

# Topology Data Analysis-Based Error Detection for Semantic Image Transmission with Incremental Knowledge-Based HARQ

Ni Fei<sup>1</sup>, Li Rongpeng<sup>1,\*</sup>, Zhao Zhifeng<sup>2</sup>, Zhang Honggang<sup>3</sup>

<sup>1</sup> College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

<sup>2</sup> Zhejiang Lab, Hangzhou 311121, China

<sup>3</sup> Faculty of Data Science, City University of Macau, Macao, China

\* The corresponding author, email: lirongpeng@zju.edu.cn

**Cite as:** F. Ni, R. Li, *et al.*, “Topology data analysis-based error detection for semantic image transmission with incremental knowledge-based harq,” *China Communications*, 2025, vol. 22, no. 1, pp. 235-255. DOI: 10.23919/JCC.fa.2023-0799.202501

**Abstract:** Semantic communication (SemCom) aims to achieve high-fidelity information delivery under low communication consumption by only guaranteeing semantic accuracy. Nevertheless, semantic communication still suffers from unexpected channel volatility and thus developing a re-transmission mechanism (e.g., hybrid automatic repeat request [HARQ]) becomes indispensable. In that regard, instead of discarding previously transmitted information, the incremental knowledge-based HARQ (IK-HARQ) is deemed as a more effective mechanism that could sufficiently utilize the information semantics. However, considering the possible existence of semantic ambiguity in image transmission, a simple bit-level cyclic redundancy check (CRC) might compromise the performance of IK-HARQ. Therefore, there emerges a strong incentive to revolutionize the CRC mechanism, thus more effectively reaping the benefits of both SemCom and HARQ. In this paper, built on top of swin transformer-based joint source-channel coding (JSCC) and IK-HARQ, we propose a semantic image transmission framework SC-TDA-HARQ. In particular, different from the conventional CRC, we introduce

a topological data analysis (TDA)-based error detection method, which capably digs out the inner topological and geometric information of images, to capture semantic information and determine the necessity for re-transmission. Extensive numerical results validate the effectiveness and efficiency of the proposed SC-TDA-HARQ framework, especially under the limited bandwidth condition, and manifest the superiority of TDA-based error detection method in image transmission.

**Keywords:** error detection; incremental knowledge-based HARQ; joint source-channel coding; semantic communication; swin transformer; topological data analysis

## I. INTRODUCTION

Recently, semantic communication (SemCom), which focuses on transmitting meaningful information rather than precise bits or symbols, emerges as a feasible approach to achieve high fidelity information delivery under lower communication consumption [1–4]. To achieve superior transmission performance and outstanding channel environmental adaptability, some infant literature adopts joint source-channel coding (JSCC) for the SemCom system. Benefiting from the booming development of deep learning (DL) models

Received: Dec. 27, 2023

Revised: Feb. 22, 2024

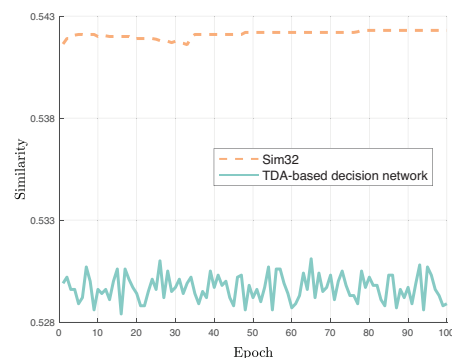
Editor: Yang Kun

in feature extraction and reconstruction, SemCom systems with DL-based JSCC (Deep JSCC) and its extended frameworks are eligible for the transmission of various kinds of data, such as image [2, 5–13], speech [14], text [15–17] and video [18, 19], and outperform the traditional design, especially under limited channel bandwidth and low signal-to-noise ratio (SNR). Considering the promising market prospect of computer vision (CV) applications and the foundational role in video transmission, the image SemCom warrants further exploration particularly.

However, due to the inherent limitations of the traditional DL model’s capacity, a notable decline in the performance of Deep JSCC emerges as the image resolution increases. Meanwhile, as one of the latest variants of transformer architecture, swin transformer shows remarkable performances on many image datasets [20]. Besides, Ref. [6], which utilizes swin transformer as the backbone to realize semantic image transmission, demonstrates considerable performance gain. Motivated by these facts, we adopt swin transformer for the extraction of sophisticated semantic features from the source image, thereby facilitating a more effective method for wireless image transmission. However, most of the aforementioned frameworks discard useful information in unacknowledged received signals and merely employ a one-shot decoding [11]. Nevertheless, since the lossy information may still embody valuable semantic parts that can assist the receiver in decoding and improving transmission efficiency, the one-shot decoding misses part of the information in rounds of transmissions while reconstructing a high-quality image.

To combat this deficiency, an incremental knowledge-based hybrid automatic repeat request (IK-HARQ) is introduced for wireless transmission [16, 21] and plays an indispensable role in SemCom [17]. Specifically, apart from the similar error detection-based acknowledgement/negative acknowledgement (ACK/NAK) feedback to the HARQ in traditional transmission systems, IK-HARQ in SemCom additionally recycles the contaminated information as the valuable incremental knowledge to supplement re-transmission decoding, which further guarantees the accuracy of the received packets [22]. Notably, unlike the bit-level error detection standard (e.g., cyclic redundancy check [CRC] [23], Hamming Code [24]), IK-HARQ in SemCom is expected to

discriminate the errors from a semantic perspective, and the corresponding DL-based error detection modules for text transmission in Refs. [16] and [17] provide an auspicious start for the future research. However, Ref. [17] makes its re-transmission decision solely on channel conditions (i.e., the value of SNR), and lacks a comprehensive understanding of source and environment complexity. Meanwhile, Ref. [16] employs a Sim32 method to calculate semantic similarity at the receiver side and make re-transmission decisions according to a preset threshold. Albeit its effectiveness in semantic text transmission, when this approach is extended to image transmission<sup>1</sup>, the detected similarity tends to be overly stiff (i.e., a variance of  $3.1151 \times 10^{-8}$  only) as is illustrated in Figure 1 and Table 1, resulting in re-transmission decisions highly vulnerable to inaccurately configured thresholds. Therefore, boosting the detection robustness necessitates a novel error detection scheme by discovering intrinsic high-level patterns.



**Figure 1.** The semantic similarity of images detected by Sim32 and TDA-based decision network during the training procedure.

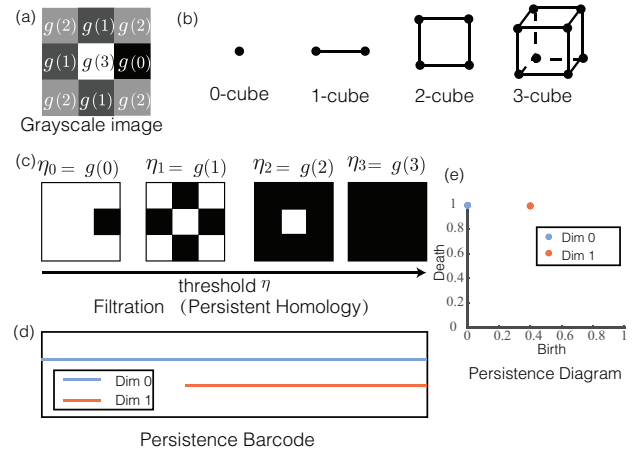
On the other hand, topological data analysis (TDA), which correlates with the inherent topological patterns of an image [25, 26], is introduced to extract intrinsic features from the source image, demonstrating robustness in error detection as shown in Figure 1 and Table 1. Specifically, TDA leverages topology and geometry to robustly infer qualitative and quantitative information about data structure. In other words, it produces recapitulative summaries or approximations with specific methods like *persistent homology* (PH) [27–31]. Taking the PH in Figure 2 as an example, a  $3 \times 3$  images  $x$  is constructed by pixels with different grayscale values  $g$ . Hence, as is shown in Figure

2(b),  $x$  can be constructed as the cubical complex  $K$  from a union of *multiple*  $d \in \{0, 1\}$ -cubes  $\xi$ , where 0-cube and 1-cube refer to vertices and edges respectively. Given a threshold  $\eta$ , when its grayscale value  $g \leq \eta$ , the corresponding pixel spans a 0-cube. Meanwhile, a 1-cube emerges when both adjacent pixels have  $g \leq \eta$ . Thus, different thresholds  $\eta$  correspond to different cubical complexes, indicating different topological features. In order to depict how  $q$ -dimensional topological features persist across different thresholds  $\eta_0 < \eta_1 < \eta_2 < \eta_3$ , the filtration of  $K$  is defined as the nested family of subspaces  $K(x, \eta_0) \subseteq K(x, \eta_1) \subseteq K(x, \eta_2) \subseteq K(x, \eta_3)$ . Along with the filtration, the topological features evolve, thus affecting the PH. As illustrated in Figure 2(c), 0-dimensional homology captures the connected components once  $\eta$  increases to  $g(0)$  while 1-dimensional homology characterizes a loop that lasts from  $\eta = g(1)$  to  $\eta = g(3)$ . As a result, Figure 2(d) records the corresponding PH as a well-defined set of disjoint half-open intervals known as a *persistence barcode* (PB), where each interval in dimension  $q \in \{0, 1\}$  symbolizes the lifespan  $[b^q, d^q]$  of a topological feature (i.e.,  $q$ -homology class) with the term of the birth time  $b^q$  and the death time  $d^q$  as the beginning and end of the interval of the  $q$ -homology. Alternatively, the same information can be represented through a *persistence diagram* (PD), as is shown in Figure 2(e), where intervals are depicted as points  $(b^q, d^q)$  in a birth-death coordinate. The extracted topological and geometric information from images competently describe a complete picture without dimensionality reduction [25], and thus potentially amplify the slight variations on core semantic information after noisy transmission. Inspired by these promising findings [27–31], there emerges a strong incentive to develop a TDA-based error detection mechanism and compensate for the deficiency of conventional error detection schemes, so as to fully ensure the gains of IK-HARQ. In this regard, preliminary results in Figure 1 and Table 1 validate the more competent capability of the TDA-based scheme.

In this paper, on top of IK-HARQ and the state-of-the-art semantic image encoder, swin transformer [20], we focus on developing a TDA-based error detection scheme, and correspondingly design a TDA-based semantic image transmission framework SC-TDA-HARQ. After highlighting key differences with highly related studies [17, 16, 19] in Table 2, SC-TDA-

**Table 1.** The Comparison of error detection results between Sim32 and proposed TDA-based decision network.

Method	Similarity	Mean	Variance
Sim32	0.5416 ~ 0.5423	0.5421	$3.1151 \times 10^{-8}$
TDA-based decision network	0.5284 ~ 0.5311	0.5297	$4.021 \times 10^{-7}$



**Figure 2.** An example of a filtered cubical complex's PH. (a) A  $3 \times 3$  grayscale image. (b) Different dimension of cubes that constitutes the cubical complex. (c) The filtration process of the image. (d) The corresponding persistence barcode. (e) The corresponding persistence diagram.

HARQ has the following three-folded merits.

- Inspired by the works IK-HARQ [17, 21], SC-TDA-HARQ can improve the image transmission efficiency with reduced semantic errors, by effectively leveraging incrementally transmitted knowledge.
- To further improve the efficiency of semantic image transmission, SC-TDA-HARQ integrates the swin transformer-based semantic coding with an additional TDA coding module to enhance image reconstruction at the receiver side. Therefore, it can exploit the advantages of the semantic architecture and TDA, and outperform competing semantics-based methods in terms of quality of reconstructed images.
- SC-TDA-HARQ incorporates a TDA-based decision network to determine re-transmissions, by more accurately evaluating the existence of semantic error in estimated images. In other words, on the condition of semantics consistency, SC-TDA-HARQ can tolerate the reconstruction incorrectness of a few pixels rather than

**Table 2.** *The comparison with highly related literature.*

Related works	Advantages	Limitations
Refs. [2, 5–8]	Superior to traditional systems under limited bandwidth and low SNR.	Contingent on one-shot decoding which discards useful information in unacknowledged signals.
Ref. [16]	Using IK-HARQ to combat the deficiency of one-shot decoding.	Stiff semantic similarity metrics for image re-transmission decisions.
Ref. [17]		Making re-transmission decisions based on SNR only and unable to deal with the complexity of delivered content.
Ours	Incorporating a TDA-based decision network to check semantic errors in received images and determine re-transmissions.	

re-transmission, promising to save more communication resources.

The rest of the paper is organized as follows. In Section II, related work is presented. The system model and the accompanied framework of SC-TDA-HARQ are introduced in Section III. In Section IV, we elaborate on the implementation details of SC-TDA-HARQ. The experimental results are demonstrated in Section V, which are followed by our conclusion in Section VI.

