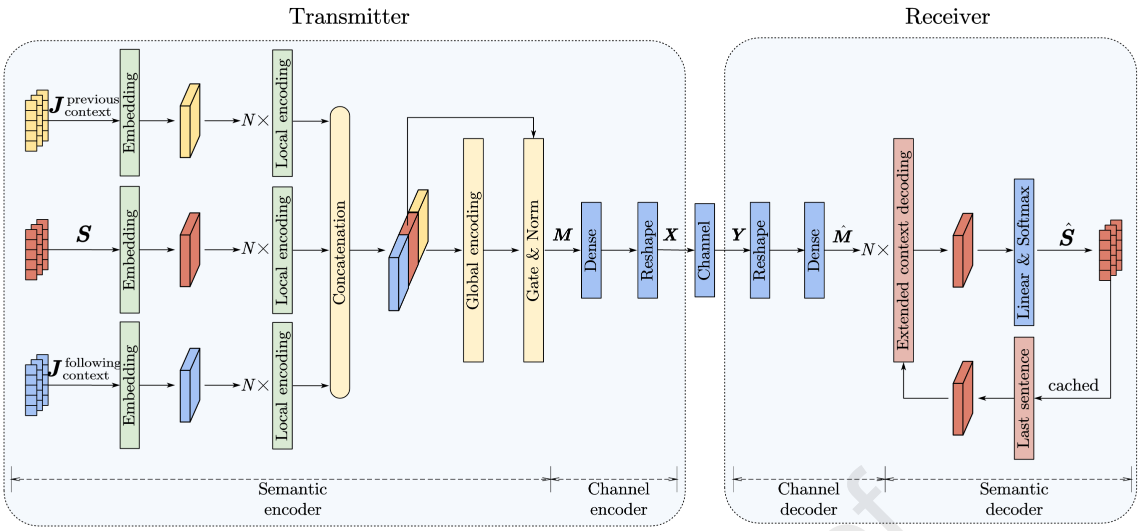
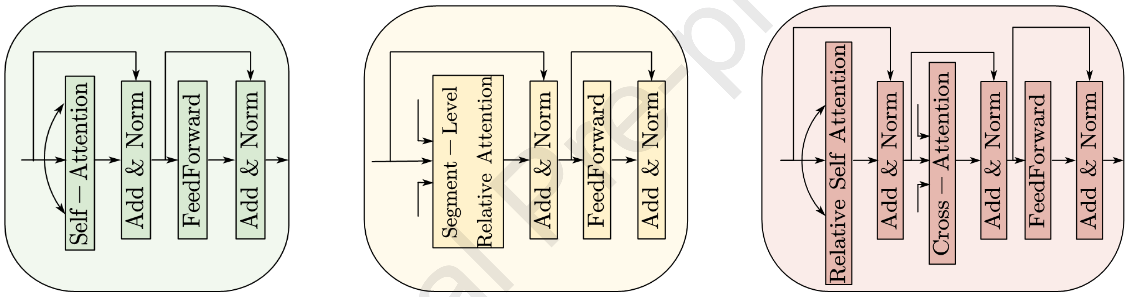


Fig. 2. The overall neural network architecture of our proposed ECSC, which consists of a semantic encoder, a channel encoder, channel, a channel decoder and a semantic decoder. Here, $N\times$ denotes N modules are stacked.



(a) Local encoding (b) Global encoding (c) Extended context decoding

Fig. 3. The detailed structure of modules that are utilized in the proposed ECSC. (a) The local encoding module for extracting context information within a sentence. (b) The global encoding module for extracting inter-sentential context information. (c) The extended context decoding module for utilizing historical communication process and recovering sentences.

Algorithm 1 Training process of the whole semantic communication system

Input: The transmitted minibatch of sentences $\boldsymbol{S}$ and the minibatch of the contextual sentences $\boldsymbol{J}_{\text{context}}$.

