

Knowledge-Enhanced Network Semantic Communication Based on Multi-Objective Optimization

Hassan Touati and Rodrigo C. de Lamare

Abstract—Semantic communication represents a paradigm shift from traditional bit-level to meaning-centric wireless communication. This paper introduces Knowledge-Enhanced Network (KE-Net) semantic communication, a novel framework that leverages external knowledge graphs (KGs) to improve semantic transmission accuracy over wireless channels. The proposed approach employs a multi-objective optimization strategy combining semantic loss, knowledge integration loss, reconstruction loss, and channel-aware loss components. Numerical results show the superior performance of KE-Net semantic communication over competing techniques.

Keywords—Semantic communication, knowledge enhancement, wireless channels, deep learning, BLEU score, multi-objective optimization, and natural language processing.

I. INTRODUCTION

Traditional wireless communication systems prioritize bit-level accuracy and transmission reliability through advanced modulation schemes and channel coding techniques. However, the exponential growth in data traffic and emergence of intelligent applications reveal fundamental limitations in conventional approaches that treat all bits equally regardless of their semantic importance. Semantic communication, inspired by Shannon’s original vision [1], represents a transformative approach that prioritizes the transmission of meaning over exact symbol reconstruction. Recent advances in deep learning and natural language processing [2], [3] have enabled practical implementations of semantic communication systems [4], [5], with comprehensive surveys highlighting the potential of this paradigm [6]. Recent studies have demonstrated the feasibility of end-to-end (E2E) semantic transmission, while subsequent works have explored various aspects, including multi-modal communication [7]. However, existing methods primarily rely on inherent neural network capabilities without leveraging external knowledge sources, limiting their ability to understand context-dependent semantics. This work introduces the Knowledge-Enhanced Network (KE-Net) for semantic communication, addressing key limitations through the integration of external knowledge graphs using Graph Attention Networks (GATs) [8] to enhance semantic understanding and a multi-objective optimization framework that balances semantic fidelity, knowledge integration, reconstruction accuracy, and channel robustness. To this end, we introduce a mathematical formulation with gradient derivations for stable training that is incorporated into the proposed KE-Net semantic communication scheme. Numerical results assess the proposed KE-Net scheme against competing approaches under realistic channel conditions considering rate-distortion trade-offs [9].

The authors are with the Centre for Telecommunication Studies, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Emails: touati.hassan@hotmail.com, delamare@puc-rio.br. Prof. de Lamare is also with the University of York, UK. CNPQ, CPQD, FAPERJ, and FAPESP supported this work.

This paper presents a short version of our extended work [10], accepted for publication in *IEEE Transactions on Vehicular Technology*, which provides a more comprehensive treatment of the KE-Net semantic communication framework, including a complete system design, extended numerical evaluation, and rate-distortion analysis. The present paper focuses on the core system model, the multi-optimization strategy, and the associated gradient derivations underlying the training procedure.

This paper is organized as follows. Section II presents the proposed KE-Net semantic communication system model. Section III details the proposed multi-objective optimization. Section IV presents and discusses the numerical results and Section V draws the conclusions.

II. PROPOSED KE-NET SEMANTIC COMMUNICATION SYSTEM MODEL

Consider a semantic communication system where a transmitter aims to convey the semantic meaning of message $\mathbf{m}$ to a receiver over a wireless channel. The received signal is expressed as:

$$\mathbf{y} = \mathbf{h} \odot \mathbf{s}_k + \mathbf{n} \quad (1)$$

where $\mathbf{s}_k, \mathbf{h}, \mathbf{y}, \mathbf{n} \in \mathbb{R}^{d_s}$ denote the knowledge-enhanced semantic feature vector, channel gain vector, received signal, and additive white Gaussian noise, respectively, with $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. This element-wise formulation assumes that each dimension of the semantic feature vector experiences an independent flat-fading channel gain, consistent with a scenario in which the physical-layer front-end has already equalized cross-dimension coupling prior to semantic decoding. This is a simplifying assumption adopted for analytical and computational tractability. A more general channel model would instead couple the feature dimensions through a full channel matrix, $\mathbf{y} = \mathbf{H}\mathbf{s}_k + \mathbf{n}$, with $\mathbf{H} \in \mathbb{R}^{d_s \times d_s}$; extending KE-Net to this general setting is left for future work.

