Few-Shot Adaptive Learning for Robust Task-Oriented Semantic Communication

Anonymous submission

Abstract

Task-oriented semantic communication (ToSC) emerges as a promising approach for executing remote inference tasks. While existing ToSC systems are generally trained under specified channel conditions, the volatile nature of real-world channel conditions poses significant adaptation challenges to ToSC. To this end, we propose an adaptive ToSC system for dynamic environments via few-shot learning in this paper. The method utilizes the data-driven mechanism named vector quantized variational autoencoder (VQ-VAE) to dynamically optimize the codebook and generate non-uniform modulation codebooks that are closely aligned with specific task objectives. In addition, few-shot learning and transfer learning techniques are adopted to facilitate efficient learning on small datasets, allowing the system to swiftly adjust its operating parameters to adapt to new communication conditions. Experimental results show that the proposed method achieves superior performance compared to traditional channel-adaptive methods, especially in environments with low signal-to-noise ratios (SNR).

Introduction

The massive artificial intelligence (AI) applications pose a major challenge to communication bandwidth. Applications such as smart city (Mylonas et al. 2021) and intelligent factory (Wang et al. 2021) demand substantial data support, and their real-time transmission could consume considerable bandwidth, risking network capacity saturation. To address the challenges posed by limited bandwidth, a new paradigm, task-oriented semantic communication (ToSC) (Xie, Qin, and Li 2022; Zhang, Wang, and et al 2024) was proposed for remote inference. ToSC can alleviate the scarcity of transmission resources by leveraging transceivers' computational resources through split learning. Moreover, when tasks are distributed across edge networks, ToSC significantly enhances the reliability and real-time performance of communication systems, due to its high flexibility. ToSC mainly uses deep neural networks (DNN) to implement joint source-channel coding (JSCC) (Choi et al. 2019) and extracts task-related continuous semantic information for efficient transmission. To transmit continuous feature data, element-wise quantization is often used. However, this approach is susceptible to distortion, as task performance depends on the collective contribution of the entire feature vector rather than the individual element. To address this problem, vector quantized variational autoencode (VQ-VAE) (Hu et al. 2023) technology is introduced, incorporating a vector quantization mechanism to map continuous features into discrete codes.

It is important to note that current ToSC systems are typically trained under the assumption of static data and channel distributions, with the implicit expectation that channel conditions will remain unchanged. However, the dynamic nature of real-world communication environments, where channel conditions are subject to fluctuation, often leads to the failure or suboptimal performance of these systems. To enhance the adaptability of the system across varying channel conditions, (Raghuram et al. 2021) suggested employing autoencoders for domain adaptation to reduce the need for frequent retraining. This method was developed for task-agnostic end-to-end communication systems relying on uniform distributed one-hot encoding, which struggles to effectively capture the potential structure and task relevance in semantic information.

Motivated by the above challenges, this paper proposes a few-shot adaptive method based on non-uniform feature codebook mapping, called FA-NFM. Specifically, to improve the system's resilience to the rapid changes of channel conditions, we adopt a mixture density network (MDN) (Garcia Marti et al. 2020) to solve the domain adaptation (DA) problem in the autoencoder. It uses a Gaussian mixture model (GMM) (Plataniotis and Hatzinakos 2017) to model channel characteristics and adopts few-shot learning to quickly capture changes in channel distribution. Based on the learned task-related features, an optimal transformation is designed at the decoder input to compensate for the distribution shift caused by channel changes, ensuring high decoding accuracy and transmission reliability. Experimental results show that this method provides reliable transmission of task-related features and demonstrates robust adaptability to varying channel conditions.