## II. RELATED WORKS

### 2.1 Deep Joint Source-Channel Coding

Conventionally, an image transmission system typically uses a separate source-dependent coding scheme (e.g., JPEG [32], JPEG2000 [33], better portable graphics (BPG) [34]) to remove the information redundancies. Subsequently, the transmitted bits undergo a source-independent channel coding scheme (e.g., low-density parity-check (LDPC) [35], Turbo code [36]) to safeguard from distortions introduced by the noisy communication channel. Despite the remarkable success and popularity, the independence between the source encoder and the channel encoder may incur restricted transmission performance and a significant reduction in the quality of reconstructed images [2]. Thus, JSCC is proposed to overcome the so-called “cliff effect” [2], and obtains graceful quality. The inchoate Deep JSCC architecture for image transmission, whose encoder and decoder are modeled by deep neural networks (DNNs) with excellent generalization ability, directly maps the pixel values to channel inputs and obtains a more reliable reconstruction of source images [2]. Afterwards, many variants of Deep JSCC emerge. For example, in order to combat channel fluctuations, Ref. [5] devises

an SNR-adaptive decoder, which uses the pilot signal sent by the transmitter to estimate the SNR and obtains preferable reconstruction quality. Furthermore, Ref. [6] proposes a spatial modulation module to scale the latent representations based on channel state information, thus enhancing the capability to deal with various channel conditions. In addition, inspired by compressed sensing, Ref. [7] focuses on reducing the number of measurements corrupted by mild noise and presents uncertainty-aware autoencoders. On the other hand, Ref. [8] proposes to combine the orthogonal frequency division multiplexing (OFDM) base-band processing blocks with the Deep JSCC framework, which is superior to traditional schemes and exhibits robustness under the conditions of mismatched channel models as well.

Notably, all the aforementioned works conduct a *one-shot decoding* method without further iterations. To verify the contribution and the potential of feedback, Ref. [9] designs practical JSCC schemes that can exploit noisy or noiseless channel feedback, and demonstrates considerable gains as well as excellent adaptability. Ref. [10] reconfirms the performance of transformer-aided Deep JSCC with feedback. On the contrary, Ref. [11] shows the reconstruction quality improvement can be attained by more iterations at the receiver on the basis of no modifications to the transmitter or requesting re-transmission. Nevertheless, the performance gains come at the expense of rapidly increased computational expenditure. In that regard, some feedback and/or re-transmissions sound more appealing.

### 2.2 HARQ for Wireless Communication System

As one of the channel feedback in conventional communication, HARQ can ensure the correctness of the



received packets by repetitively re-transmitting until receiving ACK, and is an essential part in a reliable transmission system [16]. Meanwhile, based on the feedback, adjustable rates can be attained [37, 38]. Until recently, only a few works notice the latent capability of HARQ in boosting the performance of Deep JSCC. Specifically, in Ref. [16], a network called Sim32 is introduced to detect the meaning error in the received sentences. Concurrently, inspired by Ref. [21], Ref. [17] pioneers the study of IK-HARQ by regarding previously transmitted lossy information as incremental knowledge to benefit the decoding of re-transmissions. Furthermore, in Ref. [19], a semantic video conferencing (SVC) network based on keypoint transmission is established wherein the semantic error detector exploits the fluency of the video to check the received frame. However, faced with the possibility of semantic ambiguity, a convincing error detection scheme remains under-investigated for images.

### 2.3 TDA and Its Combination with DL

Due to its versatility and robustness, TDA has been leveraged to solve various DL problems, after aligning the topological features with the vector input of DL. For instance, Refs. [39] and [40] introduce Bottleneck distance and Wasserstein distance to statistically compute the similarity of the PDs, thus measuring the topological differences of images. Alternatively, PB can be transformed into Betti curve, which is defined as the summation of all barcodes belonging to different dimensions and represents the number of topological features [25]. Persistence landscape (PL), which focuses on the importance and duration of features during the filtration process [26], belongs to another way to map the PD in a vector space. Besides, in line with the application of kernels in machine learning models, suitable kernel functions are proposed to deal with the topological data as well [41–43]. Apart from the aforementioned methods that provide the statistic treatment of PD, the topological features can be summarized in terms of persistent entropy (PE) which computes the Shannon entropy based on the probability distribution obtained from the given PD [44]. On this basis, all previous schemes contribute to significant performance gain in the classification and segmentation tasks of images [27, 45, 46]. Accordingly, it is promising to introduce TDA to SemCom as a

**Table 3.** Mainly used notations in this paper.

Notations	Description
$x \in \mathbb{R}^{H \times W \times 3}$	The source image with height $H$ and width $W$
$SC_{en}(\cdot), SC_{de}(\cdot)$	The source-channel encoder and decoder
$TDA_{en}(\cdot), TDA_{de}(\cdot)$	The TDA encoder and decoder
$x'$	The symbol encoded by $SC_{en}(\cdot)$
$x'_{TDA}$	The symbol encoded by $TDA_{en}(\cdot)$
$x''$	The symbol after compression by MLP MLP( $\cdot$ )
$y$	Transmitted symbol after power normalization
$\hat{y}$	The received signal
$\hat{x}_{TDA}$	The image reconstructed by $TDA_{de}(\cdot)$
$\hat{x}''$	Recovered symbol by MLP( $\cdot$ ) at the receiver
$\hat{x}$	The image reconstructed by $SC_{de}(\cdot)$
DEC( $\cdot$ )	The decision network that classifies $\hat{x}$ and $\hat{x}_{TDA}$
$m$	The source bandwidth
$R$	The bandwidth compression rate
$\alpha, \beta, \phi, \gamma$	Trainable parameters for $SC_{en}(\cdot), SC_{de}(\cdot), TDA_{de}(\cdot)$ and DEC( $\cdot$ ), respectively
$\zeta$	The ACK/NAK signal computed by DEC( $\cdot$ )
$(u, v)$	The position of the pixel in images
$g_B(u, v)$	The binarized function on $(u, v)$
$\mu_H(\cdot), \mu_R(\cdot)$	The height and radical filtration on $g_B(\cdot)$
$\psi, (u_c, v_c)$	A candidate direction and center for $\mu_H(\cdot)$ and $\mu_R(\cdot)$
$b, d$	Birth time and death time of topological features
$q$	The dimension of the topological features
$t$	The index of a $q$ -dimensional topological feature
AW( $p, q$ )	The $p$ -order Wasserstein amplitude calculated from $(b_t^q, d_t^q)$
AB( $q$ )	The Bottleneck distance calculated from $(b_t^q, d_t^q)$
$\varepsilon(\varsigma, q)$	The Betti curve of PB at a given threshold $\varsigma$
PL( $\lambda, \varrho, q$ )	The persistent landscape for layer $\lambda$ of PD with $\varrho \in \mathbb{N}$
HK( $\kappa, q, r, s$ )	The heat kernel of PD with respect to $(r, s) \in \mathbb{R}^2$
PE( $q$ )	The persistent entropy of PD
$\kappa$	The standard deviation for Gaussian distribution HK( $\cdot$ )

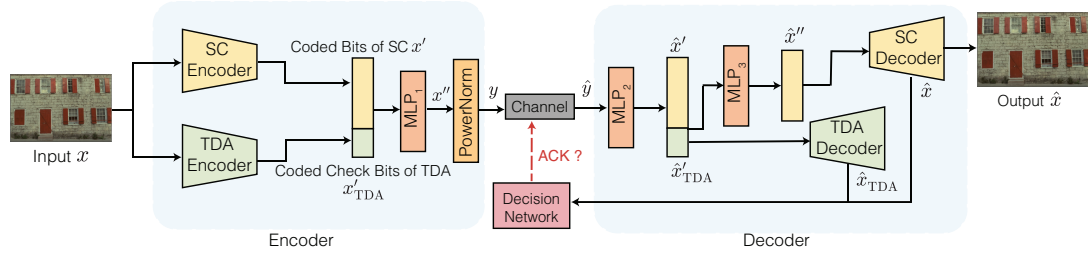
candidate solution to evaluate the credibility of reconstructed images.

## III. SYSTEM MODEL

In this section, we present the system model for semantic image transmission with SC-TDA-HARQ and enumerate key components therein, which are further shown in Figure 3. Beforehand, we summarize the mainly used notations in Table. 3.

### 3.1 Classical Model for SemCom

The fundamental SemCom system consists of sym-



**Figure 3.** The pipeline of the transformer-enabled SemCom system with IK-HARQ and TDA-based error detection.

metrical encoder and decoder in the transmitter and receiver. The source-channel (SC) encoder, denoted as  $\text{SC}_{\text{en}}(\cdot)$ , parameterized by  $\alpha$  extracts the semantic features of a 3-color (i.e., RGB) source image  $x \in \mathbb{R}^{H \times W \times 3}$  and embeds them into a semantic symbol  $x'$ , where  $H$  and  $W$  denote the height and width of an RGB image, respectively. Mathematically,

$$x' = \text{SC}_{\text{en}}(x; \alpha). \quad (1)$$

Then, a multilayer perception (MLP) module, i.e.,  $\text{MLP}_1(\cdot)$ , is used to compress the semantic symbol vector  $x'$  into transmitted symbol  $x'' \in \mathbb{C}^k$  with the assumption that  $k$  is smaller than the source bandwidth  $m := H \times W \times 3$ . Subsequently, a power normalization layer [2] is then applied to obtain the practically transmitted symbol  $y$ , that is,

$$y = \sqrt{k} \frac{x''}{\sqrt{(x'')^T x''}}, \quad (2)$$

so as to meet the average power constraint  $\mathbb{E}[y^T y] \leq k$ . Therefore, as the complex-valued transmission channel enables to transmit two real-valued elements per channel-use [11], the bandwidth compression rate can be written as

$$R = \frac{k}{2m}. \quad (3)$$

Afterwards,  $y$  undergoes the communication channel, which can be modeled as an additive white Gaussian noise (AWGN) channel or the Rayleigh fading channel. Correspondingly, the received signal  $\hat{y}$  at the decoder can be formulated as

$$\hat{y} = \begin{cases} h \cdot y + \theta, & \text{for Rayleigh fading channel;} \\ y + \theta, & \text{for AWGN channel,} \end{cases} \quad (4)$$

where  $\theta$  is the noise sampled from independent and identically distributed (i.i.d.) Gaussian distribution  $\mathcal{CN}(0, \sigma^2 \mathbf{I})$  with zero mean and variance  $\sigma^2$ . In addition, for the Rayleigh fading channel, the channel gain  $h$  follows Rayleigh distribution and can be regarded as a constant during a transmission slot. Therefore, the corresponding SNR that evaluates the channel quality can be formulated as  $10 \log_{10} \frac{\|h \cdot y\|^2}{\|\theta\|^2}$  (dB), while AWGN can be viewed as a special case of  $h = 1$ .

At the receiver, the channel output  $\hat{y}$  undergoes dual MLP modules, so as to obtain a semantic symbol vector  $\hat{x}''$ ,

$$\hat{x}'' = \text{MLP}_3(\hat{x}') = \text{MLP}_3(\text{MLP}_2(\hat{y})). \quad (5)$$

Afterwards, the semantic-channel decoder  $\text{SC}_{\text{de}}(\cdot)$  with parameters  $\beta$  tries to mitigate the physical noise for  $\hat{x}''$  and recovers the original image in a semantically accurate manner, that is,

$$\hat{x} = \text{SC}_{\text{de}}(\hat{x}''; \beta). \quad (6)$$

Correspondingly, as is shown in Figure 4(1), a loss function (e.g., mean-squared-error [MSE]) is leveraged to evaluate the distortion between  $x$  and  $\hat{x}$  and optimize the JSCC parameters  $\alpha$  and  $\beta$ ,

$$(\alpha, \beta) = \arg \min \mathcal{L}_{\text{MSE}}(x, \hat{x}), \quad (7)$$

where defined by an  $L_2$ -norm  $\|\cdot\|_2$ ,  $\mathcal{L}_{\text{MSE}}(z, \hat{z}) = \mathbb{E} \|\cdot\|_2$  accompanies with the expectation operator  $\mathbb{E}(\cdot)$ .

### 3.2 TDA-Enabled SemCom with IK-HARQ

Although Deep JSCC has shown effectiveness under low SNR or with the limited bandwidth [2], the

veracity of decoded message is still hard to be ensured, especially for the one-shot paradigm. Consequently, merging SemCom and HARQ is essential to prevent semantic misunderstandings in lossy environments. Furthermore, benefiting from the flexibility in semantics, IK-HARQ allows for efficient information reuse and presents a unified decoder for both transmissions and re-transmissions [17]. Compared to classical SemCom, as shown in Figure 5 and summarized in Algorithm 1, IK-HARQ introduces a decision network for error detection and revamps the JSCC to accomplish incremental decoding. Therefore, in line with our previous works [17], we implement IK-HARQ as a complement in our image SemCom system. Notably, we aim to develop a novel TDA-based decision network, instead of traditional methods (e.g., CRC and Sim32), by introducing additional calibration modules and forging the SC-TDA-HARQ. For example, a TDA encoder extracts the corresponding topological features within an image and constructs the inherent relationship between the image and its topological features. Meanwhile, a TDA decoder is necessary to recover the TDA pattern from noisy signals. Next, we enumerate the key components in SC-TDA-HARQ, and highlight essential changes to classical SemCom.

