

# A Theoretically-Grounded Codebook for Digital Semantic Communications

Lingyi Wang\*, Rashed Shelim\*, Walid Saad\*, and Naren Ramakrishnan†

\*Bradley Department of Electrical and Computer Engineering, Virginia Tech, Alexandria, VA, 22305, USA

†Department of Computer Science, Virginia Tech, Alexandria, VA, 22305, USA

Emails: {lingyiwang, rasheds, walids, naren}@vt.edu

**Abstract**—The use of a learnable codebook provides an efficient way for semantic communications to map vector-based high-dimensional semantic features onto discrete symbol representations required in digital communication systems. In this paper, the problem of codebook-enabled quantization mapping for digital semantic communications is studied from the perspective of information theory. Particularly, a novel theoretically-grounded codebook design is proposed for jointly optimizing quantization efficiency, transmission efficiency, and robust performance. First, a formal equivalence is established between the one-to-many synonymous mapping defined in semantic information theory and the many-to-one quantization mapping based on the codebook’s Voronoi partitions. Then, the mutual information between semantic features and their quantized indices is derived in order to maximize semantic information carried by discrete indices. To realize the semantic maximum in practice, an entropy-regularized quantization loss based on empirical estimation is introduced for end-to-end codebook training. Next, the physical channel-induced semantic distortion and the optimal codebook size for semantic communications are characterized under bit-flip errors and semantic distortion. To mitigate the semantic distortion caused by physical channel noise, a novel channel-aware semantic distortion loss is proposed. Simulation results on image reconstruction tasks demonstrate the superior performance of the proposed theoretically-grounded codebook that achieves a 24.1% improvement in peak signal-to-noise ratio (PSNR) and a 46.5% improvement in learned perceptual image patch similarity (LPIPS) compared to the existing codebook designs when the signal-to-noise ratio (SNR) is 10 dB.

**Index Terms**—Digital semantic communication systems, codebook-enabled vector quantization, information theory.

## I. INTRODUCTION

Semantic communication is emerging as a promising communication paradigm that shifts the focus from conventional bit-accurate delivery to goal-driven, meaning-preserving transmission [1]–[3]. Particularly, in an end-to-end semantic communication framework, the transmitter typically uses deep neural networks to extract semantic features from raw data based on communication goals [3]. However, existing digital baseband hardware and protocol stacks support only discrete bitstreams for channel coding and modulation [4]. Hence, it is challenging for semantic communication systems to efficiently and reliably convert the high-dimensional continuous semantic features into discrete indices with preserved critical meaning [5]. To address this

challenge, learnable codebooks for vector quantization can be applied to discretize the semantic features without losing semantic capacity [6]–[8]. Specifically, the use of a codebook allows the semantic communication system to partition the semantic feature space into a collection of disjoint subspaces and label these subspaces with a finite integer index set. Then, each semantic feature vector can be quantized to a discrete index by finding its optimally associated subspace.

Recently, a number of works studied the problem of codebook-enabled semantic communications [4]–[12]. Several prior works in [4]–[7] primarily used a codebook for compact semantic discretization to reduce communication costs by quantization loss-based end-to-end training. In [8] and [10], the authors proposed unified codebooks for multi-task semantic communications by exploiting semantic relationships among tasks. The work in [11] investigated a multilevel codebook that used multi-head octonary quantization to further compress indices. The authors in [12] proposed the multi-codebook design with each codebook trained with different bit-flip probabilities for adaptive modulation. However, despite its success in addressing some meaningful problems, this prior work [4]–[12] is limited in a number of ways:

- There is a lack of a unified theoretical foundation on existing codebooks for semantic communications. Moreover, it is hard to determine the theoretically optimal codebook size.
- Existing codebook schemes solely optimize the quantization distortion, which leads to codeword underutilization or over-concentration.
- Existing codebook schemes cannot explicitly model the impact of physical-layer bit-flip errors on semantic representations, thus lacking channel-aware distortion metrics along with robustness optimization.

The main contribution of this paper is to overcome the aforementioned limitations of existing codebooks for digital semantic communications by rethinking the codebook-enabled vector quantization from a perspective of information theory. In particular, we propose a novel theoretically-grounded codebook design for better quantization efficiency, transmission efficiency, and robust performance. First, we demonstrate that the critical one-to-many synonymous mapping defined in semantic information theory can be realized