The proposed KE-Net framework enhances traditional semantic encoding by incorporating external knowledge through knowledge graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents entities and $\mathcal{E}$ denotes relationships. KE-Net redefines semantic communication by introducing a bidirectional knowledge integration framework that operates at both the transmitter and receiver ends. In contrast to prior receiver-centric approaches [11], KE-Net leverages structured KGs and GNNs [12] to maintain semantic coherence throughout the communication process. This architecture stands out in stark contrast to GenSC [13], which relies on BART-based generative models that emphasize token-level dependencies without incorporating external knowledge. While GenSC conceptualizes semantic communication as a generative recovery task, KE-Net approaches it as a knowledge extraction and integration problem. By employing advanced entity linking and graph traversal techniques, KE-Net captures domain-specific terminology, contextual cues, and intricate

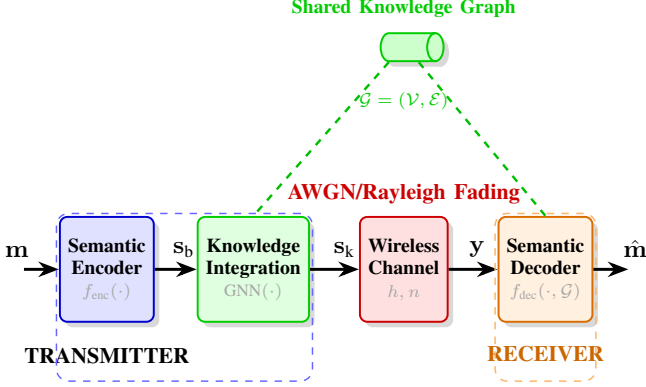


Fig. 1: KE-Net architecture showing knowledge graph integration in semantic communication.

semantic relationships. This transition from simple reconstruction to semantic reasoning marks a significant evolution in the field, setting a new standard for intelligent wireless communication systems.

The proposed KE-Net architecture consists of four main components: the semantic encoder, the knowledge integration module, the channel simulation module, and the semantic decoder. Figure 1 illustrates the overall system architecture.

1) *Semantic Encoder*: The semantic encoder utilizes a Transformer-based T architecture [14] to extract semantic features from input text. Given an input sequence $\mathbf{m} = \{m_1, m_2, \dots, m_T\}$, the encoder produces semantic embeddings through multi-head self-attention mechanisms

$$\mathbf{s}_b = T(\mathbf{m}) \in \mathbb{R}^{L \times d_s} \quad (2)$$

where $T(\cdot)$ is the mathematical mapping performed by the Transformer on the input sequence $\mathbf{m}$, where L denotes the sequence length and d_s represents the embedding dimension. The encoder incorporates positional encoding and layer normalization to capture sequential dependencies and ensure stable training dynamics. To derive a unified semantic representation for knowledge integration, we apply mean pooling across the temporal dimension:

$$\mathbf{r} = \frac{1}{L} \sum_{i=1}^L \mathbf{s}_{b,i} \in \mathbb{R}^{d_s} \quad (3)$$

where $\mathbf{s}_{b,i}$ denotes the i_{th} contextualized embedding vector.

2) *Knowledge Integration Module*: The knowledge integration module employs GNNs to incorporate relevant knowledge from external KGs. The knowledge base integrates ConceptNet 5.7 [15], WordNet 3.1 [16], and domain-specific ontologies, creating a unified knowledge graph with 2.5 million entities and 8 million relationships. The module first identifies relevant entities in the input text through named entity recognition and entity linking. For each identified entity e_i , the corresponding subgraph $\mathcal{G}_i$ is extracted from the knowledge base. A two layer Graph Attention Network (GAT) is then applied to each extracted subgraph, where node representations are updated through neighborhood aggregation. The resulting graph embedding is projected into the semantic feature space and fused with the semantic encoder and decoder through a learnable attention mechanism.

The GNN, F , processes each subgraph to generate knowledge embeddings described by

$$\mathbf{k}_i = F(\mathcal{G}_i) \in \mathbb{R}^{d_s}. \quad (4)$$

These knowledge embeddings are then fused with the base semantic features through an attention mechanism:

$$\mathbf{s}_k = \mathbf{s}_b + \sum_{i=1}^{N_k} \alpha_i \mathbf{k}_i = \mathbf{s}_b + \mathbf{c}, \quad (5)$$

where α_i represents the attention weights computed based on the relevance of each knowledge component to the input semantics.