Specifically, at the transmitter, before passing into the MLP module  $\text{MLP}_1(\cdot)$ , the semantic vector  $x'$  is first concatenated with the TDA encoding result  $x'_{\text{TDA}}$  by a TDA encoder  $\text{TDA}_{\text{en}}(\cdot)$ , which can be expressed as

$$x'_{\text{TDA}} = \text{TDA}_{\text{en}}(x). \quad (8)$$

As is depicted in Figure 4(2), all the MLP layer dimensions are modified to be compatible with the incorporation of TDA. In this manner,  $x'' \in \mathbb{C}^k$  can be re-written as

$$x'' = \text{MLP}_1([x', x'_{\text{TDA}}]). \quad (9)$$

Moreover, once receiving an NAK, the re-transmission starts unless the number of re-transmissions reaches  $N_{\text{max}}$ .

At the receiver,  $\hat{y}_j$ , which denotes the received bits from the  $j$ -th transmission, is concatenated with previously received ones, followed by another MLP module  $\text{MLP}_2(\cdot)$  parameterized by  $\rho$  to recover the semantic symbol vector  $\hat{x}'$  and the topological symbol vector

$\hat{x}'_{\text{TDA}}$ . Mathematically,

$$(\hat{x}', \hat{x}'_{\text{TDA}_j}) = \text{MLP}_2(y_{\text{buffer},j}; \rho), \quad (10)$$

where  $y_{\text{buffer},j} = [\hat{y}_1, \dots, \hat{y}_j]$ . In order to maintain consistent dimension (i.e.,  $k \times N_{\text{max}}$ ) of input,  $y_{\text{buffer},j}$  in the previous  $j < N_{\text{max}}$  transmissions will be padded with zero vectors for decoding, as is illustrated in Figure 4(3). Therefore, the transmitted knowledge is incrementally leveraged. Moreover, the MLP module  $\text{MLP}_2(\cdot)$  shall be well-trained as

$$\rho = \arg \min \mathcal{L}_{\text{MSE}}(x, \hat{x}), \quad (11)$$

so as to capably adapt to different re-transmission times. Meanwhile, the TDA decoder, denoted as  $\text{TDA}_{\text{de}}(\cdot)$  with learnable parameters  $\phi$ , can induce an image  $\hat{x}_{\text{TDA}}$  according to  $\hat{x}'_{\text{TDA}}$ ,

$$\hat{x}_{\text{TDA}} = \text{TDA}_{\text{de}}(\hat{x}'_{\text{TDA}}; \phi). \quad (12)$$

Furthermore, as is shown in Figure 4(4), the TDA decoder can be trained by

$$\phi = \arg \min \mathcal{L}_{\text{CE}}(x, \hat{x}_{\text{TDA}}), \quad (13)$$

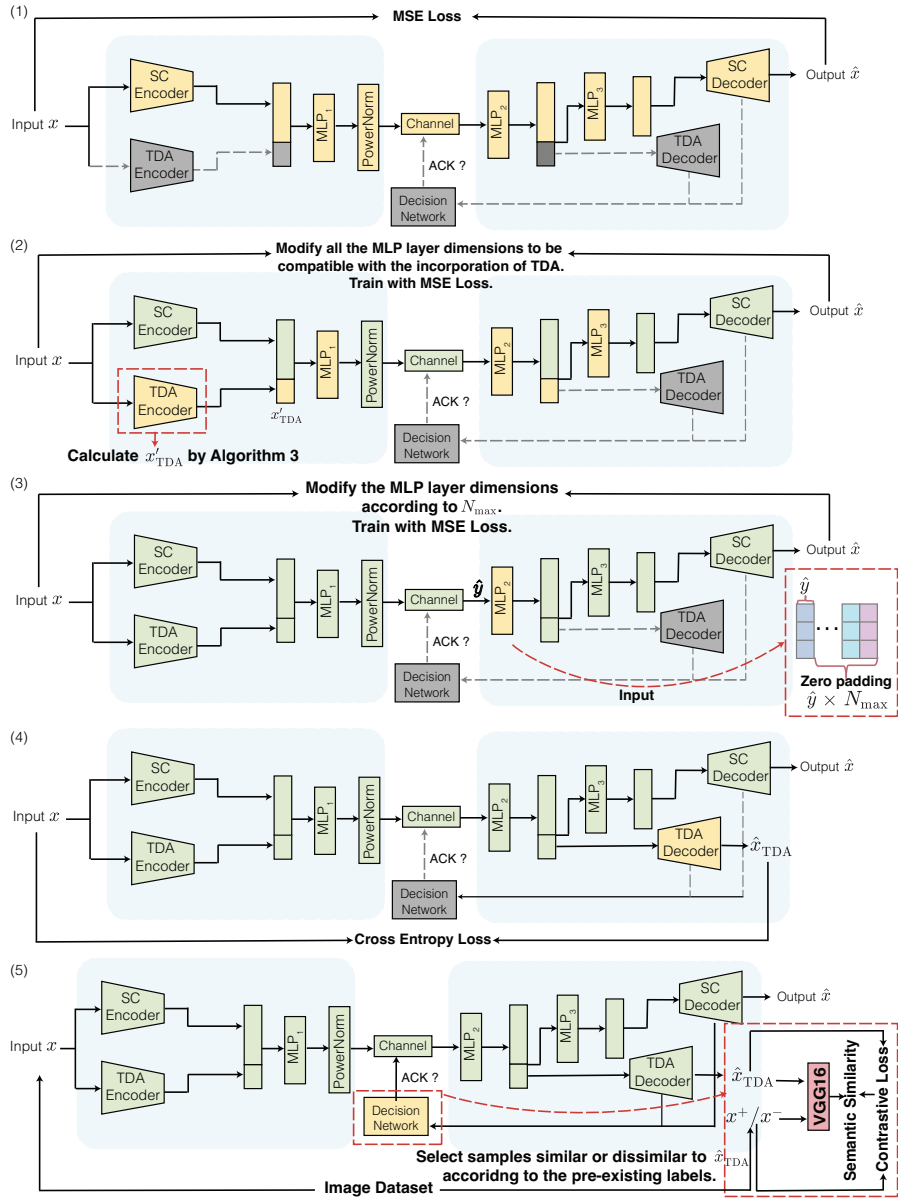
where the cross entropy function  $\mathcal{L}_{\text{CE}}(z, \hat{z}) = \mathbb{E}_z \log \hat{z}$ . The pixel values of both images are scaled to effectively transform each image into a probability distribution, so as to compute the cross entropy loss.

With the recovered  $\hat{x}$  and  $\hat{x}_{\text{TDA}}$ , as lately discussed in Section IV, a decision network will check the correctness of the decoded bits, and triggers a re-transmission when  $\zeta = 0$ . Mathematically,

$$\zeta = \text{DEC}(\hat{x}, \hat{x}_{\text{TDA}}; \gamma) = \begin{cases} 0, & \text{error detected;} \\ 1, & \text{otherwise,} \end{cases} \quad (14)$$

where  $\gamma$  is the trainable parameter in  $\text{DEC}(\cdot)$ . In particular, as demonstrated in Figure 4(5), the decision network, which builds the relationship between an image (e.g.  $x, \hat{x}$ , and  $\hat{x}_{\text{TDA}}$ ) and a decision threshold  $\chi_x$ , can be trained by

$$\gamma = \arg \min \mathcal{L}_{\text{Contrastive}, \chi_x}(x_+/x_-, \hat{x}_{\text{TDA}}), \quad (15)$$



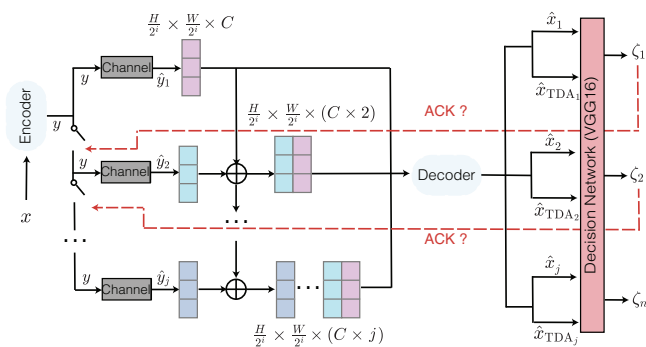
**Figure 4.** The training procedures of SC-TDA-HARQ. Modules colored in yellow are designated to be actively trained. Modules in green indicate that their weights are pre-loaded but not subject to further training during this phase. The modules shown in gray are not included in the training process at this stage.

where  $x^+$  represents images from dataset consisting of  $x$  with TDA features matching those of  $x_{TDA}$ , and  $x^-$  includes images from dataset consisting of  $x$  with TDA features different from  $x_{TDA}$ . Moreover, the contrastive loss function  $\mathcal{L}_{\text{Contrastive}, \chi_x}(x^+/x^-, \hat{x}_{TDA})$  can be formulated as

$$\mathcal{L}_{\text{Contrastive}, \chi_x}(x^+/x^-, \hat{x}_{TDA}) = \frac{1}{2} Y D^2 + \frac{1}{2} (1 - Y) \max(0, \text{margin} - D)^2, \quad (16)$$

where  $Y$  is a label specified according to  $x^+/x^-$  and  $x_{TDA}$ . Specifically, when the input pair comprises  $x^+$  and  $x_{TDA}$ , which indicates that the input belongs to the same category, we set  $Y = 1$ ; conversely, for the input  $x^-$  and  $x_{TDA}$ , we set  $Y = 0$ . In this context, a margin is established for these dissimilar pairs to differentiate them more clearly. Additionally,  $D$  represents the Euclidean distance between the input pair, serving as a quantitative measure of the similarity or disparity between the two images. On the other hand,  $\chi_x$  is a decision threshold.





**Figure 5.** The workflow of semantic IK-HARQ at the receiver.

Generally, we optimize the parameters  $\alpha$ ,  $\beta$ ,  $\phi$  and  $\gamma$  for  $SC_{en}(\cdot)$ ,  $SC_{de}(\cdot)$ ,  $TDA_{de}(\cdot)$  and  $DEC(\cdot)$  in a subsequent manner. Consistent with Figure 4, a detailed description of this training procedures is presented in Algorithm 2. On this basis, we will present the details of DNN design and TDA-enabled error detection scheme in Section IV.

---

**Algorithm 1.** The workflow of IK-HARQ.

---

**Initialization:** The transmission iteration indicator  $j = 1$ ; the maximum re-transmission times  $N_{max} = 3$ ; the ACK indicator  $\zeta = 1$ ; the incremental knowledge buffer  $y_{buffer} = \text{NULL}$ , well-trained  $SC_{en}(\cdot)$  and  $TDA_{en}(\cdot)$ .

**Input:** The transmitted image  $x$ .

**Output:** The reconstructed image  $\hat{x}$ .

- 1: Compute  $y$  based on Eq. (9) and Eq. (2) on top of well-trained  $SC_{en}(\cdot)$  and  $TDA_{en}(\cdot)$ .
  - 2: **while**  $j \leq N_{max}$  and  $\zeta = 0$  **do**
  - 3: Update the received vector  $\hat{y}_j$  after the  $j$ -th transmission with Eq. (4).
  - 4:  $y_{buffer} \leftarrow [y_{buffer}, \hat{y}_j]$ .
  - 5: Update the reconstructed images  $\hat{x}'$  and  $\hat{x}'_{TDA}$  with Eq. (6) and Eq. (12).
  - 6: Update  $\zeta$  according to Eq. (14).
  - 7:  $j = j + 1$ .
  - 8: **end while**
- 

## IV. THE IMPLEMENTATION DETAILS OF SC-TDA-HARQ

This section elaborates on the implementation details of the proposed image SemCom framework SC-TDA-HARQ, which incorporates IK-HARQ and

---

**Algorithm 2.** The training procedures of TDA-enabled SemCom with IK-HARQ.

---

**Initialization:** The maximum re-transmission times  $N_{max} = 3$ .

**Input:** The transmitted image  $x$ .

**Output:** The reconstructed image  $\hat{x}$ .

- 1: Train swin transformer-based  $SC_{en}(\cdot)$  and  $SC_{de}(\cdot)$  by Eq. (7).
  - 2: Add  $TDA_{en}(\cdot)$ . Re-train  $SC_{en}(\cdot)$  and  $SC_{de}(\cdot)$  by Eq. (7).
  - 3: Zero padding the already received information  $\hat{y}$  according to  $N_{max}$ , and train  $MLP_2(\cdot)$  by Eq. (11).
  - 4: Pretrain  $TDA_{de}(\cdot)$  and  $DEC(\cdot)$  by Eq. (13) and Eq. (15).
  - 5: Optimize the parameters  $(\alpha, \beta, \rho, \phi, \gamma)$  by loss function  $\mathcal{L}_{MSE}$ .
- 

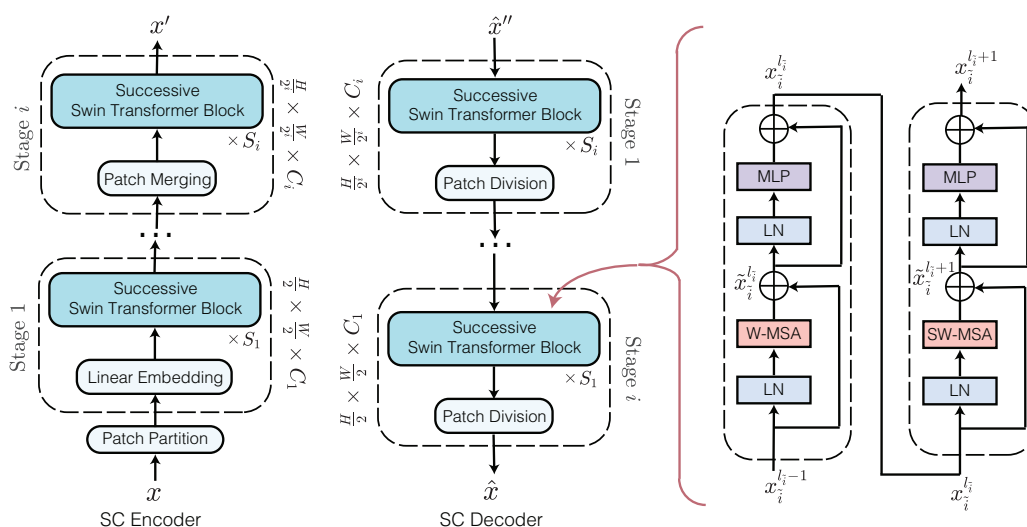
TDA-based error detection into a classical Deep JSCC paradigm.

### 4.1 Swin Transformer-Based Semantic Encoder and Decoder

As illustrated in Figure 6, we utilize swin transformer blocks [20], which manifest the excellence in extracting and recovering discriminative semantic representations of images, as the backbone of encoder  $SC_{en}(\cdot)$  and decoder  $SC_{de}(\cdot)$ .

In the encoding process, we stack several stages, each of which embodies a patch merging layer and  $S_i$  swin transformer blocks, as the encoder, to project tokens into semantic features. Specifically, as shown in Figure 6, each swin transformer block in a stage fulfills some sequence-to-sequence functionality [20], where a shifted window partitioning approach is adopted by reforming the standard multi-head self-attention (MSA) module with partition windows. Consequently, two successive swin transformer blocks  $l_i$  and  $l_i + 1$  of stage  $\tilde{i} \in \{1, \dots, i\}$  are computed according to

$$\begin{aligned}
 \tilde{x}_i^{l_i} &= \text{W-MSA}(\text{LN}(x_i^{l_i-1})) + x_i^{l_i-1}, \\
 x_i^{l_i} &= \text{MLP}(\text{LN}(\tilde{x}_i^{l_i})) + \tilde{x}_i^{l_i}, \\
 \tilde{x}_i^{l_i+1} &= \text{SW-MSA}(\text{LN}(x_i^{l_i}) + x_i^{l_i}), \\
 x_i^{l_i+1} &= \text{MLP}(\text{LN}(\tilde{x}_i^{l_i+1})) + \tilde{x}_i^{l_i+1},
 \end{aligned} \tag{17}$$



**Figure 6.** The DNN structure of the semantic encoder and decoder for encoding the source image and reconstructing the received signal.

where  $\tilde{x}_i^{l_i}$  and  $x_i^{l_i}$  represent the output feature of the (S)W-MSA module and the MLP module for block  $l_i$  at stage  $\tilde{i}$ , respectively. Notably, W-MSA( $\cdot$ ) and SW-MSA( $\cdot$ ) are multi-head self attention modules with regular and shifted windowing configurations. The joint application of W-MSA and SW-MSA contributes to learning implicit connections among partitioned windows, and significantly enhances the semantic extraction capability for images. Besides, LN( $\cdot$ ) denotes the layer normalization operation [20]. After stage  $i$ , the source image  $x \in \mathbb{R}^{H \times W \times 3}$  is gradually converted into a grid of  $\frac{H}{2^i} \times \frac{W}{2^i}$  patches, and subsequently the pixel intensities of each patch are flattened to form a sequence of vectors as a “token”.

Similarly, as illustrated in Figure 6, we adopt a symmetrical DNN structure in the SC decoder, which recovers the image from the noise-polluted representation  $\hat{x}''$ . On this basis, the functionalities in Eqs. (1) and (6) can be accomplished.

## 4.2 TDA-Based Semantic Feature Extraction

In this subsection, we talk about the implementation means of the TDA encoder  $\text{TDA}_{\text{en}}(\cdot)$ , which captures intrinsic topological features of an image  $x$ , and lays the foundation for a TDA-based error detection scheme.

Beforehand, as shown in Figure 7, the 3-color RGB image  $x \in \mathbb{R}^{H \times W \times 3}$  is transformed into a grayscale image  $x \in \mathbb{R}^{H \times W}$ , with grayscale values  $g(u, v)$  at

the pixel  $(u, v)$  where  $1 \leq u \leq H$  and  $1 \leq v \leq W$ . Then, in order to highlight essential features,  $x$  is binarized according to a preset threshold  $\nu$ , that is,

$$g_B(u, v) = \begin{cases} 1, & \text{if } g(u, v) \geq \nu; \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

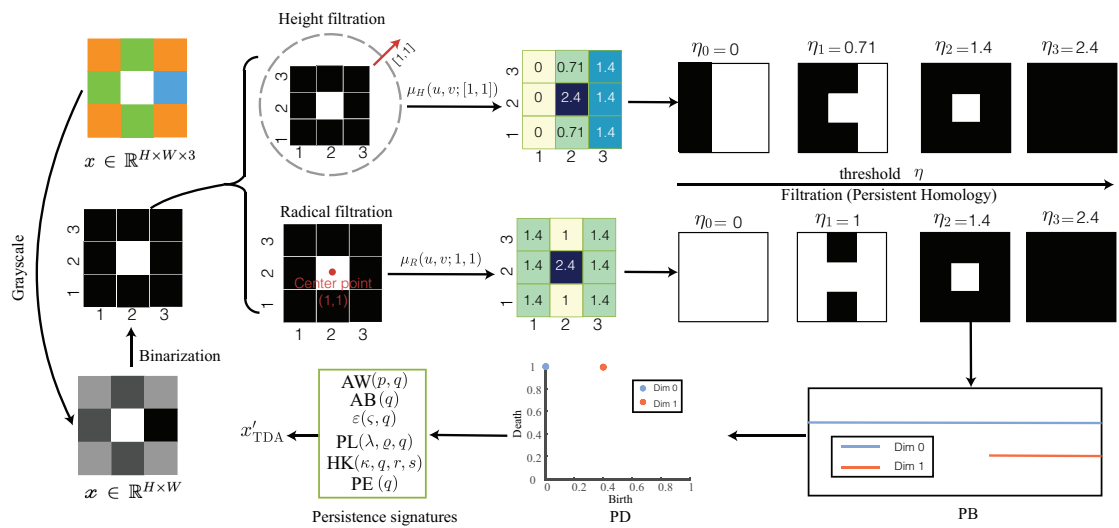
Recalling the discussions in Section I, direct grayscale filtration can merely generate few fixed results, which might be unable to comprehensively capture intact topological features. Therefore, height filtration and radial filtration, which capably incorporate the position information of the homology, are leveraged to make a multiple-perspective analysis of topological features.

- **Height filtration:** Inspired by Morse theory and the PH transform [47], the height filtration  $\mu_H : x \rightarrow \mathbb{R}$  of a 2-dimensional binary image is determined from a chosen unit direction  $\psi \in \mathbb{R}^2$  as

$$\mu_H(u, v; \psi) = \begin{cases} \langle \frac{\psi}{\|\psi\|_2}, (u, v) \rangle, & \text{if } g_B(u, v) = 1; \\ H_\infty, & \text{if } g_B(u, v) = 0, \end{cases} \quad (19)$$

where  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product and  $H_\infty$  is the filtration value of the pixel that is the farthest away from the  $\psi$ .

- **Radial filtration:** Taking the idea from Ref. [48], the radial filtration  $\mu_R$  of  $x$  is computed according to the Euclidean distance between a pixel



**Figure 7.** An example for the pipeline of the TDA encoder to extract the topological feature. In particular, after grayscale, the source image is binarized according to a preset threshold  $\nu$ . The height filtration  $\mu_H(u, v; (1, 1))$  and radical filtration  $\mu_R(u, v; 1, 1)$  are further utilized to assign corresponding values to the image. Sequentially, the filtrations are constructed to obtain the PBs and PDs of the topological features. Finally,  $x'_{TDA}$  is calculated by vectorizing PDs, in terms of persistence signatures (i.e., the Wasserstein distance  $AW(p, q)$ , Bottleneck distance  $AB(q)$ , Betti curve  $\varepsilon(\varsigma, q)$ , persistence landscape  $PL(\lambda, \varrho, q)$ , heat kernel  $HK(\kappa, q, r, s)$  and persistent entropy  $PE(q)$ ).

$(u, v)$  and a preset “center” pixel  $(u_c, v_c)$ , that is,

$$\begin{aligned} & \mu_R(u, v; u_c, v_c) \\ &= \begin{cases} \|(u_c, v_c) - (u, v)\|_2, & \text{if } g_B(u, v) = 1; \\ R_\infty, & \text{if } g_B(u, v) = 0, \end{cases} \end{aligned} \quad (20)$$

where  $R_\infty$  is the distance in terms of the pixel farthest away from the center.

Based on Eq. (19) and Eq. (20), by adjusting the direction  $\psi$  or the center  $(u_c, v_c)$ , different filtration values  $\mu(u, v)$  (either  $\mu_H(u, v)$  and  $\mu_R(u, v)$ ) can be derived from the binary image, thus significantly enhancing the diversity of the captured topological features. Subsequently, a series of cubical complexes is obtained [49]. In particular, for a specific filtration  $\mu(u, v)$  with the maximum  $\mu_{\max}$ , the nested family of subspaces satisfies  $K(x, \eta_0) \subseteq K(x, \eta_1) \subseteq \dots \subseteq K(x, \eta_n)$ , where  $0 \leq \eta_0 \leq \dots \leq \eta_i \leq \dots \leq \eta_n \leq \mu_{\max}$  and the cubical complex  $K(x, \eta_i)$  is formed with a set of cubes  $\xi_q, q \in \{0, 1, 2\}$ . During the filtration, all emerging cubes are sorted as follows: Initially, the 0-cubes (i.e., the nodes) are randomly sorted. For a  $d_1$ -cube  $\xi_{d_1}$  and a  $d_2$ -cube  $\xi_{d_2}$ , the cubes are first sorted by the ascending order of  $d_1$  and  $d_2$ . While for  $d_1 = d_2$ , the order of cubes is determined by its longest (most specific) 1-cube (i.e., the edge). As for those equal longest edges

(i.e., their longest edges emerge at the same value  $\eta$ ), the cubes are sorted in ascending order of the maximum number of nodes. Subsequently, along with the variations of  $\eta \in [0, \mu_{\max}]$ , we can achieve a boundary matrix  $\tau_{\text{filtration}}$  in which each row and column represents all ordered emerging cubes, as well as an incident matrix  $\varpi_{\text{filtration}}$  recording emerging thresholds  $\eta$  for all cubes. Furthermore, if the  $d_{q-1}$ -cube  $\xi_{d_{q-1}}$  is the face of the  $q$ -cube  $\xi_q$ , the corresponding entry in  $\tau_{\text{filtration}}$  refers to 1, otherwise null [50].

Next, by reducing  $\tau_{\text{filtration}}$  into a column-echelon form  $\tau_{\text{reduced}}$ , the  $t$ -th topological feature (i.e., the PD  $(b_t^q, d_t^q)$ ) of dimension  $q$  can be obtained by simply scanning each column from left to right [50]. Specifically, for a column contains non-zero entries (i.e., 1), when the row index corresponding to the last 1 equals the column index, it implies that the cube disappears. Therefore, the corresponding death time  $\eta_{\text{death}}$  is the same as the exact birth time for the cube corresponding to the row index and can be found by looking up  $\varpi_{\text{filtration}}$ . Otherwise, if the values of a column (i.e. the index of the ordered cubes) are all 0, we resort to lookup the emerged threshold  $\eta_{\text{birth}}$  of the cubes, and regard  $[\eta_{\text{birth}}, +\infty)$  as the interval of the cube. Correspondingly, the PB for a cube  $t$  can be represented as  $[b_t^q, d_t^q] = [\eta_{\text{birth}}, \eta_{\text{death}}]$ , while the corresponding PD

can be denoted as  $\vec{l}_t^q = (b_t^q, d_t^q)$ .

Based on the computed PD from height filtration and radial filtration, the PD can be further processed. Correspondingly, consistent with the discussions in Section 2.3, persistence signatures like the Wasserstein distance, Bottleneck distance, Betti curve, PL, and heat kernel are further adopted.

- **Wasserstein amplitude** [40]: Based on the Wasserstein distance, the Wasserstein amplitude of order  $p$  is defined according to  $L_p$  norm of the vector, that is

$$\text{AW}(p, q) = \frac{\sqrt{2}}{2} \left( \sum_t (d_t^q - b_t^q)^p \right)^{\frac{1}{p}}. \quad (21)$$

In this paper, we use  $p \in \{1, 2\}$ .

- **Bottleneck distance** [39]: The Bottleneck amplitude can be considered as a special case of  $p \rightarrow \infty$  in (21). In other words,

$$\text{AB}(q) = \frac{\sqrt{2}}{2} \sup_t (d_t^q - b_t^q). \quad (22)$$

- **Betti curve** [25]: The Betti curve of a PB is the bar number in  $q$  dimension at a given threshold  $\varsigma$

$$\varepsilon(\varsigma, q) = \sum_t \Pi_{[b_t^q, d_t^q]}(\varsigma), \quad (23)$$

where  $\Pi_{[b_t^q, d_t^q]}(\varsigma) = \begin{cases} 1, & \varsigma \in [b_t^q, d_t^q]; \\ 0, & \text{otherwise.} \end{cases}$  [51].

- **Persistent landscape** [26]: The PL is characterized by a sequence of  $\text{PL}(\lambda, \varrho, q) : \mathbb{R} \rightarrow [0, \infty)$  for layers  $\lambda = 1, 2, \dots, \varrho \in \mathbb{N}$ , which denotes the  $\lambda$ -th largest value of  $\{f(\varrho, \vec{l}_t^q, q)\}_{t=1}^{N_q}$ , with  $N_q$  indicating the number of bars in dimensions  $q$  and

$$f(\varrho, \vec{l}_t^q, q) = \begin{cases} 0, & \text{if } \varrho \notin (b_t^q, d_t^q); \\ \varrho - b_t^q, & \text{if } \varrho \in (b_t^q, \frac{b_t^q + d_t^q}{2}); \\ d_t^q - \varrho, & \text{if } \varrho \in (\frac{b_t^q + d_t^q}{2}, d_t^q). \end{cases} \quad (24)$$

Here we consider curves obtained by setting  $\lambda \in \{1, 2\}$ .

- **Heat kernel** [41]: A stable multi-scale kernel bridges with machine learning techniques by Gaussian distributions with mean  $\iota = (b_t^q, d_t^q)$  and

standard deviation  $\kappa$  (i.e.,  $\mathcal{N}(\iota, \kappa)$ ) [41]. Mathematically, denoting  $(r, s) \in \mathbb{R}^2$  as any position in the final heat map, the value of  $(r, s)$  corresponds to the summation of the output through the Gaussian distribution over every point of the PD. HK is a real-valued function on  $\mathbb{R}^2$  and can be expressed as

$$\text{HK}(\kappa, q, r, s) = \frac{1}{8\pi\kappa} \sum_t e^{-\frac{(r-b_t^q)^2 + (s-d_t^q)^2}{8\kappa}} - e^{-\frac{(r-d_t^q)^2 + (s-b_t^q)^2}{8\kappa}}, \quad (25)$$

where  $\kappa \in \{10, 15\}$ .

- **Persistent entropy** [44]: The PE of a PB takes the Shannon entropy of the persistence (lifetime) of all cycles [52]

$$\text{PE}(q) = - \sum_t \frac{\vec{l}_t^q}{L} \log\left(\frac{\vec{l}_t^q}{L}\right), \quad (26)$$

where  $L = \sum_t \vec{l}_t^q$  is the sum of all the persistences.

For a specific case of filtration  $\psi$  or  $(u_c, v_c)$ , we concatenate all abovementioned metrics as the topological persistence signatures as

$$x'_{\text{TDA}_s}(f) = [\varepsilon(\varsigma, q), \text{PL}(\lambda, \varrho, q), \text{HK}(\kappa, q, r, s), \text{AW}(p, q), \text{AB}(q), \text{PE}(q)]_f, \quad (27)$$

where the operator  $[\cdot]_f$  implies all components are computed under the filtration parameterized by  $f$  (e.g.,  $\psi$  and  $(u_c, v_c)$ ).

Furthermore, after the process of the TDA encoder,  $x'_{\text{TDA}}$  can be achieved by further concatenating Eq. (27) from  $N_\psi$  height filtration and  $N_c$  radial filtration, that is

$$x'_{\text{TDA}} = [x'_{\text{TDA}_s}(\psi_1), \dots, x'_{\text{TDA}_s}(N_\psi), x'_{\text{TDA}_s}((u_c, v_c)_1), \dots, x'_{\text{TDA}_s}((u_c, v_c)_{N_c})]. \quad (28)$$

Meanwhile, the procedures for TDA encoding have been summarized in Algorithm 3.

### 4.3 TDA-Based Decision Network for Error Detection

The stability theorem [39] proves that two similar objects possess close PDs. Consequently, it is eligible to use PD or the induced persistence signatures to measure the topological closeness between two images. Therefore, intuitively, the topological features of the reconstructed image  $\hat{x}$  at the receiver can be computed for the ACK/NAK decision making, by comparing with the transmitted TDA features, named as TDA-Computation decision-making approach. Alternatively, topological features can be regarded as a special type of hidden features and source images can be recovered accordingly by utilizing various methods like generative adversarial network (GAN) [53], convolutional neural network (CNN). Nevertheless, both means can be considerably time-consuming, and it is worthy of investigation of a computation-effective means for making error detection decisions.

Hence, we propose TDA-Recovery decision-making approach. Instead of a repetitive TDA computation or complete image recovery, we only try to establish the relationship between the transmitted topological features and source images and adopt an MLP module as the TDA decoder  $\text{TDA}_{\text{de}}$  to acquire  $\hat{x}_{\text{TDA}}(\cdot)$  from the noisy  $\hat{x}'_{\text{TDA}}$ , that is,

$$\hat{x}_{\text{TDA}} = \text{TDA}_{\text{de}}(\hat{x}'_{\text{TDA}}; \phi) = \text{MLP}(\hat{x}'_{\text{TDA}}; \phi). \quad (29)$$

In the end, the output of TDA decoder  $\text{TDA}_{\text{de}}(\cdot)$  is discriminated to yield the re-transmission decision. As shown in Figure 5, we design a decision network  $\text{DEC}(\cdot)$ , which is based on a pre-trained VGG16 network [54, 55], to capably measure the similarity between  $\hat{x}_{\text{TDA}}$  and  $\hat{x}$ . Finally, the decision network  $\text{DEC}(\cdot)$  in Eq. (14) is accomplished as

$$\zeta = \begin{cases} 1, & \text{if } \text{VGG}(\hat{x}, \hat{x}_{\text{TDA}}) \geq \chi_x; \\ 0, & \text{if } \text{VGG}(\hat{x}, \hat{x}_{\text{TDA}}) < \chi_x. \end{cases} \quad (30)$$

In essence, the re-transmission decision by  $\text{DEC}(\cdot)$  is based on the assessment of whether  $\hat{x}$  and  $\hat{x}_{\text{TDA}}$  fall into the same category, while  $\text{VGG}(\cdot)$  outputs the semantic similarity between the received content and the TDA-induced image. This re-transmission determination can be reached if the semantic similarity between

---

**Algorithm 3.** *The TDA-based encoder.*

---

**Initialization:** The number of the directions used in the height filtration  $N_\psi$ ; the number of the centers used in the radial filtration  $N_c$ ; the binarized threshold  $\nu$ .

**Input:** The transmitted image  $x$ .

**Output:** The topological feature  $x'_{\text{TDA}}$ .

- 1: Transform the image  $x \in \mathbb{R}^{H \times W \times 3}$  into a grayscale image  $x \in \mathbb{R}^{H \times W}$ , with grayscale values  $g(u, v)$  at a pixel  $(u, v)$   $1 \leq u \leq H, 1 \leq v \leq W$ .
  - 2: Binarize  $x$  with Eq. (18).
  - 3: **while**  $1 \leq s \leq N_\psi$  **do**
  - 4:   Update pixel value  $\mu(u, v)$  with Eq. (19).
  - 5:   Construct the filtration of nested cubical complex  $K$  with  $K(x, \eta_0) \subseteq K(x, \eta_1) \subseteq \dots \subseteq K(x, \eta_i) \subseteq \dots \subseteq K(x, \eta_n)$ , where  $0 \leq \eta_i \leq \mu_{\max}(u, v)$ .
  - 6:   Record all the cubes  $\xi$  emerged during the filtration and their emerged threshold  $\eta_\xi$ .
  - 7:   Sort the recorded cubes and construct a corresponding boundary matrix  $\tau_{\text{filtration}}$  and an incident matrix  $\varpi_{\text{filtration}}$ .
  - 8:   Reduce  $\tau_{\text{filtration}}$  as  $\tau_{\text{reduce}}$  and compute  $(b_t^q, d_t^q)$  according to  $\tau_{\text{reduce}}$  and  $\varpi_{\text{filtration}}$ .
  - 9:   Transform  $(b_t^q, d_t^q)$  into the vector input with Eq. (21)–Eq. (26).
  - 10: **end while**
  - 11: **while**  $1 \leq s \leq N_c$  **do**
  - 12:   Update pixel value  $\mu(u, v)$  with Eq. (20).
  - 13:   Repeat steps in Line 5 to Line 9.
  - 14: **end while**
  - 15: Update  $x'_{\text{TDA}}$  with Eq. (28).
- 

$\hat{x}$  and  $\hat{x}_{\text{TDA}}$  by  $\text{VGG}(\cdot)$  exceeds the decision threshold  $\chi_x$ . In other words,  $\hat{x}$  and  $\hat{x}_{\text{TDA}}$  are sufficiently different.

Table 4 compares the peak signal-to-noise ratio (PSNR) and running time between the TDA-Recovery and TDA-Computation decision-making approaches of CIFAR10 dataset under AWGN channel. The SNR is set to 10 dB and the compression rate is set to 1/3. The experiments are implemented in PyTorch and run on NVIDIA GeForce RTX 4090. As indicated in Table 4, while both methods yield similar PSNR performance, the former operates on a second-level time, whereas the latter requires an hour-level time. There-



**Table 4.** Comparison of PSNR and running time between TDA-Recovery and TDA-Computation decision-making approaches.

Method	PSNR	Running time
TDA-Recovery decision-making approach	39.757 dB	0.113 s
TDA-Computation decision-making approach	39.365 dB	5.6 h

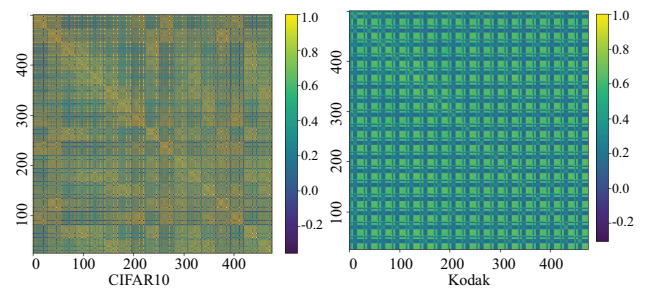
fore, it validates the effectiveness of the mentioned feature selection in terms of PSNR and computational costs.

## V. EXPERIMENTAL SETTING AND NUMERICAL RESULTS

### 5.1 Experimental Setting

We evaluate the performance of SC-TDA-HARQ on transmitting typical image datasets with resolutions spanning from  $32 \times 32$  up to 2K. Specifically, we employ the CIFAR10 dataset for training and testing of lower resolution images. Conversely, for high-resolution images, the DIV2K dataset is selected for training purposes, while the Kodak dataset is utilized for testing. Moreover, we qualify the performance using both the widely used pixel-wise metric PSNR and the perceptual metric multi-scale structural similarity (MS-SSIM). The performance comparison is primarily conducted against varying SNRs and compression rates. In terms of setting the compression rate, it is regulated by the compression dimension  $C$ . For the CIFAR10 dataset,  $C$  is selected from the set  $\{8, 16, 24, 32, 48\}$ . Meanwhile, for high-resolution datasets,  $C$  is chosen from a different set comprising of  $\{16, 32, 64, 96, 128, 196\}$ . SC-TDA-HARQ is compared with separation-based source and channel coding (i.e., BPG codec for source image compression and LDPC codec for channel coding), Deep JSCC [2] and WITT [6]. For the LDPC codec, various combinations of code rates and modulations are evaluated. In particular, (3072, 6144), (3072, 4608), and (1536, 4608) are chosen for 1/2, 2/3, and 1/3 code rates, respectively, while BPSK, 4-QAM, 16-QAM, or 64-QAM are considered for modulations [56].

During the training, images are randomly cropped into  $256 \times 256$  patches. Besides, we exploit the Adam optimizer with a learning rate of  $1 \times 10^{-4}$ , and



**Figure 8.** The Pearson correlation matrix of CIFAR10 and Kodak datasets.

the batch size is set to 128 and 16 for the CIFAR10 dataset and DIV2K dataset, respectively. Other essential settings for swin transformer in SC-TDA-HARQ are summarized in Table 5.

On the other hand, the direction  $\psi$  used in the height filtration and the center used in the radical filtration are set as Table 6, while typical settings for computing distance-based metrics or kernels are given in Section 4.2. Accordingly, the TDA encoder can generate 476 dimensions of topological features. In terms of Pearson coefficients, Figure 8 presents the correlation of these features [57], 28 of which with least correlations are selected to reduce the computational complexity without sacrificing the performance.

### 5.2 Numerical Results

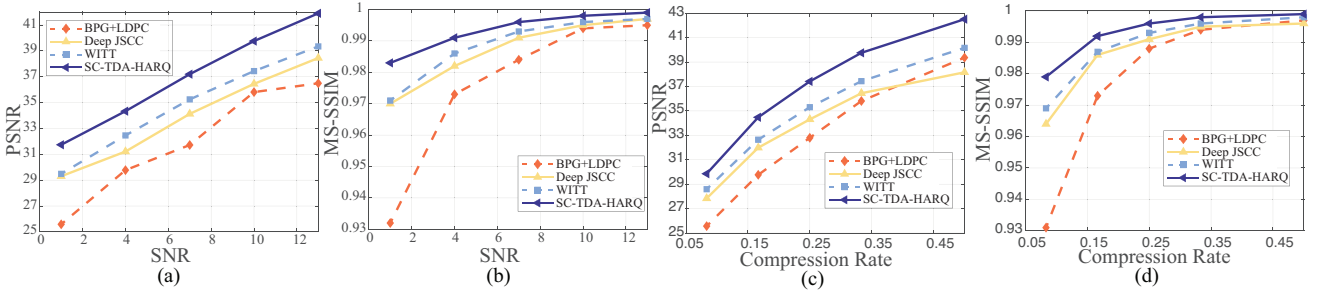
Numerically, Figure 9 shows the PSNR and MS-SSIM performance of CIFAR10 dataset over the AWGN channel. Figure 9(a) and Figure 9(b) investigates the performance versus the SNR, where the compression rate is set to 1/3. It can be observed that for the low-resolution dataset, SC-TDA-HARQ outperforms other schemes (i.e., “BPG+LDPC”, Deep JSCC and WITT) for all SNRs in terms of both PSNR and MS-SSIM. Meanwhile, the MS-SSIM performance gain decreases with the increase of the SNR, and SC-TDA-HARQ achieves the maximum gain when SNR is set to 1 dB. On the other hand, Figure 9(c) and Figure 9(d) manifests the performance versus the compression rate where the SNR = 10 dB. SC-TDA-HARQ can generally outperform both traditional and semantic communication methods by the comparison of PSNR and MS-SSIM. Moreover, the PSNR performance gap increases in direct proportion to the compression rate; while there is a significant drop in MS-SSIM performance gain with the growth of the compression rate.

**Table 5.** The default hyper-parameter settings of swin transformer.

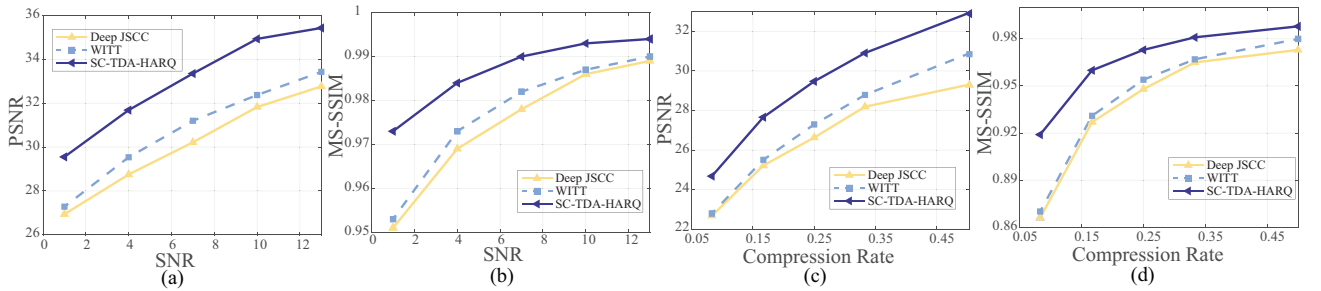
Hyperparameter	No. of Stages	Depth (No. of swin transformer blocks)	Embedding dimension of each stage	No. of heads	Window size
<b>CIFAR10</b>	2	[2, 4]	[128, 256]	[4, 8]	2
<b>DIV2K</b>	4	[2, 2, 6, 2]	[128, 192, 256, 320]	[4, 6, 8, 10]	8

**Table 6.** The default hyper-parameter settings of TDA encoder.

Filtration	Values
Height Filtration $\psi$	(1, 0), (0, 1), (1, 1), (1, -1), (-1, 1), (-1, -1), (-1, 0), and (0, -1)
Radial Filtration $(u, v)$	(23, 7), (23, 15), (23, 23), (7, 23), (7, 15), (7, 7), (15, 23), (15, 15), and (15, 7)



**Figure 9.** PSNR and MS-SSIM performance of the CIFAR10 dataset over the AWGN channel. (a)–(b) PSNR and MS-SSIM performance of the CIFAR10 dataset versus the SNR. The compression rate is set to 1/3. (c)–(d) PSNR and MS-SSIM performance of the CIFAR10 dataset versus the compression rate. The SNR is set to 10 dB.



**Figure 10.** PSNR and MS-SSIM performance of the CIFAR10 dataset over the Rayleigh channel. (a)–(b) PSNR and MS-SSIM performance of the CIFAR10 dataset versus the SNR, respectively. The compression rate is set to 1/3. (c)–(d) PSNR and MS-SSIM performance of the CIFAR10 dataset versus the compression rate. The SNR is set to 3 dB.

On the other hand, Deep JSCC achieves inferior performance than “BPG+LDPC” at SNR = 10 dB, 1/2 compression rate when transmitting CIFAR10 images.

Figure 10 demonstrates the PSNR and MS-SSIM performance of CIFAR10 dataset over the Rayleigh channel. Figure 10(a) and Figure 10(b) investigates the performance versus the SNR, where the compression rate  $R$  is set to 1/3. By Figure 10(a) and Figure 10(b), for the low-resolution dataset, SC-TDA-HARQ is superior to conventional SemCom schemes (i.e., Deep JSCC and WITT) for all SNRs in terms of both PSNR and MS-SSIM. Different from the slight MS-

SSIM gain obtained when the SNR increases to 13 dB over the AWGN channel, SC-TDA-HARQ still holds considerable gain with the growth of the SNR over the Rayleigh channel. On the other hand, Figure 10(c) and Figure 10(d) manifests the performance versus the compression rate where the SNR = 3 dB. Concerning the PSNR and MS-SSIM performance obtained by Deep JSCC and WITT, SC-TDA-HARQ shows substantial performance gain. There is also an enlarged gap in PSNR as the compression rate increases. A similar observation also applies to the MS-SSIM versus the compression rate, except that SC-TDA-HARQ

yields superior significant improvement no matter how the compression rate changes. The findings in Figure 9 and Figure 10 validate the joint advantages of TDA and IK-HARQ by making re-transmission decisions and reusing information, so as to reduce semantic misunderstanding in lossy environments. Furthermore, Figure 11 visualizes the reconstructions of the Kodak dataset under AWGN channel at SNR= 10dB and the compression rate  $R = 1/16$ . It can be observed that SC-TDA-HARQ can achieve better visual quality with the same bandwidth cost. More specifically, it avoids block artifacts and produces higher fidelity textures and details.

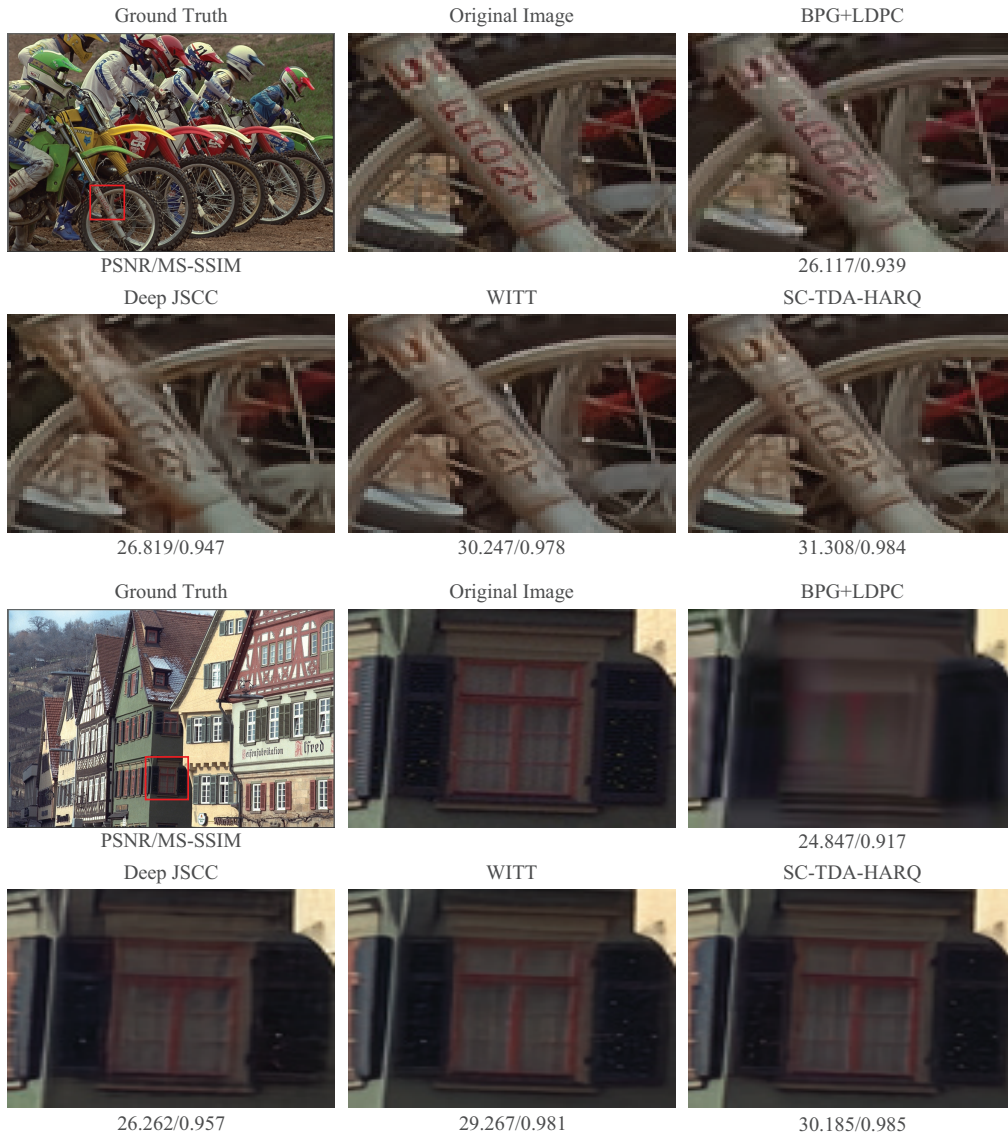
Next, we analyze the contributing impact of topological features in terms of PSNR. For CIFAR10 dataset, Figure 12(a) and Figure 12(b) demonstrates the PSNR performance versus the SNR at the compression rate  $R = 1/3$  and SNR = 10 dB over the AWGN channel, respectively. When Comparing with SC, i.e., the conventional SemCom model with swin transformer-based semantic codec, it can be observed that the TDA-enhanced SemCom (SC-TDA) outperforms SC for low-resolution image datasets, which illustrates that the incorporation of TDA benefits the encoding and decoding processing during the image transmission. As the SNR increases, the performance enhancement brought about by TDA becomes more pronounced. Meanwhile, with the fixed SNR, TDA achieves the ascending performance gap with the growth of the compression rate as well. Figure 12(c) and Figure 12(d) illustrate the variations of PSNR versus the SNR at the compression rate  $R = 1/3$  and SNR = 3 dB over the Rayleigh channel, respectively. Similar to the performance under AWGN channel, the incorporation of TDA obtains performance improvement to some extent, and the maximum performance gain is obtained at SNR = 10 dB. Meanwhile, regardless of channel environments, the combination of IK-HARQ still obtains a considerable gain in the transmission of low resolution images.