1: Semantic encoder: $S_\beta(\boldsymbol{S}, \boldsymbol{J}_{\text{context}}) \rightarrow \boldsymbol{M}$.
2: Channel encoder: $C_\alpha(\boldsymbol{M}) \rightarrow \boldsymbol{X}$.
3: Transmit $\boldsymbol{M}$ over the channel.
4: Channel decoder: $C_\delta^{-1}(\boldsymbol{Y}) \rightarrow \hat{\boldsymbol{M}}$.
5: Semantic decoder: $S_\chi^{-1}\left(\hat{\boldsymbol{M}}, \hat{\boldsymbol{S}}_{\text{previous}}\right) \rightarrow \hat{\boldsymbol{S}}$.
6: Update and cache the minibatch of previous decoded sentences $\hat{\boldsymbol{S}}_{\text{previous}}$.
7: Compute loss function $\mathcal{L}_{\text{CE}}$.
8: Train $\alpha, \beta, \delta, \chi$ by Gradient descent.
9: **return** The whole network $S_\beta, C_\alpha, C_\delta^{-1}, S_\chi^{-1}$.

3.2. Semantic encoding strategy

In this section, details of the neural network modules at the semantic encoder are introduced. To make use of inter-sentential context, in addition to the typical Transformer encoder, our semantic encoder also adopts segment embedding, segment-level relative attention, and a gated context fusion mechanism.

3.2.1. Segment embedding

When converting the input tokens into vectors, word position embedding is used in a typical Transformer encoder to represent the order of words in a sentence [26]. Since the input in our model contains several sentences, segment embedding is used to represent the order of sentences. Let $w_{k,i}$ be the i^{th} word in k^{th} sentence and $x_{k,i}$ be the vector representation of $w_{k,i}$. Therefore, $x_{k,i}$ can be denoted as follows: [22]

$$x_{k,i} = E[w_{k,i}] + E_k^s + E_i^s, \tag{9}$$

where $E[w_{k,i}]$ represents the word embedding of $w_{k,i}$, E_k^s is the k^{th} sentence embedding, and E_i^s is the i^{th} word position embedding.

3.2.2. Segment-level relative attention

In a typical Transformer encoder, self-attention helps to extract features within a sentence. Encoding representation of $w_{k,i}$ under self-attention can be denoted as [26]

$$\begin{aligned} h_{k,i} &= \mathrm{SelfAttn}(q,k,v) \\ &= \mathrm{SelfAttn}\left(x_{k,i}, x_{k,j}, x_{k,j}\right) \\ &= \sum_{j=1}^{m} \mathrm{softmax}\left(\frac{\left(W^{Q} x_{k,i}\right)\left(W^{K} x_{k,j}\right)^{T}}{\sqrt{d_h}}\right) W^{V} x_{k,j}, \end{aligned} \tag{10}$$

where q, k, v are queries, keys and values, respectively, $x_{k,j}$ is the vector representation of the j^{th} word in the k^{th} sentence, d_h is the dimension size of an attention head, W^Q, W^K, W^V are linear projection matrices for the queries, keys and values, respectively.

Let the parameter vector $\gamma^{*}_{k-\kappa}$ denote the relative distance between the k^{th} sentence and the κ^{th} sentence. To extend the attention from token level to sentence level, $\gamma^{*}_{k-\kappa}$ is used in key and value, providing intersentential clues. Assuming that the contextual sentence set J_{context} contains N sentences and the κ^{th} sentence contains m_κ words, encoding representation of $w_{k,i}$ under segment-level relative attention can be described as [33]

$$\begin{aligned} h_{k,i} &= \mathrm{SegAttn}(q,k,v) \\ &= \mathrm{SegAttn}\left(x_{k,j}, x_{\kappa,j}, x_{\kappa,j}\right) \\ &= \sum_{\kappa=0}^{N} \sum_{j=1}^{m_s} \mathrm{softmax}\left(\frac{\left(W^{Q} x_{k,j}\right)\left(W^{K} x_{\kappa,j} + \gamma^{K}_{k-\kappa}\right)^{T}}{\sqrt{d_h}}\right) \times \\ &\left(W^{V} x_{\kappa,j} + \gamma^{V}_{k-\kappa}\right), \end{aligned} \tag{11}$$

where $x_{\kappa,j}$ is the vector representation of the j^{th} word in the κ^{th} sentence. Through segment-level relative attention, sentences in the contextual sentence set could participate in encoding of the current sentence.

3.2.3. Gated context fusion

After extracting both the local and global context information, the final semantic representation h_{final} is decided by the weights g assigned by the gated mechanism, shown as follows: [22]

$$h_{\text{final}} = (1-g) \odot h_k^{L} + g \odot h_k^{G}, \tag{12}$$

where h_k^L is the local encoding representation, h_k^{G} is the global encoding representation, and $\odot$ indicates element-wise product. Concretely, g is obtained by a linear layer, shown as follows:

$$g = \sigma\left(W_g \left[{h_k}^{L} \oplus {h_k}^{G}\right]\right), \tag{13}$$

where W_g is a learnable linear transformation, $\oplus$ denotes the concatenation operation, and $\sigma(\cdot)$ is the sigmoid activation which leads the fusion gate value to be between 0 to 1.