3) *Channel Simulation Module*: To ensure robustness across various channel conditions, the training process incorporates realistic channel models including AWGN and Rayleigh fading. The channel simulation module applies channel effects during training. For AWGN channels, we have

$$\mathbf{y} = \mathbf{s}_k + \mathbf{n}. \quad (6)$$

For Rayleigh fading channels, we employ

$$\mathbf{y} = \mathbf{h} \odot \mathbf{s}_k + \mathbf{n}, \quad (7)$$

where each element h_j , $j = 1, \dots, d_s$, is drawn i.i.d. from a Rayleigh distribution with scale parameter $\sigma_h = 1/\sqrt{2}$, giving unit average power ($\mathbb{E}[h_j^2] = 1$). The channel gain vector $\mathbf{h}$ is assumed constant over each transmitted feature vector (block fading) and independently redrawn thereafter. Noise power σ^2 is set according to the target SNR in Section IV.

4) *Semantic Decoder*: The semantic decoder reconstructs the original semantic meaning from the received noisy features. The decoder employs a Transformer-based architecture with cross-attention mechanisms to handle the knowledge-enhanced representations:

$$\hat{\mathbf{m}} = \mu(\mathbf{y}, \mathcal{G}), \quad (8)$$

The decoder μ incorporates beam search for improved reconstruction quality and supports variable-length output generation.

III. PROPOSED MULTI-OBJECTIVE OPTIMIZATION

In this section, we present the proposed multi-objective optimization strategy. We first detail the adopted multi-objective loss function, and then we describe the proposed optimization algorithm along with the expressions of the gradient terms.

A. Multi-Objective Loss Function

The KE-Net training process optimizes a multi-objective loss function that balances semantic fidelity, knowledge integration, reconstruction accuracy, and channel robustness. The total loss is defined as:

$$L_{\text{total}} = \alpha L_{\text{semantic}} + \beta L_{\text{knowledge}} + \gamma L_{\text{reconstruction}} + \delta L_{\text{channel}}, \quad (9)$$

where L_{semantic} , $L_{\text{knowledge}}$, $L_{\text{reconstruction}}$ and L_{channel} are the semantic, knowledge, reconstruction and channel losses, respectively, and α , β , γ and δ are their weights, respectively. The details of the total loss terms are given in what follows.

1) *Semantic Loss*: The primary evaluation metric is the BLEU score [17], computed for n-gram orders 1 through 4:

$$\text{BLEU}_n = BP \cdot \exp\left(\frac{1}{N} \sum_{i=1}^N \log p_i\right) \quad (10)$$

The BLEU evaluation metric incorporates a brevity penalty (BP) to discourage translations that are too short, which may inflate precision while omitting important content. The penalty is defined as

$$BP = \begin{cases} 1, & \text{if } c > r \\ e^{(1-\frac{c}{r})}, & \text{if } c \leq r \end{cases} \quad (11)$$

where c is the length of the candidate translation and r is the length of the reference translation. This ensures that overly brief outputs are penalized proportionally. In addition to BP, BLEU evaluates translation quality using n-gram precision, typically from 1-gram to 4-gram, to capture lexical accuracy and contextual fluency. For example, in the sentence: " Semantic communication improves efficiency. " 1-gram :["Semantic", "communication", "improves", "efficiency"]; 2-gram : ["Semantic communication", "communication improves", "improves efficiency"]; 3-gram : ["Semantic communication improves", "communication improves efficiency"]; and p_i represents the precision for the i-th. The semantic loss is formulated based on BLEU scores, which quantify the quality of semantic reconstruction:

$$L_{\text{semantic}} = -\frac{1}{N} \sum_{n=1}^4 w_n \cdot \log(\text{BLEU}_n), \quad (12)$$

where w_n denotes the weights for different n-gram orders, typically set to $w_n = 1/4$ for balanced evaluation.

2) *Knowledge Loss*: The knowledge loss ensures effective integration of external knowledge:

$$L_{\text{knowledge}} = \sum_{i=1}^{N_t} -w_i [t_i \cdot \log \hat{t}_i + (1 - t_i) \cdot \log(1 - \hat{t}_i)], \quad (13)$$

where t_i and $\hat{t}_i$ represent the ground truth and predicted knowledge relevance scores, and w_i are importance weights.