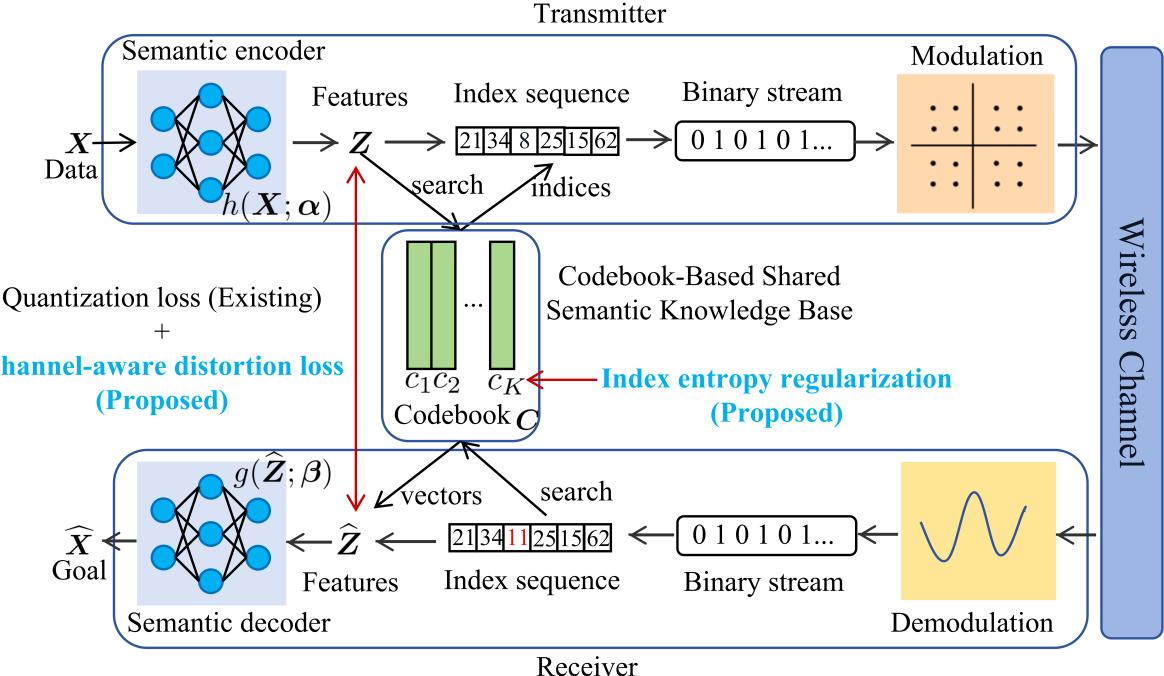


Fig. 1. Illustration of a codebook-enabled digital semantic communication framework.

through the many-to-one quantization mapping over the codebook's Voronoi partitions. Our perspective provides an unified theoretical-engineering model for codebook-enabled semantic communications. Then, we derive the mutual information between semantic features and their quantized indices. To maximize the semantic information carried by discrete indices, we propose an entropy-regularized quantization loss with empirical estimation to ensure balanced codeword utilization. Moreover, we characterize the channel-induced semantic distortion under bit-flip errors and the optimal distortion-aware codebook size. Driven by the channel-induced semantic distortion, we propose a novel channel-aware distortion loss to address semantic distortion induced by bit-flip errors. Extensive simulations on an image reconstruction task demonstrate that the proposed theoretically-grounded codebook achieves a 24.1% improvement in peak signal-to-noise ratio (PSNR) and a 46.5% improvement in learned perceptual image patch similarity (LPIPS) compared to the existing codebook designs when the signal-to-noise ratio (SNR) is 10 dB.

The rest of this paper is organized as follows. In Section II, we present the codebook-enabled digital semantic communications. Then, we propose a novel theoretically-grounded codebook in Section III. In Section IV, we demonstrate the simulation results and analysis. Finally, conclusions are drawn in Section V.

## II. CODEBOOK FOR DIGITAL SEMANTIC COMMUNICATIONS

As shown in Fig. 1, we consider a codebook-enabled digital semantic communication network, in which, without

loss of generality, the source data consists of images. Let the image dataset be  $\mathcal{X}$ , with each image  $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^{H \times W \times O}$ , where  $H$ ,  $W$  and  $O$  are the height, width, and the channel size of each input image, respectively. The learnable codebook for vector quantization is given by  $\mathbf{C} \triangleq [c_1, c_2, \dots, c_k, \dots, c_K] \subset \mathbb{R}^N$ , where  $K$  is the codebook size, and each element  $c_k \in \mathbb{R}^N$  is called a *semantic codeword* with dimension size  $N$  and index  $k$ . The codebook  $\mathbf{C}$  serves as a shared, constant knowledge base for the transmitter and the receiver.