3.3. Semantic decoding strategy

In this section, details of the neural network modules at the semantic decoder are introduced. our semantic decoder combines Transformer-XL and cross-attention to utilize information from the last received sentence and the current received sentence [23].

Typical Transformer decoder is able to view the previous predicted word within the sentence, while Transformer-XL decoder can view the previous predicted sentence. Transformer-XL utilizes a segment-level recurrence mechanism and a relative positional encoding scheme to enable the semantic decoder to capture longer-range dependency. Concretely, hidden states of the previous sentence will be cached and reused for the current sentence, shown as follows:

$$\tilde{h}_k = \left[\mathrm{SG}\left(\hat{h}_{k-1}\right) \oplus \hat{h}_k\right], \tag{14}$$

where $\hat{h}_k$ represents the hidden state of current sentence, $\hat{h}_{k-1}$ denotes the hidden state of last sentence, and $\mathrm{SG}(\cdot)$ is the operation of the stop gradient. Relative self-attention is a variant of self-attention. We apply relative self-attention to extract and propagate previous information, which is denoted as follows [23]:

$$\begin{aligned} \bar{h}_k &= \mathrm{RelAttn}\left(q,k,v\right) \\ &= \mathrm{RelAttn}\left(\hat{h}_k, \tilde{h}_k, \tilde{h}_k\right). \end{aligned} \tag{15}$$

After extracting context information from previous sentence, we add a cross-attention layer to the Transformer-XL decoder layer to utilize the information of the current received sentence, and the final representation is shown as follows [26]:

$$\hat{h}_{\text{final}} = \mathrm{SelfAttn}(\bar{h}_k, \hat{h}_k, \hat{h}_k). \tag{16}$$

4. Simulation results

In this section, simulation results are provided to verify the effectiveness of the proposed model.

4.1. Simulation settings

The proceedings of the European Parliament [34] is a standard corpus freely available on the Internet and has been widely used in NLP. Reference [10] utilized the English version of the proceedings of the European Parliament and split it into training and test datasets for the semantic communication system. Therefore, the English version of the proceedings of the European Parliament was also adopted as the dataset in this paper for better comparison. In [10], sentences longer than 30 words are removed, and thus the context becomes incoherent. To obtain a coherent context, sentences longer than 30 words are retained in this paper, which raises higher requirements for interpreting long sentences in communication systems.

In the experiment, we set three typical Transformer encoders, one global Transformer encoder, and three

Table 1
Parameters of our proposed model.

	Layer Name	Units	Activation
Transmitter (Encoder)	3×Typical Transformer Encoder	128(8heads)	Linear
	1×Global Transformer Encoder	128(8heads)	Linear
	Dense	128	Sigmod
	Dense	256	ReLU
	Dense	16	ReLU
Channel	AWGN/Rayleigh	None	None
Receiver (Decoder)	Dense	256	ReLU
	Dense	128	ReLU
	3×Extended Context Transformer decoders	128(8heads)	Linear
	Prediction Layer	Dictionary Size	Softmax

extended context Transformer decoders with 8 heads. The channel encoder and decoder are set as dense with 16 units and 128 units, respectively. A contextual sentence set $\boldsymbol{J}_{\text{context}}$ is configured to contain 10 sentences. More concrete settings can be found in Table 1. This paper compares ECSC with the existing semantic communication algorithms and the traditional scheme under additive white Gaussian noise (AWGN) channels and Rayleigh fading channels. In the simulation, the perfect channel state information (CSI) is assumed to be obtained for all schemes. For brevity, we denote DeepSC scheme with three typical Transformer layers and five typical Transformer layers as “DeepSC*3” and “DeepSC*5”, respectively. For traditional methods, we adopt Huffman codes for source coding, the convolutional codes for channel coding, and 8PSK for modulation as the baseline (Baselines: Huffman + CC).