3) *Reconstruction Loss*: The reconstruction loss measures the fidelity of semantic feature reconstruction:

$$L_{\text{reconstruction}} = \frac{1}{|S|} \sum_{j=1}^{|S|} \|\mathbf{s}_j - \hat{\mathbf{s}}_j\|_2^2 \quad (14)$$

This loss ensures that the semantic features are accurately preserved through the transmission process.

4) *Channel-Aware Loss*: The channel-aware loss accounts for channel impairments during training:

$$L_{\text{channel}} = \mathbb{E}_{\mathbf{h} \sim p(\mathbf{h})} [\|\mathbf{y} - \mathbf{h} \odot \hat{\mathbf{y}}\|_2^2] \quad (15)$$

This loss term ensures robust performance across different channel realizations.

During training, the expectation is approximated using Monte Carlo sampling over channel realizations. For each mini-batch, independent channel coefficients are randomly generated according to the assumed channel distribution (AWGN or Rayleigh fading), and the empirical average over the mini-batch is used to estimate the expectation.

B. Optimization Strategy

The training process employs the Adam optimizer [18] with dynamic loss coefficient adaptation. The loss coefficients are adjusted during training to balance different objectives:

$$\alpha_t = \alpha_0 \cdot \exp(-\lambda_1 \cdot t) \quad (16)$$

$$\beta_t = \beta_0 \cdot (1 + \tanh(\lambda_2 \cdot (t - t_0))) \quad (17)$$

The optimization process continues until convergence, defined as:

$$\frac{|L_{\text{total}}^{(t)} - L_{\text{total}}^{(t-1)}|}{L_{\text{total}}^{(t-1)}} < \epsilon_{\text{conv}} \quad (18)$$

where $\epsilon_{\text{conv}} = 1 \times 10^{-5}$ is the convergence threshold. The Adam optimizer parameters are updated according to:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (19)$$

where $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected moment estimates.

C. Gradient Derivations and Optimization Analysis

The gradient of the total loss function with respect to model parameters θ is computed as:

$$\nabla_{\theta} L_{\text{total}} = \alpha \nabla_{\theta} L_{\text{semantic}} + \beta \nabla_{\theta} L_{\text{knowledge}} + \gamma \nabla_{\theta} L_{\text{reconstruction}} + \delta \nabla_{\theta} L_{\text{channel}} \quad (20)$$

For the semantic loss gradient:

$$\nabla_{\theta} L_{\text{semantic}} = -\frac{1}{N} \sum_{n=1}^4 \frac{w_n}{\text{BLEU}_n} \nabla_{\theta} \text{BLEU}_n \quad (21)$$

We then obtain the final expression:

$$\frac{\partial L_{\text{semantic}}}{\partial \theta} = -\frac{1}{N} \sum_{n=1}^N w_n \left[\frac{1}{BP} \cdot \frac{\partial BP}{\partial \theta} + \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i} \cdot \frac{\partial p_i}{\partial \theta} \right] \quad (22)$$

The knowledge loss gradient involves the GNN parameters θ_{GNN} :

$$\begin{aligned} \frac{\partial L_{\text{knowledge}}}{\partial \theta_{\text{GNN}}} &= \sum_{i=1}^{N_t} -w_i \cdot \frac{\partial}{\partial \theta_{\text{GNN}}} [t_i \log(\hat{t}_i) + (1 - t_i) \log(1 - \hat{t}_i)] \\ &= \nabla_{\theta_{\text{GNN}}} L_{\text{knowledge}} = \sum_{i=1}^{N_t} w_i \left(\frac{t_i - \hat{t}_i}{\hat{t}_i(1 - \hat{t}_i)} \right) \nabla_{\theta_{\text{GNN}}} \hat{t}_i \end{aligned} \quad (24)$$

This gradient is essential in optimizing binary classification models and is commonly found in logistic regression and neural networks. The reconstruction loss gradient with respect to encoder parameters θ_{enc} :

$$\nabla_{\theta_{\text{enc}}} L_{\text{reconstruction}} = \frac{2}{|S|} \sum_{j=1}^{|S|} (\mathbf{s}_j - \hat{\mathbf{s}}_j) \nabla_{\theta_{\text{enc}}} (\mathbf{s}_j - \hat{\mathbf{s}}_j) \quad (25)$$

Algorithm 1 summarizes the complete KE-Net training procedure, integrating the semantic encoding, knowledge fusion, channel simulation, and multi-objective loss computation steps described above.