At the transmitter, the semantic encoder  $h(\mathbf{X}; \alpha)$  with parameters  $\alpha$  extracts semantic features  $\mathbf{Z} \in \mathbb{R}^{M \times N}$  from the raw image  $\mathbf{X}$ . The learnable codebook  $\mathbf{C}$  then serves to map vector-based semantic representations onto discrete symbol representations in the digital communication system. Particularly, each semantic vector  $\mathbf{z}_m \in \mathbf{Z} \subset \mathbb{R}^N$  with index  $m \leq M$  is quantized by finding its nearest semantic codeword  $\mathbf{c}_k \in \mathbf{C}$ . The *quantization mapping*  $q(\cdot)$  by the nearest-neighbor algorithm in the codebook-enabled semantic space will be

$$q(\mathbf{z}_m) = \arg \min_{k \in \{1, \dots, K\}} \|\mathbf{z}_m - \mathbf{c}_k\|_2. \quad (1)$$

Then, with each vector  $\mathbf{z}_m \in \mathbf{Z}$  mapped to an index by (1), the index sequence of semantic features  $\mathbf{Z}$  is obtained by

$$\mathbf{S} = [s_m]_{m=1}^M \in \{1, \dots, K\}^M, \quad s_m = q(\mathbf{z}_m). \quad (2)$$

The codebook-enabled vector quantization discretizes the original floating-point semantics into compact integer indices, which stand in for the continuous feature. The resulting index sequence is then processed by standard digital communication blocks, including channel coding, mapping of constel-

lation symbols, modulation over carriers, and transmission over the physical channel.

The preimage of mapping  $q$  is represented by

$$\begin{aligned} q^{-1}(k) &= \{\mathbf{z} \in \mathbb{R}^N : q(\mathbf{z}) = k\} \\ &= \{\mathbf{z} \in \mathbb{R}^N : \|\mathbf{z} - \mathbf{c}_k\|_2 \leq \|\mathbf{z} - \mathbf{c}_j\|_2, \forall j \neq k\}, \end{aligned} \quad (3)$$

which is a Voronoi region of  $\mathbb{R}^N$  including infinite distinct semantic vectors. This Voronoi region can be considered as a *semantic feature cluster*, which can be obtained by

$$\mathbb{R}^N = \bigcup_{k=1}^K q^{-1}(k), \quad q^{-1}(k) \cap q^{-1}(\ell) = \emptyset \quad (k \neq \ell). \quad (4)$$

(3) and (4) partition the semantic feature space into disjoint Voronoi regions, thus enabling quantization by mapping each feature vector to a unique representative codeword. Hence, the codebook induces an equivalence-class partition  $\mathcal{Z}/q = \{q^{-1}(1), \dots, q^{-1}(K)\}$  of the continuous semantic space. The receiver demodulates and channel-decodes the received symbols to recover the integer index sequence  $\hat{\mathcal{S}} = [\hat{s}_m]_{m=1}^M$  and recovers the semantic information  $\hat{\mathbf{z}}_m = \mathbf{c}_{\hat{s}_m} = q^{-1}(\hat{s}_m)$ . The recovered semantic features  $\hat{\mathbf{Z}} \in \mathbb{R}^{M \times N}$  are input into the semantic decoder  $g(\cdot; \beta)$  with parameters  $\beta$  to accomplish the semantic task by  $\hat{\mathbf{X}} = g(\hat{\mathbf{Z}}; \beta)$ . To train the learnable codebook, the quantization loss [4]–[12] is minimized by

$$\mathcal{L}^{\text{qua}} = \mathbb{E}_{\mathbf{Z} \sim p_{\mathbf{Z}}} \left[ \left\| \sum \mathbf{z} - \mathbf{c}_{q(\mathbf{z})} \right\|^2 \right]. \quad (5)$$

Although (5) can measure the semantic distortion caused by the quantization mapping, it cannot ensure the maximum semantic information carried by each index, s.e., efficient semantic codeword design, and explicitly evaluate the semantic distortion from the physical channel noise.

### III. THEORETICALLY-GROUNDED CODEBOOK

In this section, we first rethink the quantization mapping of the codebook from a perspective of semantic information theory, as shown in Fig. 2. Then, a theoretically-grounded codebook is proposed with the entropy regularization for the maximum mutual information between semantic features and their quantized discrete indices, along with a novel channel-aware semantic distortion loss to alleviate the impact of physical channel noise.