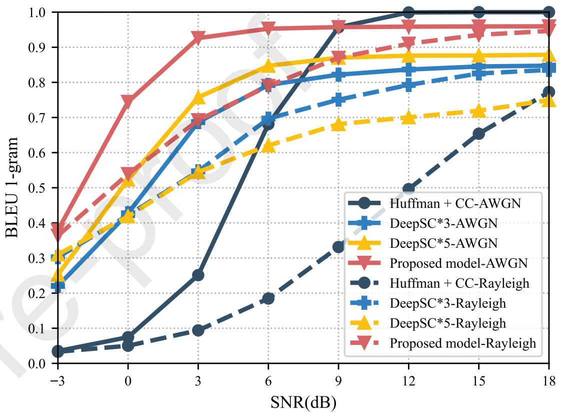


Fig. 4. BLEU 1-gram score versus SNR over the AWGN channel and the Rayleigh fading channel.

4.2. Performance comparisons

Fig.4 shows the BLEU-1gram score versus SNR over the AWGN channel and the Rayleigh fading channel. It can be seen that our proposed model outperforms all other schemes over the Rayleigh fading channel across the entire SNR region. Due to severe impacts of Rayleigh fading, the score of the "Huffman + CC" scheme improves slowly with the increase of the SNR and do not reach 0.8 when the SNR is 18dB. Meanwhile, both “DeepSC*3” and “DeepSC*5” schemes are more competitive and achieve higher scores than the "Huffman + CC" scheme especially in low SNR region over Rayleigh channels. Besides, the proposed ECSC achieves higher BLEU scores than DeepSC*3 and DeepSC*5 in both AWGN and Rayleigh channel, which indicates the effectiveness of the inter-sentential context in improving the reliability of the communication systems. In AWGN channels, the BLEU 1-gram scores of our proposed model increase rapidly and reach above 0.9 when the SNR is above 3dB. To obtain the same score, the "Huffman + CC" scheme requires an additional 6dB power consumption. In high SNR region, the "Huffman + CC" scheme achieves a perfect score of 1 when SNR is 12dB due to the protection of channel coding, but the score of our model is slightly lower than the traditional method. This is because deep-learning based methods not only learn the semantic representation of sentences but also compress the dimensions of semantic representation. In AWGN channels, DeepSC*5 performs better than DeepSC*3 but still worse than ECSC, which demonstrates that the extended context-based mechanisms introduced by our proposed model is more effective than the original structure of DeepSC assisted with attention layers.

Fig.5 draws the relationship between the METEOR score and the SNR over the AWGN channel and the Rayleigh fading channel. Compared with BLEU-1 gram score, METEOR takes synonymy, stemming, and paraphrasing into consideration and utilizes external knowledge sources, thus paying more attention to sentence fluency. As observed, the METEOR scores of the ECSC outperform all other schemes in all SNR region over Rayleigh channels. Besides, the METEOR scores of ECSC perform similar to their BLEU 1-gram scores in both channel conditions. Nevertheless, the METEOR scores of the "Huffman + CC" scheme are different. The traditional method approaches 1 at 9dB in terms of BLEU 1-gram over the AWGN channels while only achieving 0.6 in terms of METEOR, showing less capability in maintaining sentence fluency. Moreover, it can be observed

Table 2
Representative decoded sentences and numbers of wrong words by different methods over AWGN channels when SNR is 3dB.

Different methods	Representative decoded sentences	Number of wrong words
Transmitted sentence	It has meant tighter constraints and ceilings for budgetary spending and for that reason an adaptation has been necessary.	0
ECSC	It has meant tighter constraints and ceilings for budgetary spending and for that reason an adaptation has been necessary.	0
DeepSC*5	It has meant developed skills and reverse for budgetary spending and for that reason an guilty has been necessary.	4
DeepSC*3	It has meant to counter and appeared for humanitarian review and social that reason an assets has been necessary.	7
Huffman + CC	Iemmpi thereby fr sdil egwiconvosci mhn elnd eloginey txpepirinenm dor onrcil ifncorpunity funatiittob taranteu.	19

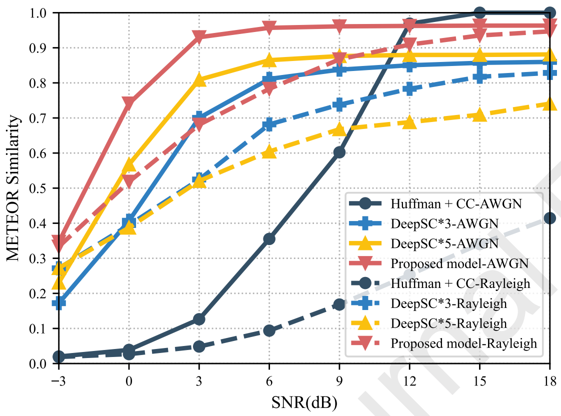


Fig. 5. METEOR score versus SNR over the AWGN channel and the Rayleigh fading channel.