Related Work

Recent progress in ToSC has introduced innovative methods to enhance communication efficiency and adaptability. For example, (Hu et al. 2023) and (Wang et al. 2024a) proposed a semantic bit quantization (SBQ)-driven resource allocation paradigm, which optimizes semantic communication qual-

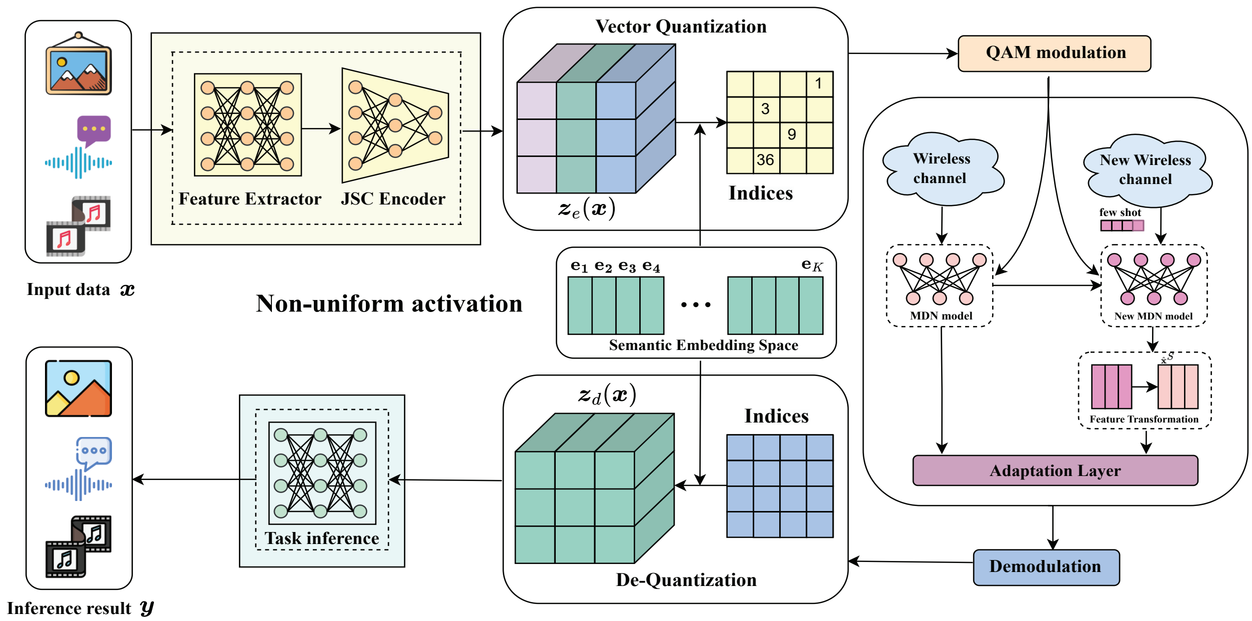

Figure 1: Architecture of the Proposed FA-NFM System

ity of service (SC-QoS) through hybrid uniform-nonuniform quantization. Complementing this, (Fu et al. 2023) introduced semantic evolution, an automated approach to generate high-quality semantic information, reducing the need for complex network structures.

To further minimize communication overhead, (Li et al. 2024) proposed extracting task-focused subgraphs that preserve critical information. Additionally, (Zhang and Guo 2024) developed a multi-rate task-oriented communication (MR-ToC) framework, which dynamically adapts to variations in data rates, improving robustness. Beyond efficiency, privacy remains a key concern. (Wang et al. 2024b) addressed this with the Information Bottleneck and Adversarial Learning (IBAL) framework, which safeguards user privacy against model inversion attacks.

For channel adaptivity, recent advancements have focused on addressing the challenges of bias elimination and domain adaptation in few-shot learning. (Zhang et al. 2019) introduced a variational Bayesian framework utilizing stochastic variational inference to approximate bias-eliminated, class-specific sample distributions, significantly improving few-shot learning performance. Methods like DANN (Ganin et al. 2016) employ adversarial training to align source and target distributions by learning shared representations. While effective, these approaches often require extensive source-domain labels and are computationally intensive, making them less suitable for rapid test-time adaptation. Few-shot DA frameworks, such as (Motiian et al. 2017), alleviate the dependency on labeled data but focus primarily on training-time adaptation, leaving the critical challenge of test-time adaptation insufficiently addressed.