#### A. Relationships Between Synonymous Mapping And Quantization Mapping

Semantic information theory [13] is based on a consensus that there is a one-to-many synonymous mapping  $f$  between the semantic source  $\tilde{U}$  and the syntactic source  $U$  to generate the raw message, as shown in Fig. 2. For instance, the semantic concept of “happiness” can be represented by various syntactic realizations such as text information “joy” and smiling images. Let the syntactic set be  $U = \{u_1, \dots, u_j, \dots, u_J\}$  and the semantic set be

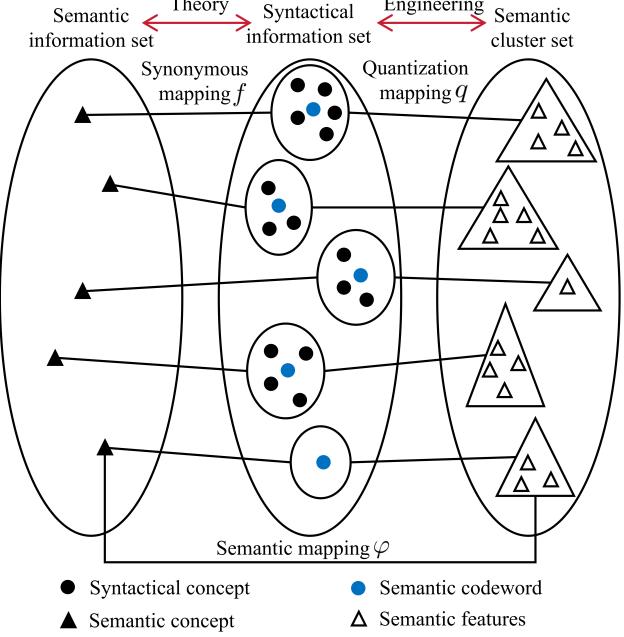


Fig. 2. A showcase of the unified synonymous mapping and quantization mapping from a joint perspective of theory and engineering.

$\tilde{U} = \{\tilde{u}_1, \dots, \tilde{u}_{\tilde{j}}, \dots, \tilde{u}_{\tilde{J}}\}$ ,  $\tilde{J} \ll J$ , where  $u_j$  represents a syntactic realization and  $\tilde{u}_{\tilde{j}}$  represents a semantic concept. Then, the synonymous mapping is given by  $f : \tilde{U} \rightarrow U$ . The synonymous mapping  $f$  induces an equivalence-class partition  $U/f = \{f(\tilde{u}_1), \dots, f(\tilde{u}_{\tilde{j}}), \dots, f(\tilde{u}_{\tilde{J}})\}$ , which can be obtained by

$$U = \bigcup_{\tilde{j}=1}^{\tilde{J}} f(\tilde{u}_{\tilde{j}}), \quad U_f(\tilde{u}_{\tilde{j}}) \cap U_f(\tilde{u}_{\tilde{m}}) = \emptyset \quad (\tilde{u}_{\tilde{j}} \neq \tilde{u}_{\tilde{m}}). \quad (6)$$

In practice, the learnable codebook realizes the synonymous mapping  $f$  by the quantization mapping  $q$ . In particular, as discussed in (1)–(4), the learnable codebook  $C \triangleq [c_1, c_2, \dots, c_k, \dots, c_K] \subset \mathbb{R}^N$ , partitions the continuous semantic feature space  $\mathbb{R}^N$  into  $K$  Voronoi cells, where each semantic codeword  $c_k$  provides a *standard syntax* for a set of synonymous syntaxes  $f(\tilde{u}_{\tilde{j}})$ . Since the semantic information is embedded in and cannot be directly separated from its syntactic representations, the semantic encoder is applied to infer a latent meaning representation with semantic features  $\mathbf{z}$  from the observed data  $\mathbf{X}$ . However, in practice, the same semantic concept can produce a semantic feature cluster, represented by  $q^{-1}(k)$ , due to morphological variability and observation noise. This reflects the fact that a single meaning can manifest through diverse syntactic expressions. During quantization mapping, regardless of its low-level syntactic variations, any semantic feature vector in the Voronoi region of  $c_k$  is assigned the same integer index  $k$ . In this way, the codebook-enabled vector quantization acts as the many-to-one mapping from syntactic expressions to a single semantic

symbol. Let the one-to-many semantic mapping between semantic concept  $\tilde{u}$  and semantic features  $z$  be  $\varphi : \tilde{U} \rightarrow \mathbb{R}^N$ . The semantic mapping  $\varphi$  means that one semantic concept can behave different features with different contexts, and coding noise from source data. The  $q \circ \varphi$  and  $f$  coincide on every semantic concept as

$$f(\tilde{u}) = U_k \iff \varphi(\tilde{u}) \in q^{-1}(k) \iff q(\varphi(\tilde{u})) = k. \quad (7)$$

Hence, from the perspective of information theory, the codebook can preserve the theoretical advantages of synonymous diversity with Voronoi regions partitioned by the quantization mapping, and enhances the transmission efficiency by providing discrete semantic representations that can be directly deployed in existing digital communication systems.

### B. Entropy Regularized Indices

Now, we derive the entropy regularization based on mutual information between semantic features and their quantized indices to maximize semantic information carried by discrete indices. Let  $p_{\mathbf{Z}}(z)$  be the density of the continuous semantic feature vector  $z \in \mathbb{R}^N$ , and then we have  $p_{\mathbf{Z}, \mathbf{S}}(z, k) = p_{\mathbf{Z}}(z) \mathbb{I}\{q(z) = k\}$ , where  $\mathbb{I}(x)$  is a binary-valued indicator function that equals to 1 if the condition  $x$  holds true and 0 otherwise. By definition of information theory, the mutual information between semantic features and their quantized indices is

$$\begin{aligned} s(\mathbf{Z}; \mathbf{S}) &= \sum_{k=1}^K \int p_{\mathbf{Z}, \mathbf{S}}(z, k) \log \frac{p_{\mathbf{Z}, \mathbf{S}}(z, k)}{p_{\mathbf{Z}}(z) p_{\mathbf{S}}(k)} dz \\ &= H(\mathbf{S}) - H(\mathbf{S} | \mathbf{Z}) = H(\mathbf{S}), \end{aligned} \quad (8)$$

where  $H(\mathbf{S} | \mathbf{Z}) = 0$  since the index sequence  $\mathbf{S}$  is determined given semantic features  $\mathbf{Z}$  by the many-to-one quantization mapping  $q$ . Let the occurrence probability of each index be  $\pi_k = \Pr[q(z) = k]$ . Equivalently, the coding rate can be rewritten as

$$\begin{aligned} s(\mathbf{Z}; \mathbf{S}) &= \int p_{\mathbf{Z}}(z) \log \frac{p(s = q(z) | z)}{\pi_{q(z)}} dz \\ &= \int p_{\mathbf{Z}}(z) \log \frac{1}{\pi_{q(z)}} dz = -\mathbb{E}[\log \pi_{\mathbf{S}}] \leq \log K, \end{aligned} \quad (9)$$

where the equality holds when  $\pi_1 = \pi_2 = \dots = \pi_K = 1/K$ . Hence, to increase the effective semantic information carried by the discrete indices and improve the transmission efficiency, we need to ensure every codeword is equitably accessible, which can be realized by introducing the entropy regularization of indices. However, in practice, the actual distribution  $p_{\mathbf{Z}}(z)$  is unknown, and, thus, it is hard to directly calculate the entropy  $H(\mathbf{S})$  for end-to-end codebook training. Here, we utilize the batch samples  $\mathbf{Z} \sim p_{\mathbf{Z}}$  to do empirical estimation, and the estimated entropy  $\hat{H}(\mathbf{S})$  of indices can be represented by

$$\hat{H}(\mathbf{S}) = -\sum_{k=1}^K \hat{\pi}_k \log \hat{\pi}_k, \quad (10)$$

with

$$\hat{\pi}_k = \mathbb{E}_{\mathbf{Z} \sim p_{\mathbf{Z}}} \left[ \sum_{z \in \mathbf{Z}} \mathbb{I}\{q(z) = k\} \right]. \quad (11)$$

Then, the estimated mutual information between semantic features and their quantized indices can be rewritten as

$$\hat{s}(\mathbf{Z}; \mathbf{S}) = \hat{H}(\mathbf{S}) = -\mathbb{E}_{\mathbf{Z} \sim p_{\mathbf{Z}}} \left[ \sum \log \hat{\pi}_{q(z)} \right]. \quad (12)$$

Hence, the quantization loss with entropy regularization for end-to-end codebook training can be represented by

$$\mathcal{L}^{\text{reg}} = \mathbb{E}_{\mathbf{Z} \sim p_{\mathbf{Z}}} \left[ \left\| \sum \mathbf{z} - \mathbf{c}_{q(z)} \right\|^2 \right] - \gamma \hat{H}(\mathbf{S}), \quad (13)$$

where  $\gamma > 0$  is a parameter that captures the tradeoff between quantization fidelity and index-entropy. The gradient with respect to the empirical frequencies  $\hat{\pi}_k$  is

$$\frac{\partial(-\hat{H}(\mathbf{S}))}{\partial \hat{\pi}_k} = 1 + \log \hat{\pi}_k, \quad (14)$$

which drives the occurrence probability  $\hat{\pi}_k$  of each index toward the uniform distribution  $1/K$ .

### C. Physical Channel-Induced Semantic Distortion

Now, we explore the semantic distortion over the physical channel. In particular, each index  $s \in \{1, \dots, K\}$  is encoded into a  $L$ -length binary sequence  $b \in \{0, 1\}^L$ , where  $L = \lceil \log_2 K \rceil$ , and  $\lceil x \rceil$  returns the smallest integer  $\geq x$ . We consider the indices are transmitted over a memoryless binary symmetric channel with bit-flip probability  $p$ . Let  $\hat{b}$  be the received binary vector, and  $d_H(b, \hat{b})$  be Hamming distance. Then, the error-weight distribution of the binary index sequence can be obtained by

$$\Pr(d_H(b, \hat{b}) = w) = \binom{L}{w} p^w (1-p)^{L-w}, w \leq L, \quad (15)$$

and the overall index-error probability can be represented by

$$P_e = \Pr(\hat{b} \neq b) = 1 - \Pr(d_H(b, \hat{b}) = 0) = 1 - (1-p)^L. \quad (16)$$

The conditional probability with random labeling can be approximately represented by

$$\Pr(\hat{s} = \ell | s = k) = \begin{cases} 1 - P_e, & \ell = k, \\ \frac{P_e}{K-1}, & \ell \neq k. \end{cases} \quad (17)$$

Then, the physical channel-induced semantic distortion based on average squared-error is obtained by

$$\begin{aligned} D_{\text{ch}} &= \sum_{k=1}^K \pi_k \left[ (1 - P_e) \cdot 0 + \sum_{\ell \neq k} \frac{P_e}{K-1} \|\mathbf{c}_k - \mathbf{c}_{\ell}\|_2^2 \right] \\ &= P_e \sum_{k=1}^K \pi_k \bar{\Delta}_k^2, \end{aligned} \quad (18)$$

with

$$\bar{\Delta}_k^2 = \frac{1}{K-1} \sum_{k \neq \ell} \|\mathbf{c}_k - \mathbf{c}_{\ell}\|_2^2. \quad (19)$$

Thus, the total semantic distortion induced by the physical channel noise can be expressed as

$$D_S(K, p) = \underbrace{\mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} \left[ \left\| \sum \mathbf{z} - \mathbf{c}_{q(\mathbf{z})} \right\|^2 \right]}_{\text{quantization loss}} + D_{\text{ch}}(K, p). \quad (20)$$

In traditional communication systems, a codebook is designed to map fixed-length bit sequences to modulation symbols, typically by maximizing minimum Euclidean distance or optimizing channel capacity under bit-error probability  $p$ . However, in the semantic communication framework, the codebook is constructed to quantize high-dimensional features to accurately convey the semantic information and directly minimize semantic distortion  $D_S(K, p)$  rather than bit-level errors. In this context, the optimal codebook size  $K^*$  is therefore chosen to balance semantic fidelity against bandwidth cost. Hence, given the bit-error probability  $p$  and the bandwidth weighting factor  $\lambda$ , the optimal codebook size for semantic communications can be selected by

$$K^* = \arg \min_K \{D_S(K, p) + \lambda R(K)\}, \quad (21)$$

where  $R(K) = M \log_2 K$  is the total bit-rate.

In addition to the selection of the optimal codebook size, the conventional codebooks [4]–[7] for semantic communications only minimize the quantization error and cannot capture the impact of physical-layer bit flips on digital communication. Thus, existing semantic codebooks experience severe semantic distortion or even total information loss under high error-rate conditions. Driven by channel-induced semantic distortion (13), we further introduce the generalized channel-aware distortion loss, which can be represented by

$$\mathcal{L}^{\text{ch}} = \mathbb{E}_p [D_{\text{ch}}(K, p)] = \mathbb{E}_p \left[ P_e \sum_{k=1}^K \pi_k \bar{\Delta}_k^2 \right], \quad (22)$$

which imposes structured constraints on the codebook to enhance the robustness of semantic communications. The total loss function for the codebook can be represented by

$$\mathcal{L}^C = \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} \left[ \left\| \sum \mathbf{z} - \mathbf{c}_{q(\mathbf{z})} \right\|^2 \right] + \omega \mathcal{L}_{\text{ch}} - \gamma \hat{H}(\mathbf{S}), \quad (23)$$

where  $\omega \in (0, 1)$  is the weight factor for the channel-aware semantic distortion.

Fundamentally, the proposed entropy regularization and the channel-aware semantic distortion loss enable the codebook to find an optimal space partition scheme. In particular, the entropy regularization of discrete indices encourages more fair partitions of the semantic feature space based on the occurrence probability of each index, thus maximizing the semantic information carried by each index. The channel-induced distortion loss reduces the spatial distance between the original features and distorted quantization based on the possibility of physical-layer bit-flip errors, which alleviates semantic distortion and enhances semantic robustness.

## IV. SIMULATION RESULTS AND ANALYSIS

### A. Simulation Setup

For our simulations, an image semantic communication network with the reconstruction task is considered with a single-antenna transmitter and a single-antenna receiver, where the semantic encoder and decoder are based on vector quantized-variational autoencoder (VQ-VAE) [14] with the same training parameters provided in [8]. The end-to-end training loss of the semantic network for image reconstruction can be represented by

$$\mathcal{L}(\alpha, \beta, C) = \underbrace{\left\| G_{\ell}^j(\mathbf{X}) - G_{\ell}^j(\widehat{\mathbf{X}}) \right\|_2^2}_{\text{reconstruction loss}} + \mathcal{L}^C. \quad (24)$$

The reconstruction loss is based on the pretrained VGG16 network [15] to capture the semantic similarity in the latent space, where  $G_{\ell}^j$  is the Gramm matrix, represented by

$$G_{\ell}^j(\mathbf{X}) = \frac{1}{H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \ell_j(\mathbf{X})_{o,h,w} \ell_j(\mathbf{X})_{o',h,w}, \quad (25)$$

and  $\ell_j(\cdot)$  represents the first  $j$  layers of VGG16 with parameter  $\ell$ ,  $o \leq O_j$ . We set  $j = 8$ ,  $\omega = 0.1$  and  $\gamma = 0.1$ . The modulation scheme is set to 64-QAM [11], and the codebook size is set to  $K = 256$ . We adopt VOC-2012 as the training database, which consists of 11,530 images of 20 classes. We introduce the JPEG+LDPC scheme based on joint photographic experts group (JPEG) and 1/2-rate low-density parity-check (LDPC) code, the BPG+LDPC scheme based on better portable graphics (BPG) and the 1/2-rate LDPC code, and the VQ-VAE scheme [8] that only considers quantization loss as the benchmark schemes. LPIPS [16] and PSNR are used to evaluate the quality of reconstructed images.

Fig. 3 and Fig. 4 respectively show the LPIPS performance and the PSNR performance of different schemes across different SNRs in a Rayleigh fading channel. Compared to the existing VQ-VAE-based scheme, the proposed theoretically-grounded codebook achieves a 24.1% improvement in PSNR and a 46.5% improvement in LPIPS compared to the existing VQ-VAE method when SNR is 10 dB. These improvements can be attributed to the proposed index-entropy regularization and channel-aware semantic distortion loss, which is further demonstrated by the ablation experiments. In particular, we independently introduce the regularization item and channel-aware distortion loss to the VQ-VAE-based schemes, respectively named “VQ-VAE + Index Entropy” and “VQ-VAE + Channel-Aware”. Under low-SNR conditions with high bit-flip rates, the quantized indices of the semantic features can be distorted over the physical channel. In this context, the channel-aware semantic distortion loss can reduce the spatial distance between the original features from the transmitter and the distorted features at the receiver based on distortion probability. The distortion loss suppresses the semantic distortion caused by error bits, thus enhancing the semantic robustness. In the context of high SNRs with accurate symbol

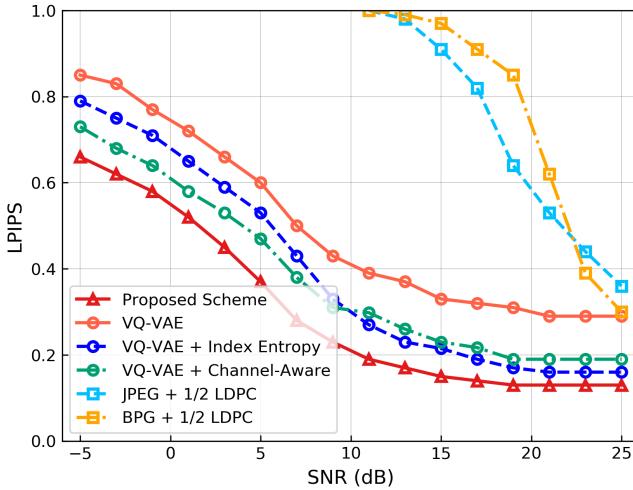


Fig. 3. The LPIPS performance of different schemes across different SNRs.

transmission, the entropy regularization of discrete indices encourages more fair partitions of the semantic feature space based on the occurrence probability of each index, and maximizes the information carried by each index, thus improving transmission efficiency. Hence, the channel-aware semantic distortion loss and entropy regularization can enable an efficient codebook design by the optimal space partitions.

## V. CONCLUSION

In this paper, we have rethought the codebook-enabled vector quantization for digital semantic communications through an information-theoretic perspective and introduced a novel, theoretically-grounded codebook design that jointly enhances quantization efficiency, transmission efficiency, and robustness. We have demonstrated that the one-to-many synonymous mapping of semantic information theory can be realized through the Voronoi regions partitioned by the quantization mapping, thus unifying abstracted semantic information theory with the codebook-based implementation. Building on this foundation, we have derived the mutual information between high-dimensional semantic features and their quantized discrete indices, and have introduced the index-entropy regularization. The entropy-regularized quantization loss has encouraged the balanced codeword utilization and maximized the semantic information carried by each index. We have further analyzed the impact of physical-layer bit-flip errors on semantic distortion, proposed the channel-aware semantic distortion loss for robust performance, and presented the optimal codebook size under error-rate and bandwidth constraints. Extensive simulations on image reconstruction tasks have demonstrated that the proposed theoretically-grounded codebook achieves 24.1% improvement in PSNR and 46.5% improvement in LPIPS compared to the existing VQ-VAE designs when SNR is 10 dB.

## REFERENCES

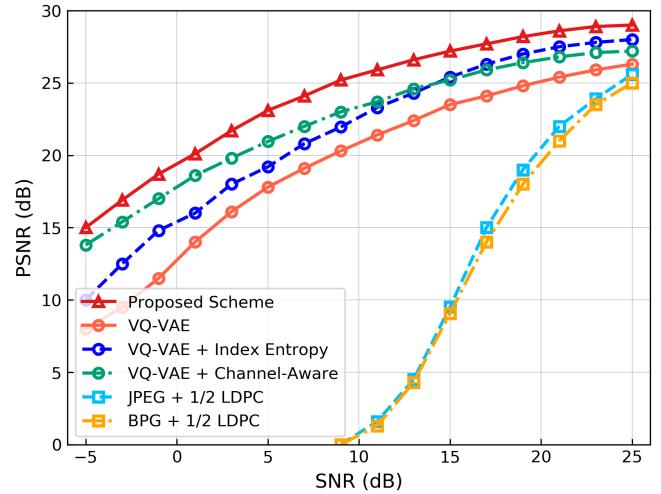


Fig. 4. The PSNR performance of different schemes across different SNRs.

- [1] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. Vincent Poor, “Less data, more knowledge: Building next-generation semantic communication networks,” *IEEE Commun. Surveys Tuts.*, vol. 27, no. 1, pp. 37–76, 2025.
- [2] W. Saad, O. Hashash, C. K. Thomas, C. Chaccour, M. Debbah, N. Mandayam, and Z. Han, “Artificial general intelligence (AGI)-native wireless systems: A journey beyond 6G,” *Proceedings of the IEEE*, pp. 1–39, 2025, to appear.
- [3] L. Wang, W. Wu, F. Zhou, Z. Yang, Z. Qin, and Q. Wu, “Adaptive resource allocation for semantic communication networks,” *IEEE Trans. Commun.*, vol. 72, no. 11, pp. 6900–6916, 2024.
- [4] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, “Robust semantic communications with masked vq-vae enabled codebook,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 8707–8722, 2023.
- [5] Y. Huh, H. Seo, and W. Choi, “Universal joint source-channel coding for modulation-agnostic semantic communication,” *IEEE Journal on Selected Areas in Communications*, 2025, to appear.
- [6] K. Ye, M. Gong, S. Wang, and D. Feng, “Low-rate semantic communication with codebook-based conditional generative models,” *arXiv preprint arXiv:2504.04977*, 2025.
- [7] H. Zhang, M. Tao, Y. Sun, and K. B. Letaief, “Improving learning-based semantic coding efficiency for image transmission via shared semantic-aware codebook,” *IEEE Trans. Commun.*, vol. 73, no. 2, pp. 1217–1232, 2025.
- [8] L. Wang, W. Wu, F. Zhou, F. Tian, Q. Wu, and W. Saad, “A unified hierarchical semantic knowledge base for multi-task semantic communication,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2024, pp. 2937–2943.
- [9] Y. Li, Z. Jin, T. Song, X. Song, and J. Hu, “Robust task-oriented communication with semantic-aware masking and discrete codebook,” in *Proc. IEEE Wirel. Commun. Netw. Conf. (WCNC)*, 2025, pp. 1–6.
- [10] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, and X. Tao, “A unified multi-task semantic communication system for multimodal data,” *IEEE Trans. Commun.*, vol. 72, no. 7, pp. 4101–4116, 2024.
- [11] Y. Zhou, Y. Sun, G. Chen, X. Xu, H. Chen, B. Huang, S. Cui, and P. Zhang, “Moc-rvq: Multilevel codebook-assisted digital generative semantic communication,” *arXiv preprint arXiv:2401.01272*, 2024.
- [12] J. Shin, Y. Oh, J. Park, J. Park, and Y.-S. Jeon, “Esc-mvq: End-to-end semantic communication with multi-codebook vector quantization,” *arXiv preprint arXiv:2504.11709*, 2025.
- [13] K. Niu and P. Zhang, *The Mathematical Theory of Semantic Communication*. Springer Nature, 2025.
- [14] A. van den Oord, O. Vinyals, and k. kavukcuoglu, “Neural discrete representation learning,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30, 2017.
- [15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2015.
- [16] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2018, pp. 586–595.