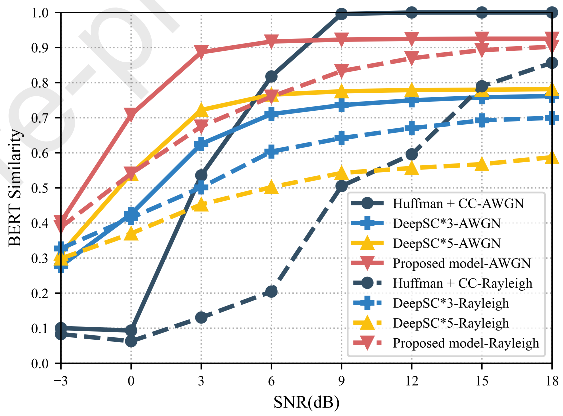


Fig. 6. Similarity score based on BERT versus SNR over the AWGN channel and the Rayleigh fading channel.

that DeepSC*5 does not have performance gain when compared with DeepSC*3 in Rayleigh channels. This means that increasing attention layers on the original structure of DeepSC in Rayleigh channels may be harder for the neural network to cope with the channel distortions.

Fig.6 plots the relationship between the similarity score based on BERT and the SNR over the AWGN channel and the Rayleigh fading channel. The BERT model, pre-trained by billions of sentences, can produce different semantic representations in different contexts and has been applied to a wide range of NLP problems. Thus, the similarity score based on BERT is closer to human evaluation. The trends of the similarity score based on BERT for all methods are similar to the trends of BLEU 1-gram scores and METEOR scores. However, compared with BLEU 1-gram scores and METEOR scores, the similarity scores based on BERT of DeepSC*3 and DeepSC*5 both decrease by 10% in high SNR regions, denoting deficiencies of DeepSC*3 and DeepSC*5 in recovering the meaning of long sentences. Contrasted to the Fig.4 and Fig.5, the score gap between ECSC and other schemes is further enlarged in Fig.6, which indicates that the proposed ECSC has more advantages in extracting and recovering the semantic features of the text.

Fig.7 plots BLEU 4-gram score versus SNR over the AWGN channel and the Rayleigh fading channel. To verify the effectiveness of our semantic decoding strategy, we add "DeepSC*3 + extended context decoding" scheme for comparison, which introduces the proposed extended context semantic decoding strategy into the DeepSC*3 system. Additionally, it is worth noting that the "DeepSC*3 + extended context decoding" scheme can be used in the situation of transmitting sentence by sentence. In Fig.7, the "DeepSC*3 + extended context decoding" scheme outperforms the "DeepSC*3" scheme in both AWGN and Rayleigh channel, which reflects that historical communication text is helpful to infer and interpret the current sentence. Besides, the performance of the ECSC model is better than that of the "DeepSC*3 + extended context

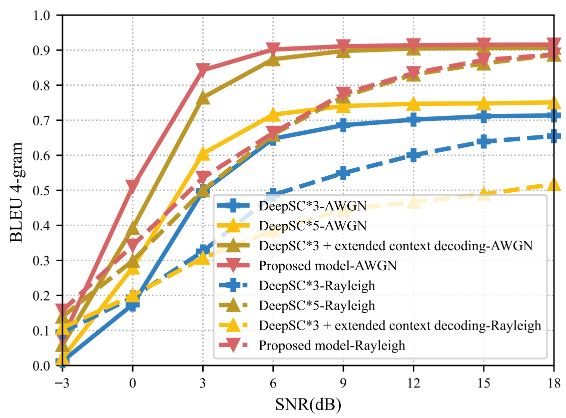


Fig. 7. BLEU 4-gram score versus SNR over the AWGN channel and the Rayleigh fading channel.