Methodology

Overall Design

We propose an adaptive discrete ToSC system consisting of an encoder, codebook, modulation, physical channel, channel adaptor, demodulation, and task inference module.

1) Feature Encoding The system input $x \in \mathbb{R}^n$ is processed by a feature extractor and a joint source-channel (JSC) encoder to obtain a continuous representation $\mathbf{z}_e(x) \in \mathbb{R}^D$:

$$z_e(x) = T(x; \theta), \tag{1}$$

where $T(x; \theta)$ combines the feature extractor and JSC encoder with parameters θ.

2) Discrete Codebook Mapping To ensure compatibility with digital systems, vector quantization maps $z_c(x)$ to a discrete codebook $\mathcal{Z}_c = \{e_j \in \mathbb{R}^D \mid j = 1, \ldots, K\}$:

$$z_c(x) = \arg\min_{e_j} \|z_e(x) - e_j\|_2, \quad \forall e_j \in \mathcal{Z}_c. \tag{2}$$

3) Codebook Training The codebook $\mathcal{Z}_c$ is jointly trained with the encoder and task inference module. To handle the non-differentiability of $\arg\min$, we use a straight-through estimator, approximating the forward propagation loss $\mathcal{L}_{\text{VQ}}$(van den Oord, Vinyals, and kavukcuoglu 2017) as:

$$\mathcal{L}_{\text{VQ}} = \|\text{sg}[z_e(x)] - e_j\|_2^2 + \gamma \|z_e(x) - \text{sg}[e_j]\|_2^2, \tag{3}$$

where γ is a hyperparameter and sg$[\cdot]$ represents the stop-gradient operation.

4) Digital Modulation and Demodulation The discrete representation is transmitted via digital modulation, mapping the codebook indices to constellation symbols. The received signal $\hat{z}$ is:

$$\hat{z} = g(h(z)), \tag{4}$$

where $h(\cdot)$ is the channel function, and $g(\cdot)$ is the demodulation function. The receiver then performs demapping to obtain $z_d(z)$ based on $\hat{z}$.

Channel Modeling

The channel (i.e., propagation medium and transceiver imperfections) can be represented as a stochastic transfer function that transforms its input $\mathbf{z} \in \mathbb{R}^d$ to an output $\mathbf{x} \in \mathbb{R}^d$. In order to learn the encoder and decoder networks using stochastic gradient descent (SGD)-based optimization, it is necessary to have a differentiable backward path from the decoder to the encoder through the channel. We address this by learning a parametric generative model of the channel $P_{\theta_c}(\mathbf{x} \mid \mathbf{z})$ (with parameters θ_c) that closely approximates the true channel conditional density $p(\mathbf{x} \mid \mathbf{z})$. In this work, we model the conditional density of the channel using a set of m Gaussian mixtures:

$$P_{\theta_c}(x \mid z) = \sum_{i=1}^{k} \pi_i(z) N\big(x \mid \mu_i(z), \Sigma_i(z)\big), \tag{5}$$

where k is the number of components, $\mu_i(\mathbf{z}) \in \mathbb{R}^d$ is the mean vector, $\Sigma_i(\mathbf{z}) \in \mathbb{R}^{d\times d}$ is the (symmetric, positive-definite) covariance matrix, and $\pi_i(\mathbf{z}) \in [0,1]$ is the prior probability of the i-th component, satisfying $\sum_{i=1}^{k} \pi_i(\mathbf{z}) = 1$. The prior probabilities $\pi_i(\mathbf{z})$ are expressed using the softmax function:

$$\pi_i(\mathbf{z}) = e^{\alpha_i(\mathbf{z})} \Big/ \sum_{j=1}^{k} e^{\alpha_j(\mathbf{z})}, \quad \forall i \in [k], \tag{6}$$