On the other hand, the effects of integrating TDA features to Kodak dataset is also investigated. Figure 13(a) and Figure 13(b) demonstrates the PSNR performance versus the SNR at the compression rate  $R = 1/16$  and SNR = 10 dB over the AWGN channel, respectively. It can be observed that the SC-TDA mostly outperforms SC, which indicates that the incorporation of TDA benefits the encoding and decoding

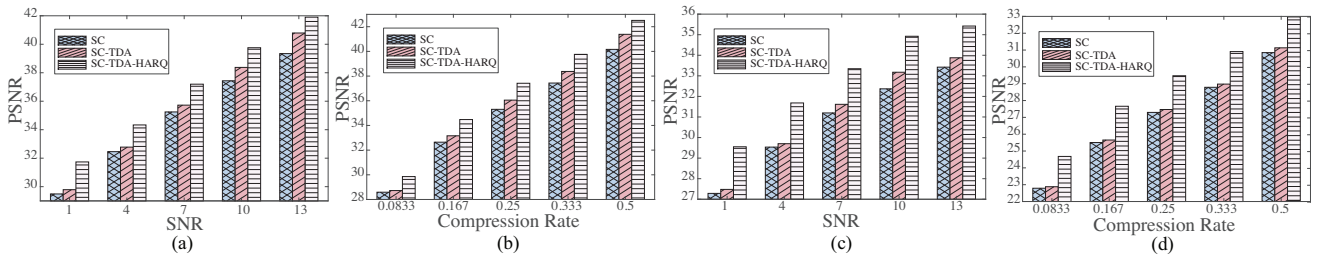
processing during the image transmission to some extent. Figure 13(c) and Figure 13(d) illustrate the PSNR performance versus the SNR at the compression rate  $R = 1/3$  and SNR = 3 dB over the Rayleigh channel, respectively. Consistent with the performance under AWGN channel, the inclusion of TDA features leads to some improvement in performance, with the most significant gain observed at an SNR = 10 dB. Moreover, Figure 13 shows that though SC-TDA sporadically under-performs than SC possibly limited 28-dimensional TDA features, regardless of the channel conditions, the incorporation of IK-HARQ consistently achieves notable gains in the transmission of high-resolution images. In other words, the combination of TDA and IK-HARQ can consistently boost the transmission performance.

Next, we investigate the impact of different dimensions of TDA features on the PSNR performance. In Figure 14, we present a comparative analysis of PSNR performances for the Kodak dataset, using the SC-TDA model with different dimensions of TDA features. Observations from Figure 14 reveal that the SC-TDA model employing 28-dimensional TDA features achieves a PSNR performance on par with that of the SC-TDA model utilizing 476-dimensional (i.e., full-dimensional) features. Notably, this comparable performance is attained with significantly less transmission resource usage, highlighting the efficiency of the SC-TDA model when it incorporates optimally TDA features. This finding underscores the potential of TDA features in enhancing image transmission quality while conserving transmission resources. Additionally, in light of the instances where SC-TDA does not surpass SC, as illustrated in Figure 13, Figure 14 indicates that increasing the dimensions of the TDA features, which enhances the complexity or richness of TDA features, could address this issue to some extent. Moreover, as demonstrated in Figure 15, the SC-TDA-HARQ model consistently outperforms the SC-TDA model, and SC-TDA-HARQ with 476-dimensional features shows slightly performance improvement over that with 28-dimensional features. Thus, it can be safely concluded that under sufficient length of check bits, substantial performance improvements can be anticipated.

We also compare our topological error detection method with other scheme like Sim32 in Appendix. Given the importance of establishing a suitable thresh-



**Figure 11.** Examples of visual comparison of Kodak dataset under AWGN channel at  $SNR = 10$  dB and  $R = 1/16$ .

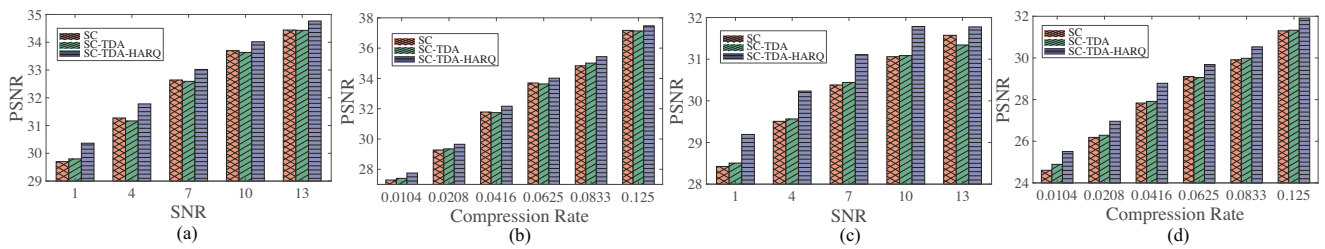


**Figure 12.** The PSNR comparison of the CIFAR10 dataset versus the SNR and compression rate with the assistance of TDA. (a) The compression rate is set to  $1/3$  under AWGN channel. (b) The SNR is set to 10 dB under AWGN channel. (c) The compression rate is set to  $1/3$  under Rayleigh channel. (d) The SNR is set to 3 dB under Rayleigh channel.

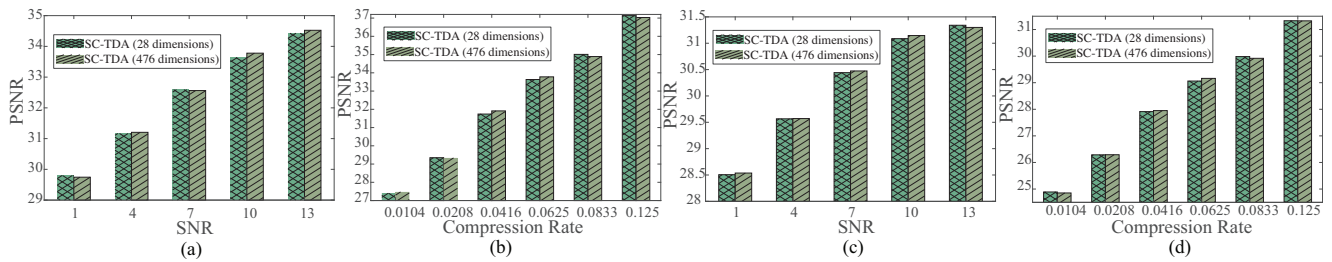
old for re-transmission by observing changes in semantic similarity, we undertake a comparative analysis. This analysis focuses on the variations in semantic

similarity across different batches of images, sampled at various stages during the training process. As is illustrated in Figure 1 and Table 1, Sim32 is less quali-

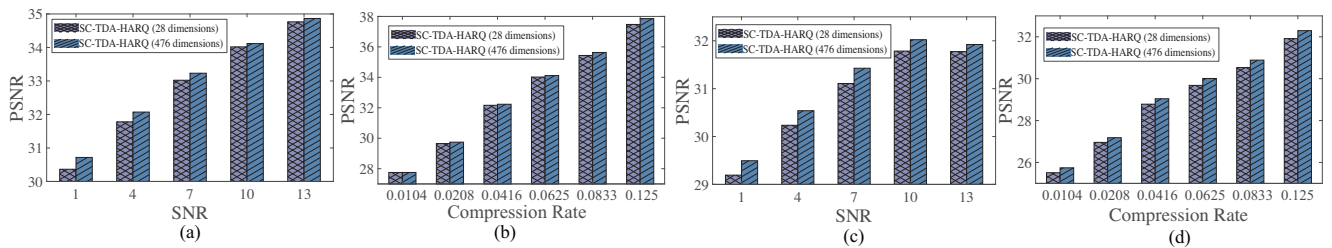




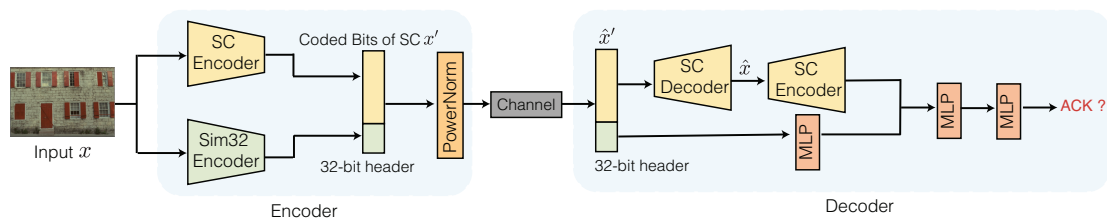
**Figure 13.** The PSNR comparison of the *Kodak* dataset versus the SNR and compression rate with the assistance of TDA. (a) The compression rate is set to 1/16 under AWGN channel. (b) The SNR is set to 10 dB under AWGN channel. (c) The compression rate is set to 1/16 under Rayleigh channel. (d) The SNR is set to 3 dB under Rayleigh channel.



**Figure 14.** The PSNR comparison of the *Kodak* dataset versus the SNR and compression rate, considering different dimensions of TDA features in the SC-TDA model. (a) The compression rate is set to 1/16 under AWGN channel. (b) The SNR is set to 10 dB under AWGN channel. (c) The compression rate is set to 1/16 under Rayleigh channel. (d) The SNR is set to 3 dB under Rayleigh channel.



**Figure 15.** The PSNR comparison of the *Kodak* dataset versus the SNR and compression rate, considering different dimensions of TDA features in the SC-TDA-HARQ model. (a) The compression rate is set to 1/16 under AWGN channel. (b) The SNR is set to 10 dB under AWGN channel. (c) The compression rate is set to 1/16 under Rayleigh channel. (d) The SNR is set to 3 dB under Rayleigh channel.



**Figure 16.** Pipeline of SC-HARQ with Sim32.

ified for the decision of re-transmission, as the similarity computed by Sim32 remains almost unchanged under various channel conditions. On the contrary, SC-

TDA-HARQ achieves more robust performance compared to Sim32, thus validating the positive effect of TDA in making the re-transmission decision.



## VI. CONCLUSION

In this paper, we have proposed a high-efficiency framework SC-TDA-HARQ to improve the performance of wireless image transmission. In particular, with the assistance of TDA, the SC-TDA-HARQ framework is built upon the swin transformer to extract hierarchical image representations. Furthermore, SC-TDA-HARQ utilizes IK-HARQ to fully use incrementally knowledge from multiple transmissions, while TDA-based error detection scheme is leveraged to determine the conditions for re-transmissions. Extensive results have demonstrated that SC-TDA-HARQ outperforms both conventional communication and classical SemCom methods, with manifested performance gain under different SNRs and compression rates. Meanwhile, due to the capability to leverage inner topological and geometric information embedded images, TDA significantly improves robustness compared to other error detection schemes like Sim32. In the future, we will investigate efficient means to compute TDA, so as to better tackle SemCom issues.

## ACKNOWLEDGEMENT

This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFE0200600, in part by the National Natural Science Foundation of China under Grant 62071425, in part by the Zhejiang Key Research and Development Plan under Grant 2022C01093, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR23F010005, in part by the National Key Laboratory of Wireless Communications Foundation under Grant 2023KP01601, and in part by the Big Data and Intelligent Computing Key Lab of CQUPT under Grant BDIC-2023-B-001.

## NOTES

<sup>1</sup>The details to apply Sim32 for computing semantic similarity are given in Appendix.

## APPENDIX

**Introduction of similarity computation in Figure 1 and Table 1:** As is depicted in Figure 16, for Sim32,

the transmission process involves two parallel encoding paths corresponding to the SC encoder and the Sim32 encoder, which are responsible for generating the coded bits for the SC and a 32-bit header, respectively [16]. Specifically, the Sim32 encoder consists of an SC encoder-alike swin transformer and an MLP layer. These encoded symbols are then concatenated and transmitted through the communication channel. On the receiver side, the SC decoder reconstructs the SC coded bits. The reconstructed content is then subjected to a downsampling process using the SC encoder and combined with the symbols recovered from the 32-bit header by the Sim32 decoder. This amalgamated vector subsequently passes through additional MLP layers. The final output from these MLP layers serves as an indicator of the semantic similarity between the received content and the 32-bit header, effectively measuring how closely the received information matches the original source content [16].