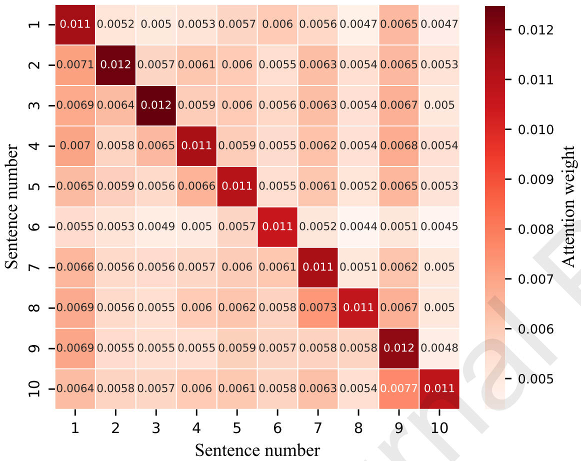


Fig. 8. An example of sentence-to-sentence average attention weights of context-related sentences at the global encoding module.

decoding" scheme, especially under low SNR, which illustrates that it is optimal to use context information at both semantic encoder and semantic decoder.

To show the decoding performance of the proposed ECSC and baselines more intuitively, representative decoded sentences are shown in Table 2. In Table 2, the representative decoded sentence contains 19 words. It can be observed that the proposed ECSC interprets the sample sentence without errors while the number of wrong words in DeepSC*5 and DeepSC*3 are 4 and 7, respectively.

To further explore whether the global context information is utilized by the proposed model, visualizations of attention scores are presented in Fig.8-9.

Fig.8 describes an example of sentence-to-sentence average attention weights of a set of context-related sentences at the global encoding layer. Table 3 shows the set of context-related sentences corresponding to Fig.8. Sentence-to-sentence average attention weights

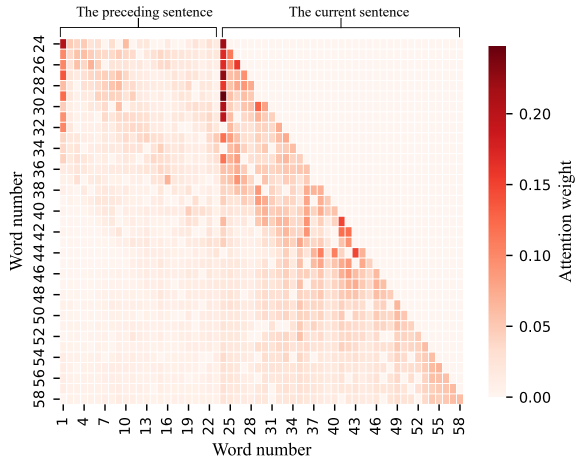


Fig. 9. An example of word-to-word attention weights of the current sentence and its preceding sentence at the extended context decoding module under one attention head.

at the global encoding module are calculated by

$$\alpha_k^{\kappa} = \frac{1}{mm_{\kappa}} \sum_{i=0}^{m} \sum_{j=1}^{m_{\kappa}} softmax\left(\frac{(W^Q x_{k,i})(W^K x_{\kappa,j} + \gamma_{k-\kappa}^K)^T}{\sqrt{d_h}}\right). \tag{17}$$

where α_k^{κ} denotes the average attention weight that the word in the k^{th} sentence gives to the word in the κ^{th} sentence. In Fig.8, each row represents the average attention weights given by the current sentence to other sentences. Each column represents the average attention weights given by other sentences to the current sentence. From Fig.8, we can see that the current sentence plays an important role in itself because the weights on the diagonal are generally high. Furthermore, a few sentences are generally given higher weights by other 9 sentences, such as the 1st and 9th sentence in Fig.8. From Table 3, we can see that this paragraph is about a question and answer about a certain disease, where the 1st sentence is the key sentence about questions and the 9th sentence is a critical response to questions. This may illustrate that sentences given higher weights could contain keywords and be important to the sentence set. Besides, in Fig.8, the previous sentence tends to be given higher weights than the sentence far away from the current sentence, which denotes more contextual relations in surrounding sentences.

Fig.9 draws an example about word-to-word attention weights of the current sentence and its preceding sentence at the extended context decoding module under one attention head. In the Fig.9, each row represents the attention weights given by the current word to other words, which similarly sums to 1. Each column represents the attention weights given by other words to the current word. In Fig.9, the current sentence contains 35 words (24th -58th word) while the preceding sentence contains 23 words (1st -23rd word). The weights in the top right corner of Fig.9 are all 0 because in the decoding process, the next word

Table 3
The set of context-related sentences corresponding to Fig.8.