where $\alpha_i(\mathbf{z}) \in \mathbb{R}$ are the component prior logits. To capture all parameters, we define the parameter vector of component i as:

$$\phi_i(\mathbf{z})^T = [\alpha_i(\mathbf{z}), \mu_i(\mathbf{z})^T, \mathrm{vec}(\Sigma_i(\mathbf{z}))^T], \tag{7}$$

wherevec$(\Sigma_i(\mathbf{z}))$ is the vectorized form of the unique entries of the symmetric covariance matrix $\Sigma_i(\mathbf{z})$ (e.g., upper triangular or lower triangular elements).

Finally, the combined parameter vector for all components is defined as:

$$\phi(\mathbf{z})^T = [\phi_1(\mathbf{z})^T, \ldots, \phi_k(\mathbf{z})^T]. \tag{8}$$

A mixture density network (MDN) (Bishop 1994), which combines a feedforward network with Gaussian mixtures, is employed to capture complex conditional distributions.

Fine-Tuning with Gaussian Mixture Models

To adapt an existing GMM-based channel model to new channel conditions, we adjust the model parameters through affine transformations to ensure statistical consistency with the new channel distribution. The fine-tuning adjustments includes the following key steps:

1) Affine Transformation of Mean Vectors Adjust each Gaussian component's mean vector $\mu_i(z)$ with a transformation matrix A_i and a bias vector b_i as follows:

$$\hat{\mu}_i(z) = A_i \mu_i(z) + b_i. \tag{9}$$

2) Affine Transformation of Covariance Matrices Modify the covariance matrices $\Sigma_i(z)$ with a scaling factor matrix C_i as follows:

$$\hat{\Sigma}_i(z) = C_i \Sigma_i(z) C_i^T. \tag{10}$$

3) Weight Adjustment of Gaussian Components Update the mixing weights $\alpha_i(z)$ with scaling and offset parameters β_i and γ_i as follows:

$$\hat{\alpha}_i(z) = \beta_i \alpha_i(z) + \gamma_i. \tag{11}$$

This method presents two principal advantages: firstly, it circumvents the need to retrain the entire network, significantly reducing computational cost; second, the transformation parameters for fine-tuning are far fewer than the original network parameters, necessitating less data for effective adaptation. As supported by (García Martí et al. 2020), such affine transformations in GMMs not only efficient but also preserve the model's analytical traits, rendering them well-suited for dynamic channel modeling.

Inverse Transformation for Feature Alignment

To address domain adaptation in low-sample scenarios, we introduce an efficient inverse feature transformation, g^{-1}, which aligns target input features x^t with source feature distributions. This transformation is based on an affine adjustment of Gaussian components, avoiding encoder retraining. The transformation is defined as:

$$\hat{\mathbf{x}}^s = g_{z_i}^{-1}(\mathbf{x}^t) := \mathbf{C}_i^{-1}\big(\mathbf{x}^t - \mathbf{A}_i \mu_i(z) - \mathbf{b}_i\big) + \mu_i(z), \tag{12}$$

where C_i is the covariance matrix of the i-th Gaussian component, A_i is the linear transformation matrix, b_i is the bias vector, and $\mu_i(z)$ is the mean vector. This formulation enables domain alignment by projecting target features back into the source distribution.

To solve for the posterior distribution, the target posterior $P_{\hat{\theta}_c}(z, i \mid \mathbf{x}^t)$ is computed as:

$$P_{\hat{\theta}_c}(z, i \mid \mathbf{x}^t) = \frac{p(z)\hat{\pi}_i(z) N(\mathbf{x}^t \mid \hat{\mu}_i(z), \hat{\Sigma}_i(z))}{\sum_{z'} \sum_j p(z')\hat{\pi}_j(z') N(\mathbf{x}^t \mid \hat{\mu}_j(z'), \hat{\Sigma}_j(z'))}, \tag{13}$$

where $P_{\hat{\theta}_c}(z, i \mid \mathbf{x}^t)$ represents the conditional probability, $\hat{\theta}_c$ denotes the model parameters, $p(z)$ refers to the prior probability of symbol z, and $\hat{\pi}_i(z)$ denotes the probability of selecting component i given the symbol z, $N(\mathbf{x}^t \mid \hat{\mu}_i(z), \hat{\Sigma}_i(z))$ represents the probability density function of $\mathbf{x}^t$ under the given symbol z and component i, $\hat{\mu}_i(z)$ is the mean vector, and $\hat{\Sigma}_i(z)$ is the covariance matrix.

The denominator represents the summation of the numerator over all possible z' and j, ensuring that the conditional probabilities are properly normalized to sum to 1. This normalization factor guarantees that the result is a valid probability distribution.

The affine transformation also defines the regularized inverse feature mapping as:

$$\begin{aligned}\mathbf{g}^{-1}(\mathbf{x}^t) &:= \mathbb{E}_{P_{\hat{\theta}_c}(z,i|\mathbf{x})}\left[\mathbf{g}_{z_i}^{-1}(\mathbf{x}^t) \mid \mathbf{x}^t\right] \\ &= \sum_{z\in\mathcal{Z}} \sum_{i\in[k]} P_{\hat{\theta}_c}(z, i \mid \mathbf{x}^t) \mathbf{g}_{z_i}^{-1}(\mathbf{x}^t), \end{aligned} \tag{14}$$

which ensures that the target feature is mapped optimally to the source domain under the posterior distribution. This approach, based on Gaussian mixture alignment, provides a practical solution for adapting varying channels while minimizing computational overhead.

Experiments

In this section, we conduct comprehensive experiments to demonstrate the superiority of the proposed FA-NFM method. In this study, we focus on the task of image reconstruction, utilizing the CIFAR-10 dataset as our experimental data and adopting the peak signal-to-noise ratio (PSNR) as the metric to assess adaptability. A 16-QAM constellation is employed for modulation, with a codebook size of 256. Channel variations are simulated based on standard wireless communication models: (i) additive Gaussian noise (AWGN), (ii) Ricean fading, and (iii) uniform fading (Goldsmith 2005). The initial channel condition is established with an SNR of 14 dB, and the system's performance is evaluated across a range of test SNR values, specifically $SNR_{\text{test}} \in \{0\ dB, 2\ dB, 4\ dB, 6\ dB, 8\ dB, 10\ dB\}$. For comparative analysis, we have chosen four following cutting-edge methodologies as our benchmarks:

- **Systerm Without Adjustment (STWA):**semantic information of images using VQ-VAE without any adjustments to the model after channel variations. It serves as a benchmark to assess the degradation in image reconstruction performance caused by channel variations.
- **Fine-Tune & Random (FTR):**(Raghuram et al. 2021) This method utilizes a small number of randomly generated uniform one-hot encodings to first fine-tune the entire MDN model. Subsequently, the fine-tuned channel model is used to generate data for further optimization of the entire ToSC model.
- **Selective Noise Adaptation (SNA):** This method fine-tunes the entire TOSC system by selecting a small amount of unknown noise data. The goal is to improve the system's adaptability and robustness in complex environments.
- **Adaption & Random (ATR):**(Raghuram et al. 2021) This method employs a small number of randomly generated uniform one-hot encodings to create an MDN model with target channel characteristics through an affine transformation of the original channel model. Subsequently, an adaptive layer is introduced to perform inverse affine transformations on the data.

Table 1: PSNR performance under conditions transitioning from Uniform fading to Ricean fading

SNR (dB)	STWA	ATR	FTR	SNA	FA-NFM
10	17.81	18.36	17.69	17.50	**18.57**
8	17.40	17.70	17.15	16.75	**17.98**
6	16.83	17.19	16.73	16.35	**17.35**
4	16.11	16.66	16.24	15.94	**16.67**
2	15.22	17.86	17.63	16.97	**18.50**

Table 2: PSNR performance under conditions transitioning from AWGN to Uniform fading

SNR (dB)	STWA	ATR	FTR	SNA	FA-NFM
10	15.48	15.89	16.05	**16.66**	16.53
8	14.48	15.52	15.36	15.61	**16.32**
6	13.50	15.10	14.84	14.44	**15.32**
4	12.68	14.35	14.46	13.37	**15.07**
2	11.93	14.27	14.21	12.97	**14.45**

Table 3: PSNR performance under conditions transitioning from Ricean fading to Uniform fading

SNR (dB)	STWA	ATR	FTR	SNA	FA-NFM
10	16.09	16.87	16.77	16.98	**17.00**
8	15.02	16.30	16.29	15.96	**16.61**
6	14.02	15.79	15.74	14.81	**15.93**
4	13.11	15.11	15.38	14.24	**15.63**
2	12.35	14.95	15.14	13.98	**15.18**

The experimental results in Tables 1-3 demonstrate that our proposed method exhibits superior stability compared to fine-tuning and non-adaptive approaches, while significantly enhancing the quality of image reconstruction.

We further evaluated the impact of few-shot learning adjusting the number of adaptive samples assigned to each constellation symbol. Figure 2 presents the relationship between the target sample per class and the PSNR of the reconstructed images, with experiments conducted under conditions transitioning from an AWGN channel with SNR=14dB to a uniform fading channel with SNR=6dB. The results indicate that when the number of adaptive samples reaches 20, our proposed method demonstrates a significant improvement in PSNR, highlighting its effectiveness in scenarios with a limited number of samples.

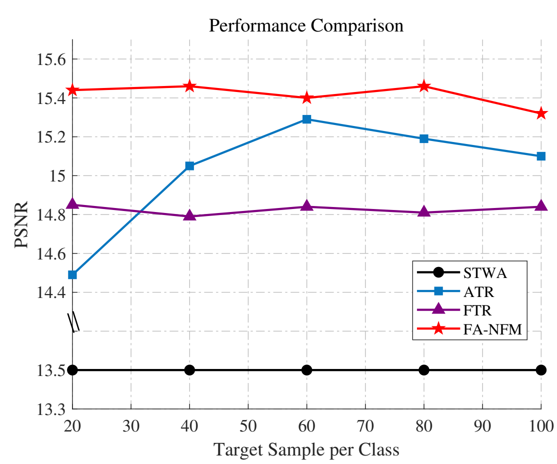

Figure 2: PSNR vs. Adaptive Sample Count

Conclusion

In this study, we investigated the challenge of channel adaptation in ToSC systems and proposed a few-shot adaptive solution based on non-uniform codebook feature mapping, named FA-NFM. Our approach involves designing a VQ-VAE system tailored for task-oriented communication, leveraging a data-driven mechanism to dynamically optimize the codebook that is tightly coupled to the specific task objective. To enhance adaptability to channel variation, few-shot learning is employed to achieve effective adaptation under different channel conditions. Experimental results show that our approach achieves superior performance in image reconstruction tasks while requiring fewer resources compared to existing methods. Looking ahead, FA-NFM has the potential to provide an effective solution for channel adaptation in AI-native communication systems.

References

Bishop, C. M. 1994. Mixture density networks.

Choi, K.; Tatwawadi, K.; Grover, A.; Weissman, T.; and Ermon, S. 2019. Neural Joint Source-Channel Coding. In *Proceedings of the 36th International Conference on Machine Learning*, 1182–1192. PMLR.