## References

- [1] Z. Lu, R. Li, *et al.*, “Semantics-empowered communications: A tutorial-cum-survey,” *IEEE Commun. Surv. Tutor.*, 2023.
- [2] E. Bourtsoulatze, D. Kurka, *et al.*, “Deep joint source-channel coding for wireless image transmission,” *IEEE Trans. Cognit. Commun. Netw.*, 2019, vol. 5, no. 3, pp. 567–579.
- [3] D. Gündüz, Z. Qin, *et al.*, “Beyond transmitting bits: Context, semantics, and task-oriented communications,” *IEEE J. Sel. Areas Commun.*, 2022, vol. 41, no. 1, pp. 5–41.
- [4] E. Strinati and S. Barbarossa, “6g networks: Beyond shannon towards semantic and goal-oriented communications,” *Comput. Netw.*, 2021, vol. 190, p. 107930.
- [5] M. Ding, J. Li, *et al.*, “Snr-adaptive deep joint source-channel coding for wireless image transmission,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021.
- [6] K. Yang, S. Wang, *et al.*, “Witt: A wireless image transmission transformer for semantic communications,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2023.
- [7] A. Grover and S. Ermon, “Uncertainty autoencoders: Learning compressed representations via variational information maximization,” in *Proc. Int. Conf. Artif. Intell. Stat.*, 2019.
- [8] M. Yang, C. Bian, *et al.*, “Deep joint source channel coding for wireless image transmission with ofdm,” in *Proc. IEEE Int. Conf. Commun.*, 2021.
- [9] D. Kurka and D. Gündüz, “Deepjpsc-f: Deep joint source-channel coding of images with feedback,” *IEEE J. Sel. Areas Inf. Theory*, 2020, vol. 1, no. 1, pp. 178–193.
- [10] H. Wu, Y. Shao, *et al.*, “Transformer-aided wireless image transmission with channel feedback,” *arXiv preprint arXiv:2306.09101*, 2023.

- [11] C. Lee, X. Hu, *et al.*, “Deep joint source-channel coding with iterative source error correction,” in *Proc. Int. Conf. Artif. Intell. Stat.*, 2023.
- [12] Z. Zhang, Q. Yang, *et al.*, “Wireless transmission of images with the assistance of multi-level semantic information,” in *Proc. Int. Symp. Wireless Commun. Syst.*, 2022.
- [13] K. Ma, Y. Shi, *et al.*, “A deep learning based broadcast approach for image semantic communication over fading channels,” *China Communications*, 2024, vol. 21, no. 7, pp. 78–94.
- [14] Z. Weng and Z. Qin, “Semantic communication systems for speech transmission,” *IEEE J. Sel. Areas Commun.*, 2021, vol. 39, no. 8, pp. 2434–2444.
- [15] N. Farsad, M. Rao, *et al.*, “Deep learning for joint source-channel coding of text,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018.
- [16] P. Jiang, C. Wen, *et al.*, “Deep source-channel coding for sentence semantic transmission with harq,” *IEEE Trans. Commun.*, 2022, vol. 70, no. 8, pp. 5225–5240.
- [17] Q. Zhou, R. Li, *et al.*, “Adaptive bit rate control in semantic communication with incremental knowledge-based harq,” *IEEE Open J. Commun. Soc.*, 2022, vol. 3, pp. 1076–1089.
- [18] F. Zhai, Y. Eisenberg, *et al.*, “Joint source-channel coding for video communications,” *Handb. Image Video Process.*, 2005, pp. 1065–1082.
- [19] P. Jiang, C. Wen, *et al.*, “Wireless semantic communications for video conferencing,” *IEEE J. Sel. Areas Commun.*, 2022, vol. 41, no. 1, pp. 230–244.
- [20] Z. Liu, Y. Lin, *et al.*, “Swift transformer: Hierarchical vision transformer using shifted windows,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021.
- [21] D. Kurka and D. Gündüz, “Successive refinement of images with deep joint source-channel coding,” in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, 2019.
- [22] B. Makki, T. Svensson, *et al.*, “Finite block-length analysis of the incremental redundancy harq,” *IEEE Wireless Commun. Lett.*, 2014, vol. 3, no. 5, pp. 529–532.
- [23] G. Castagnoli, S. Brauer, *et al.*, “Optimization of cyclic redundancy-check codes with 24 and 32 parity bits,” *IEEE Trans. Commun.*, 1993, vol. 41, no. 6, pp. 883–892.
- [24] R. Hamming, “Error detecting and error correcting codes,” *Bell Syst. Tech. J.*, 1950, vol. 29, no. 2, pp. 147–160.
- [25] F. Chazal and B. Michel, “An introduction to topological data analysis: Fundamental and practical aspects for data scientists,” *Front. Artif. Intell.*, 2021, vol. 4, p. 108.
- [26] P. Bubenik, *et al.*, “Statistical topological data analysis using persistence landscapes,” *J. of Mach. Learn. Res.*, 2015, vol. 16, no. 1, pp. 77–102.
- [27] H. Adams, T. Emerson, *et al.*, “Persistence images: A stable vector representation of persistent homology,” *J. Mach. Learn. Res.*, 2017, vol. 18.
- [28] S. Balakrishnan, A. Rinaldo, *et al.*, “Minimax rates for homology inference,” in *Proc. Int. Conf. Artif. Intell. Stat.*, 2012.
- [29] T. Bonis, M. Ovsjanikov, *et al.*, “Persistence-based pooling for shape pose recognition,” in *Proc. Int. Conf. Comput. Topol. Image Context Workshop*, 2016, pp. 19–29.
- [30] Z. Cang and G. Wei, “Topologynet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions,” *PLoS computational biology*, 2017, vol. 13, no. 7.
- [31] F. Chazal, M. Glisse, *et al.*, “Convergence rates for persistence diagram estimation in topological data analysis,” in *Proc. Int. Conf. Mach. Learn.*, 2014.
- [32] G. Wallace, “The jpeg still picture compression standard,” *IEEE Trans. Consum. Electron.*, 1992, vol. 38, no. 1, pp. 18–34.
- [33] A. Skodras, C. Christopoulos, *et al.*, “The jpeg 2000 still image compression standard,” *IEEE Signal Process. Mag.*, 2001, vol. 18, no. 5, pp. 36–58.
- [34] D. Yee, S. Soltaninejad, *et al.*, “Medical image compression based on region of interest using better portable graphics (bpg),” in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2017.
- [35] L. Song, S. Yu, *et al.*, “Low-density parity-check codes: Highway to channel capacity,” *China Communications*, 2023, vol. 20, no. 2, pp. 235–256.
- [36] J. Hagenauer, “The turbo principle: Tutorial introduction and state of the art,” in *Proc. Int. Symp. Turbo Codes Relat. Top.*, 1997.
- [37] V. Chandé, H. Jafarkhani, *et al.*, “Joint source-channel coding of images for channels with feedback,” in *Proc. Inf. Theory Workshop*, 1998.
- [38] J. Lu, A. Nosratinia, *et al.*, “Progressive joint source-channel coding in feedback channels,” in *Proc. IEEE Data Compression Conf.*, 1999.
- [39] C. David, H. Edelsbrunner, *et al.*, “Stability of persistence diagrams,” in *Proc. Annu. Symp. Comput. Geom.*, 2005.
- [40] Y. Mileyko, S. Mukherjee, *et al.*, “Probability measures on the space of persistence diagrams,” *Inverse Probl.*, 2011, vol. 27, no. 12, p. 124007.
- [41] J. Reininghaus, S. Huber, *et al.*, “A stable multi-scale kernel for topological machine learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015.
- [42] R. Kwitt, S. Huber, *et al.*, “Statistical topological data analysis—a kernel perspective,” in *Adv. Neural Inf. Process. Syst.*, 2015.
- [43] G. Kusano, Y. Hiraoka, *et al.*, “Persistence weighted gaussian kernel for topological data analysis,” in *Proc. Int. Conf. Mach. Learn.*, 2016.
- [44] H. Chintakunta, T. Gentimis, *et al.*, “An entropy-based persistence barcode,” *Pattern Recog.*, 2015, vol. 48, no. 2, pp. 391–401.
- [45] C. Hofer, R. Kwitt, *et al.*, “Deep learning with topological signatures,” in *Adv. Neural Inf. Process. Syst.*, 2017.
- [46] C. Hofer, R. Kwitt, *et al.*, “Learning representations of persistence barcodes,” *J. Mach. Learn. Res.*, 2019, vol. 20, no. 126, pp. 1–45.
- [47] K. Turner, S. Mukherjee, *et al.*, “Persistent homology transform for modeling shapes and surfaces,” *Inf. Inference J. IMA*, 2014, vol. 3, no. 4, pp. 310–344.
- [48] L. Kanari, P. Dlotko, *et al.*, “A topological representation of branching neuronal morphologies,” *Neuroinf.*, 2018, vol. 16, pp. 3–13.
- [49] A. Garin and G. Tauzin, “A topological ‘reading’ lesson: Classification of mnist using tda,” in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, 2019.
- [50] N. Otter, M. Porter, *et al.*, “A roadmap for the computation of persistent homology,” *EPJ Data Sci.*, 2017, vol. 6, pp. 1–38.
- [51] C. Pun, K. Xia, *et al.*, “Persistent-homology-based machine

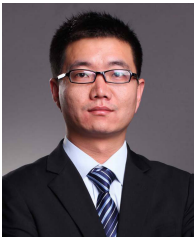
learning and its applications—A survey,” *arXiv preprint arXiv:1811.00252*, 2018.

- [52] N. Atienza, L. Escudero, *et al.*, “Persistent entropy: A scale-invariant topological statistic for analyzing cell arrangements,” *arXiv.org*, *arXiv: 1902.06467*, 2019.
- [53] I. Goodfellow, P. Jean, *et al.*, “Generative adversarial nets,” in *Adv. Neural Inf. Process. Syst.*, 2014.
- [54] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Represent.*, 2015.
- [55] Y. Li, X. Zhang, *et al.*, “Siamvgg: Visual tracking using deeper siamese networks,” *arXiv preprint arXiv:1902.02804*, 2019.
- [56] H. Yoo, L. Dai, *et al.*, “On the role of vit and cnn in semantic communications: Analysis and prototype validation,” *arXiv preprint arXiv:2306.02759*, 2023.
- [57] I. Cohen, Y. Huang, *et al.*, “Pearson correlation coefficient,” *Noise Reduct. Speech Process.*, 2009, pp. 1–4.

## Biographies



**Ni Fei** (Ph.D) is currently pursuing the doctor’s degree with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. Her research interest includes algebraic topology and deep learning.



**Li Rongpeng** (Senior Member, IEEE) is currently an Associate Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He was a Research Engineer with the Wireless Communication Laboratory, Huawei Technologies Company, Ltd., Shanghai, China, from August 2015 to September 2016. He was a Visiting Scholar with the Department of Computer Science and Technology, University of Cambridge, Cambridge, U.K., from February 2020 to August 2020. His research interest currently focuses on networked intelligence for communications evolving (NICE). He received the Wu Wenjun Artificial Intelligence Excellent Youth Award in 2021 and the sponsorship “the National Postdoctoral Program for Innovative Talents” in 2016, respectively. He serves as an Editor for *China Communications*.



**Zhao Zhifeng** (Senior Member, IEEE) received the B.E. degree in computer science, the M.E. degree in communication and information systems, and the Ph.D degree in communication and information systems from the PLA University of Science and Technology, Nanjing, China, in 1996, 1999, and 2002, respectively. From 2002 to 2004, he acted as a Post-Doctoral Researcher with Zhejiang University, Hangzhou, China, where his researches were focused on multimedia next-generation networks (NGNs) and softswitch technology for energy efficiency. From 2005 to 2006, he acted as a Senior Researcher with the PLA University of Science and Technology, where he performed research and development on advanced energy-efficient wireless router, *ad-hoc* network simulator, and cognitive mesh networking test-bed. From 2006 to 2019, he was an Associate Professor with the College of Information Science and Electronic Engineering, Zhejiang University. Currently, he is with the Zhejiang Lab, Hangzhou as the Chief Engineering Officer. His research areas include software defined networks (SDNs), wireless network in 6G, computing networks, and collective intelligence. He is the Symposium Co-Chair of ChinaCom 2009 and 2010. He is the Technical Program Committee (TPC) Co-Chair of the 10th IEEE International Symposium on Communication and Information Technology (ISCIT 2010).



**Zhang Honggang** (Fellow, IEEE) is a Professor with the Faculty of Data Science, City University of Macau, Macao, China. He was the founding Chief Managing Editor of *Intelligent Computing*, a Science Partner Journal, as well as a Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He was an Honorary Visiting Professor with the University of York, York, U.K., and an International Chair Professor of Excellence with the Université Européenne de Bretagne and Supélec, France. He has coauthored and edited two books: *Cognitive Communications: Distributed Artificial Intelligence (DAI), Regulatory Policy & Economics, Implementation* (John Wiley & Sons) and *Green Communications: Theoretical Fundamentals, Algorithms and Applications* (CRC Press), respectively. His research interests include cognitive radio networks, semantic communications, green communications, machine learning, artificial intelligence, intelligent computing, and Internet of Intelligence. He is a co-recipient of the 2021 IEEE Communications Society Outstanding Paper Award and the 2021 IEEE Internet of Things Journal Best Paper Award. He was the leading Guest Editor for the Special Issues on Green Communications of the *IEEE Communications Magazine*. He served as a Series Editor for the *IEEE Communications Magazine* (Green Communications and Computing Networks Series) from 2015 to 2018 and the Chair of the Technical Committee on Cognitive Networks of the IEEE Communications Society from 2011 to 2012. He is the Associate Editor-in-Chief of *China Communications*.