Sentence number	Sentence
1	In the interests of transparency and information to members, we need to know the extent of the problem of legionella.
2	Could you provide us with information on that?
3	What procedures have been put in place?
4	Has the air conditioning system now been switched off?
5	Can we therefore assume that an operation has been undertaken to disinfect the system?
6	Could you confirm this and, indeed, could parliament's presidency confirm that there is no problem with asbestosis which, of course, would be in breach of our own EU directive passed over a decade ago.
7	I personally am not in a position to answer any of your questions.
8	I have no information on this subject.
9	You will have to refer the matter to the president and to the quaestors.
10	I will certainly draw attention to your remarks.

is not visible and not predictable when decoding the current word. It can be found that words of the preceding sentence participate in the decoding process of the current sentence. Particularly, the attention mechanism may lead to focus on some particular position with special interests. For example, the 1st word and the 24th word, which are the first words of the preceding sentence and the current sentence respectively, are both given higher weights by the 24th -31st word, illustrating their importance in decoding. Additionally, the attention weights initially focus on the preceding sentence, but start to shift to the current sentence with the decoding process moving on.

The computational complexities of the ECSC, DeepSC*3 and DeepSC*5 are compared in the Table 4. The number of parameters to be trained in the neural network (Params) is used to represent the space complexity. Both the number of floating-point operations in the neural network (FLOPs) and the average processing runtime for each sentence (Runtime) are used to represent the time complexity. Concretely, Params and FLOPs are obtained under the help of package “NNI” [35]. Runtime is obtained by transmitting 5380 sentences from the transmitter to the receiver under different methods. Besides, all simulations were conducted by a computer with Inter Xeon Silver 4110 CPU@2.1 GHz and NVIDIA GeForce GTX 2080TI. Compared with DeepSC*3, the proposed ECSC improves the performance by increasing the computational complexity, since Params, FLOPs and Runtime of the proposed ECSC are all higher than DeepSC*3. Furthermore, it can be seen from Table 4 that the complexity of the proposed ECSC is lower than DeepSC*5.

Table 4
Comparisons of the computational complexity for different methods.

Different methods	Space complexity	Time complexity	
	Params	FLOPs	Runtime
ECSC	5254672	5246976	47.76ms
DeepSC*5	5927808	5885952	50.01ms
DeepSC*3	5004672	4968448	38.08ms

5. Conclusion

In this paper, an extended context-based semantic communication system for text transmission was proposed to improve the reliability of the communication system. To fully explore the background knowledge for improving the reliability, our proposed model utilizes the context information within and between sentences to extract and recover semantic features. Specifically, local context information and global context information were fused in our semantic encoding strategy to obtain better semantic representation. Moreover, Historical communication text was utilized in our semantic decoding strategy to help refer to and decode the current received sentence. Performance comparisons and visualizations demonstrate that the proposed ECSC improved the performance in terms of BLEU score, METEOR score and the similarity score based on BERT, especially in the low SNR regime.

However, it is worth noting that the pretrained semantic communication model may not be adaptive to changes of channel propagation environments, and meanwhile, the retraining process of the network is time-consuming. Based on the above consideration, we will consider to utilize transfer learning in the future work, to make the proposed model applicable to various communication scenarios.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant No.61931020, U19B2024, 62171449, and 62001483 and in part by the science and technology innovation Program of Hunan Province under Grant No.2021JJ40690.

References

[1] W. Weaver, C. Shannon, Recent contributions to the mathematical theory of communication, ETC A Rev. Gen. Semant. 10 (4) (1953) 261–281.

[2] L. Floridi, Outline of a theory of strongly semantic information, Minds Mach. 14 (2004) 197–221.

[3] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, J. A. Hendler, Towards a theory of semantic communication, in: 2011 IEEE Network Science Workshop, 2011, pp. 110–117.

[4] B. Güler, A. Yener, A. Swami, The semantic communication game, IEEE Trans. Cogn. Commun. Netw. 4 (4) (2018) 787–802.

[5] X. Luo, H.-H. Chen, Q. Guo, Semantic communications: Overview, open issues, and future research directions, IEEE Wirel. Commun. 29 (1) (2022) 210–219.

[6] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, K. Huang, What is semantic communication? a view on conveying meaning in the era of machine intelligence, J. Commun. Netw. 6 (4) (2021) 336–371.

[7] Y. Zhang, H. Zhao, J. Wei, J. Zhang, M. F. Flanagan, J. Xiong, Context-based semantic communication via dynamic programming, IEEE Transactions on Cognitive Communications and Networking (2022) 1–1.