Fu, Q.; Xie, H.; Qin, Z.; Slabaugh, G.; and Tao, X. 2023. Vector Quantized Semantic Communication System. *IEEE Wireless Communications Letters*, 12(6): 982–986.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; and Lempitsky, V. 2016. Domain-adversarial Training of Neural Networks. *Journal of Machine Learning Research*, 17(59): 1–35.

Garcia Marti, D.; Palacios Beltrán, J.; Lacruz, J. O.; and Widmer, J. 2020. A mixture density channel model for deep learning-based wireless physical layer design. In *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 53–62.

García Martí, D.; Palacios Beltrán, J.; Lacruz, J. O.; and Widmer, J. 2020. A Mixture Density Channel Model for Deep Learning-Based Wireless Physical Layer Design. In *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 53–62. New York, NY, USA: Association for Computing Machinery. ISBN 9781450381178.

Goldsmith, A. 2005. *Wireless Communications*. New York: Cambridge University Press.

Hu, Q.; Zhang, G.; Qin, Z.; Cai, Y.; Yu, G.; and Li, G. Y. 2023. Robust Semantic Communications with Masked VQ-VAE Enabled Codebook. *IEEE Transactions on Wireless Communications*, 22(12): 8707–8722.

Li, S.; Wang, Y.; Guo, S.; and Feng, C. 2024. Task-Oriented Communication for Graph Data: A Graph Information Bottleneck Approach. *IEEE Transactions on Cognitive Communications and Networking*, 1–1.

Motiian, S.; Jones, Q.; Iranmanesh, S.; and Doretto, G. 2017. Few-Shot Adversarial Domain Adaptation. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Mylonas, G.; Kalogeras, A.; Kalogeras, G.; Anagnostopoulos, C.; Alexakos, C.; and Muñoz, L. 2021. Digital Twins From Smart Manufacturing to Smart Cities: A Survey. *IEEE Access*, 9: 143222–143249.

Plataniotis, K. N.; and Hatzinakos, D. 2017. Gaussian Mixtures and Their Applications to Signal Processing. *Advanced Signal Processing Handbook*, 89–124.

Raghuram, J.; Zeng, Y.; Martí, D. G.; Ortiz, R. R.; Jha, S.; Widmer, J.; and Banerjee, S. 2021. Few-shot Domain Adaptation for End-to-End Communication. *arXiv preprint arXiv:2108.00874*.

van den Oord, A.; Vinyals, O.; and kavukcuoglu, k. 2017. Neural Discrete Representation Learning. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Wang, B.; Tao, F.; Fang, X.; Liu, C.; Liu, Y.; and Freiheit, T. 2021. Smart Manufacturing and Intelligent Manufacturing: A Comparative Review. *Engineering*, 7(6): 738–757.

Wang, L.; Wu, W.; Zhou, F.; Yang, Z.; Qin, Z.; and Wu, Q. 2024a. Adaptive Resource Allocation for Semantic Communication Networks. *IEEE Transactions on Communications*, 72(11): 6900–6916.

Wang, Y.; Guo, S.; Deng, Y.; Zhang, H.; and Fang, Y. 2024b. Privacy-Preserving Task-Oriented Semantic Communications Against Model Inversion Attacks. *IEEE Transactions on Wireless Communications*, 23(8): 10150–10165.

Xie, H.; Qin, Z.; and Li, G. Y. 2022. Task-Oriented Multi-User Semantic Communications for VQA. *IEEE Wireless Communications Letters*, 11(3): 553–557.

Zhang, A.; and Guo, S. 2024. Learning Multi-Rate Task-Oriented Communications Over Symmetric Discrete Memoryless Channels. *IEEE Communications Letters*, 28(10): 2303–2307.

Zhang, A.; Wang, Y.; and et al. 2024. On the Utility-Informativeness-Security Trade-Off in Discrete Task-Oriented Semantic Communication. *IEEE Communications Letters*, 28(6): 1298–1302.

Zhang, J.; Zhao, C.; Ni, B.; and et al. 2019. Variational Few-shot Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, 1685–1694.