D. Computational Complexity Analysis

The Transformer encoder/decoder has complexity $\mathcal{O}(L^2 d_s)$ per sample. For the knowledge integration module, each two-layer GAT pass over a subgraph $\mathcal{G}_i$ costs $\mathcal{O}(|\mathcal{V}_i| d_s^2 + |\mathcal{E}_i| d_s)$; with N_k linked entities, the total overhead is $\mathcal{O}(N_k d_s^2)$, which remains modest since $N_k \ll L$ and subgraphs are restricted to bounded local neighborhoods.

Algorithm 1 KE-Net Training Algorithm**Require:** Dataset $\mathcal{D}$, knowledge graph $\mathcal{G}$, hyperparameters**Ensure:** Trained KE-Net parameters θ^*

```

1: Initialize  $\theta_0$  randomly, optimizer parameters
2: for epoch = 1 to  $N_{\text{epochs}}$  do
3:   for batch in  $\mathcal{D}$  do
4:     Extract features:  $\mathbf{s}_m = \mathbf{T}(\mathbf{x})$ 
5:     Link entities:  $\mathcal{G}_{\text{rel}} = \text{EntityLink}(\mathbf{m}, \mathcal{G})$ 
6:     Embed knowledge:  $\mathbf{k} = \mathbf{F}(\mathcal{G}_{\text{rel}})$ 
7:     Fuse:  $\mathbf{s}_k = \text{Attention}(\mathbf{s}_b, \mathbf{k})$ 
8:     Channel:  $\mathbf{y} = \text{Channel}(\mathbf{h} \odot \mathbf{s}_k + \mathbf{n})$ 
9:     Decode:  $\hat{\mathbf{m}} = \mu(\mathbf{y}, \mathcal{G})$ 
10:    Compute  $L_{\text{total}}$ , update  $\theta$ , adapt weights
11:   end for
12:   if convergence met then
13:     break
14:   end if
15: end for

```

IV. NUMERICAL RESULTS

The performance evaluation of the proposed KE-Net scheme against competing techniques employs a dataset of 50,000 sentences constructed from three publicly available corpora: AG News [19] (40%), Sentiment140 [20] (35%), and the arXiv dataset [21] (25%). The samples are randomly selected from each corpus, preprocessed using the same tokenization and normalization pipeline, and partitioned into training (70%), validation (15%), and test (15%) sets. The identical dataset and preprocessing procedure are used for all compared methods to ensure fair and reproducible evaluation. The performance of the proposed framework is evaluated using BLEU-1 to BLEU-4 scores. The training parameters include a learning rate of 10^{-4} , batch size of 32, the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and loss weights of $\alpha = 1.0$, $\beta = 0.5$, $\gamma = 0.3$, and $\delta = 0.2$.

Table I summarizes the performance metrics obtained for KE-Net, GenSC, and DeepSC. As shown in Fig. 2 and Fig. 5, KE-Net achieves substantial gains in both 1-gram and 4-gram BLEU scores over DeepSC under AWGN conditions, with the relative improvement being more pronounced for higher-order n-grams. KE-Net consistently outperforms GenSC and DeepSC across all SNR values (0–18 dB) and n-gram configurations, with AWGN channels yielding higher BLEU scores than fading channels for all three systems. As expected, BLEU scores degrade as the n-gram order increases, reflecting the increasing difficulty of preserving longer semantic sequences intact over noisy wireless channels. Figs. 3 and 4 show the corresponding 2-gram and 3-gram BLEU performance. The relative ranking of the three systems remains consistent with the 1-gram and 4-gram cases, with KE-Net maintaining a clear advantage over GenSC and DeepSC under both AWGN and fading conditions. These intermediate n-gram results confirm KE-Net’s superior preservation of semantic structure across varying channel conditions.

Notably, the performance gap between systems widens under adverse channel conditions, with KE-Net exhibiting the greatest robustness to channel impairments among the three methods across all n-gram orders. This robustness advantage highlights the benefit of incorporating knowledge-graph-based

TABLE I: Performance Comparison: Average BLEU Scores and Improvements

Method	Average BLEU Score		
	1-gram	4-gram	Overall
KE-Net	0.837	0.565	0.693
GenSC	0.785	0.505	0.638
DeepSC	0.568	0.280	0.407
Improvement vs GenSC	6.6%	7.8%	8.6%
Improvement vs DeepSC	47.4%	101.8%	70.3%

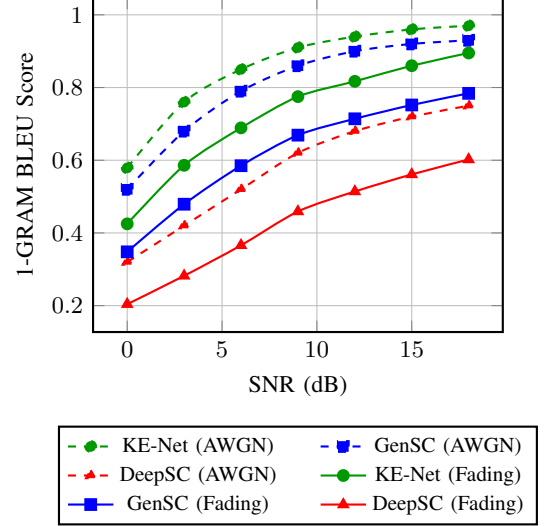


Fig. 2: 1-GRAM BLEU Score Performance Comparison for AWGN and Rayleigh Fading Channels

semantic reasoning for reliable performance in practical wireless deployment scenarios.

A. Ablation Study

Table II presents an ablation study of the proposed KE-Net framework. The results show that knowledge graph integration improves semantic reconstruction performance, while the proposed bidirectional transmitter-receiver integration achieves the highest BLEU scores.

V. CONCLUSIONS

This paper introduced KE-Net, a knowledge-enhanced semantic communication framework. KE-Net integrates external knowledge graphs into the encoding process via graph neural networks. It further applies multi-objective optimization with mathematically derived gradients to jointly balance semantic, Knowledge, reconstruction, and channel losses. Numerical results show that KE-Net achieves robust performance under realistic channel conditions and consistently outperforms existing methods across multiple BLEU-based evaluation metrics. Extending KE-Net to Portuguese would require a Portuguese knowledge base for entity linking.

Future research directions include extending the framework to multi-modal communication scenarios, developing more efficient knowledge integration mechanisms, and exploring applications in emerging 6G wireless technologies. The proposed KE-Net framework is a significant advancement toward practical semantic communication systems.

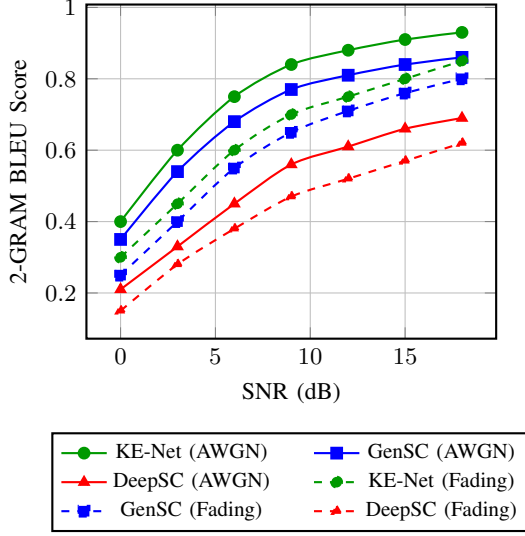


Fig. 3: 2-GRAM BLEU performance under AWGN and Fading channels

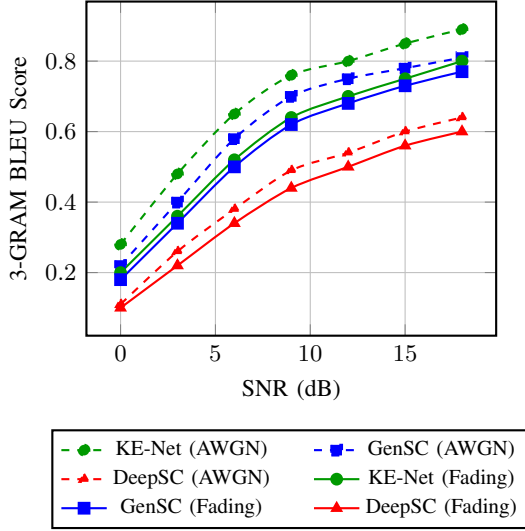


Fig. 4: 3-GRAM BLEU performance under AWGN and Fading channels

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [2] H. Strobelt, S. Gehrmann, M. Behrisch, A. Perer, H. Pfister and A. M. Rush, "Seq2seq-Vis: A Visual Debugging Tool for Sequence-to-Sequence Models," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 353–363, Jan. 2019.
- [3] Y. Goldberg, "Neural network methods for natural language processing," *Synth. Lect. Hum. Lang. Technol.*, vol. 10, no. 1, pp. 1–309, 2017.
- [4] 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.
- [5] Z. Weng, Z. Qin and G. Y. Li, "Semantic Communications for Speech Signals," *ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada*, pp. 1–6, 2021.
- [6] Y. Sun, L. Zhang, D. Niyato, M. A. Imran, "Wireless Semantic Communications: Concepts, Principles, and Challenges," Wiley Telecommunications, 2025.
- [7] Y. Weng, C. Zhang, P. Zhang, and L. Yang, "Semantic communication with adaptive universal transformer," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 632–636, 2022.

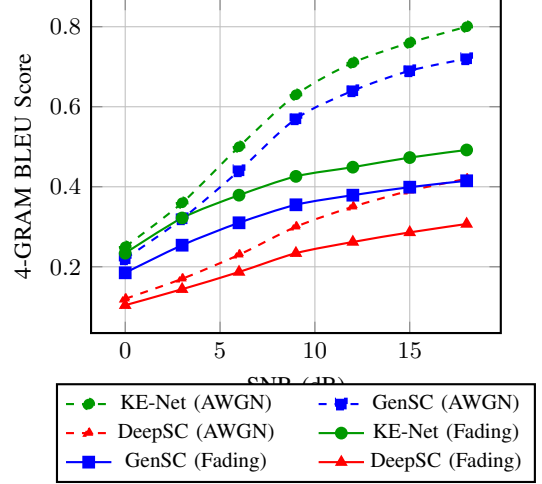


Fig. 5: 4-GRAM BLEU Score Performance Comparison for AWGN and Rayleigh Fading Channels

TABLE II: Ablation study of KE-Net (6 dB, AWGN).

Variant	BLEU-1	BLEU-4
KE-Net (TX+RX KG)	0.850	0.500
TX-side KG only	0.835	0.482
RX-side KG only	0.827	0.474
Without KG	0.798	0.446

- [8] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," in *Proc. International Conference on Learning Representations (ICLR)*, 2018.
- [9] D. Li, J. Huang, C. Huang, X. Qin, H. Zhang and P. Zhang, "Fundamental Limitation of Semantic Communications: Neural Estimation for Rate-Distortion," *Journal of Communications and Information Networks*, vol. 8, no. 4, pp. 303–318, Dec. 2023.
- [10] H. Touati and R. C. de Lamare, "Knowledge-Enhanced Network Semantic Communication Based on Multi-Objective Optimization," *IEEE Trans. Veh. Technol.*, 2026.
- [11] B. Wang, R. Li, J. Zhu, Z. Zhao and H. Zhang, "Knowledge Enhanced Semantic Communication Receiver," *IEEE Communications Letters*, vol. 27, no. 7, pp. 1794–1798, July 2023.
- [12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, Jan. 2021.
- [13] M.-K. Chang, C.-T. Hsu and G.-C. Yang, "GenSC: Generative Semantic Communication Systems Using BART-Like Model," *IEEE Communications Letters*, vol. 28, no. 10, pp. 2298–2302, Oct. 2024.
- [14] A. Vaswani et al., "Attention is all you need," *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pp. 5998–6008, 2017.
- [15] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An open multilingual graph of general knowledge," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4444–4451.
- [16] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [17] 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. Linguistics (ACL)*, Philadelphia, PA, USA, 2002, pp. 311–318.
- [18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [19] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 28, 2015, pp. 649–657.
- [20] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Stanford University, CS224N Project Report, 2009.
- [21] C. Clement, M. Kreutz, R. Homann, and J. Beckmann, "On the use of arXiv as a dataset," *arXiv preprint arXiv:1905.00075*, 2019.