[8] N. Farsad, M. Rao, A. Goldsmith, Deep learning for joint source-channel coding of text, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, 2018, pp. 2326–2330.

[9] E. Bourtsoulatze, D. Burth Kurka, D. Gündüz, Deep joint source- channel coding for wireless image transmissio, IEEE Trans. Cogn. Commun. Netw. 5 (3) (2019) 567–579.

[10] H. Xie, Z. Qin, G. Y. Li, B. H. Juang, Deep learning enabled semantic communication systems, IEEE Trans. Signal Process. 69 (2021) 2663–2675.

[11] E. C. Strinati, S. Barbarossa, 6G networks: beyond shannon towards semantic and goal-oriented communications, Comput. Netw. 190 (2021) 1389–1286.

[12] Y. Wang, M. Chen, W. Saad, T. Luo, S. Cui, H. V. Poor, Performance optimization for semantic communications: An attention-based learning approach, in: 2021 IEEE Global Communications Conference, 2021, pp. 1–6.

[13] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, R. Q. Hu, Cognitive semantic communication systems driven by knowledge graph, ArXiv abs/2202.11958.

[14] S. Jiang, Y. Liu, Y. Zhang, P. Luo, K. Cao, J. Xiong, H. Zhao, J. Wei, Reliable semantic communication system enabled by knowledge graph, Entropy 24 (6).

[15] D. Huang, X. Tao, F. Gao, J. Lu, Deep learning-based image semantic coding for semantic communications, in: IEEE Global Communications Conference, 2021, pp. 1–6.

[16] Z. Weng, Z. Qin, Semantic communication systems for speech transmission, IEEE J. Sel. Areas Commun. 39 (8) (2021) 2434–2444.

[17] H. Xie, Z. Qin, G. Y. Li, Task-oriented multi-user semantic communications for VQA, IEEE Commun. Lett. 11 (3) (2022) 553–557.

[18] P. Xiang, Q. Zhijin, H. Danlan, X. Tao, L. Jianhua, L. Guangyi, P. Chengkang, A robust deep learning enabled semantic communication system for text, ArXiv abs/2206.02596.

[19] Wireless innovation forum top 10 most wanted wireless innovations. `https://www.wirelessinnovation.org/assets/workproducts/Reports`, 2015.

[20] K. Tang, H. Zhang, B. Wu, W. Luo, W. Liu, Learning to compose dynamic tree structures for visual contexts, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6619–6628.

[21] N. N. Yusof, A. Mohamed, S. Abdul-Rahman, A review of contextual information for context-based approach in sentiment analysis, Int. J. Mach. Learn. Comput. 8 (2018) 399–403.

[22] Z. Zheng, X. Yue, S. Huang, J. Chen, A. Birch, Towards making the most of context in neural machine translation, in: International Joint Conference on Artificial Intelligence, 2020, pp. 3983–3989.

[23] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. Le, R. Salakhutdinov, Transformer-XL: Attentive language models beyond a fixed-length context, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 2978–2988.

[24] Q. Liu, M. J. Kusner, P. Blunsom, A survey on contextual embeddings, arXiv abs/2003.07278.

[25] B. Ahmadnia, B. Dorr, Enhancing phrase-based statistical machine translation by learning phrase representations using long short-term memory network, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing, 2019, pp. 25–32.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, Attention is all you need, in: the 31st International Conference on Neural Information Processing Systems, 2017, pp. 5998–6008.

[27] Y. Tay, M. Dehghani, D. Bahri, D. Metzler, Efficient transformers: A survey, ACM Comput. Surv.

[28] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 4171–4186.

[29] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. J. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, ArXiv abs/2005.14165.

[30] Z. Tan, S. Wang, Z. Yang, G. Chen, X. Huang, M. Sun, Y. Liu, Neural machine translation: A review of methods, resources, and tools, AI Open 1 (2020) 5–21.

[31] K. Papineni, S. Roukos, T. Ward, 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.

[32] S. Banerjee, A. Lavie, METEOR: An automatic metric for MT evaluation with improved correlation with human judgments, in: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, 2005, pp. 65–72.

[33] P. Shaw, J. Uszkoreit, A. Vaswani, Self-attention with relative position representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 464–468.

[34] P. Koehn, Europarl: A parallel corpus for statistical machine translation, MT summit. 5 (2005) 79–86.

[35] Microsoft, Neural network intelligence. `https://github.com/microsoft/nni`,2021.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: