Semantic Communication System Based on Semantic Slice Models Propagation

Chen Dong, Haotai Liang, Xiaodong Xu, *Senior Member, IEEE*, Shujun Han, *Member, IEEE*, Bizhu Wang, and Ping Zhang, *Fellow, IEEE*

***Abstract*— Traditional communication systems treat messages' semantic aspects and meaning as irrelevant to communication, revealing its limitations in the era of artificial intelligence (AI), such as communication efficiency and intent-sharing among different entities. Through broadening the scope of the traditional communication system and the AI-based encoding techniques, in this manuscript, we present a novel semantic communication system, which involves the essential semantic information exploration, transmission and recovery for more efficient communications. Compared to other state-of-the-art semantic communication-related works, our proposed semantic communication system is characterized by the "flow of the intelligence" via the propagation of the model. Besides, the concept of semantic slice-models (SeSM) is proposed to enable flexible model-resembling under the different requirements of the model performance, channel situation and transmission goals. Specifically, a layer-based semantic communication system for images (LSCI) is built on the simulation platform to demonstrate the feasibility of the proposed system and a novel semantic metric called semantic service quality (SS) is proposed to evaluate the semantic communication systems. We evaluate the proposed system on Cityscapes and Open Images datasets, resulting in averaged 10% and 2% bit rate reduction over JPEG and JPEG2000, respectively. In comparison to LDPC, the proposed channel coding scheme can averagely save 2dB and 5dB in AWGN channel and Rayleigh fading channel, respectively.**

***Index Terms*— Semantic communication, models propagation, LSCI, SeSM, SS.**

I. Introduction

With the proliferation of intelligent applications, there is an unavoidable necessity for a significant number of data transmissions these days. And the transmission rate is approaching the Shannon limit once more, posing several serious challenges to the existing communication system's needs for low delay and high data transmission rate, prompting researchers to consider the next generation communication system's paradigm.

The existing source channel coding method has a big problem in a communication system that only cares about symbol efficiency and accuracy measured in bits, with limited bandwidth resources but low latency requirements, due to the continual explosion of data. By filtering redundant information and extracting the meaning of effective information, semantic communication achieves the goal of greatly reducing communication traffic [1], [2].

The semantic information is sent by the transmitter, and the meaning of the semantic information is interpreted by the receiver, as inspired by Shannon and Weaver's semantic level communication [3]. Furthermore, artificial intelligence's ability to grasp human knowledge in limited scenarios is gradually improving. As one of the most important artificial intelligence technologies, deep learning has achieved great success in comprehending language, audio, pictures, and other information sources. In natural language processing (NLP), neural network approaches typically convey the semantic aspects of the language indirectly using low-dimensional and dense vectors. A big corpus-based pre-training model (PTM) has been used in a variety of NLP tasks [4], [5], [6], [7]. The pre-trained embedding code captures the high-level meaning of the NLP task's words and phrases, avoiding the need to grasp the language on the symbols right once. Most deep learning models directly output the findings of semantic analysis, such as target recognition, target detection, segmentation, and scene understanding, when it comes to image semantic analysis [8], [9], [10], [11], [12], [13]. These tasks are capable of assigning semantic tags accurately. But it's worth noting that the recent rise of Generative Adversarial Neural Networks (GAN) [11] has improved image semantics interpretability. StyleGAN [12], for example, can control various attributes in order to generate related style pictures. Voice semantics is most commonly used in speech recognition and human-computer interaction systems, as well as speech-to-text systems. The initial technological framework of the voice recognition system, which didn't care about speaking speed or tone and merely extracted the text of the speech, has been completely transformed by Microsoft's context-dependent DNN-HMM research [13] results on large vocabulary speech recognition.

Manuscript received 15 April 2022; revised 28 July 2022; accepted 9 September 2022. Date of publication 17 November 2022; date of current version 19 December 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB2902102, in part by the National Natural Science Foundation of China under Grant 61871045, in part by the Fundamental Research Funds for the Central Universities under Project 2021RC01 and Project 2022RC15, and in part by the Academician Expert Open Fund of Beijing Smart-Chip Microelectronics Technology Company Ltd., under Project SGITZXDTKJJS2201045. *(Corresponding author: Haotai Liang.)*

Chen Dong, Shujun Han, and Bizhu Wang are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: dongchen@bupt.edu.cn; hanshujun@bupt.edu.cn; wangbizhu_7@bupt.edu.cn).

Haotai Liang is with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunication, Beijing 100876, China (e-mail: lianghaotai@bupt.edu.cn).

Xiaodong Xu and Ping Zhang are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the Department of Broadband Communication, Peng Cheng Laboratory, Shenzhen, Guangdong 518055, China (e-mail: xuxiaodong@bupt.edu.cn; pzhang@bupt.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3221948.

Digital Object Identifier 10.1109/JSAC.2022.3221948

0733-8716 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

Simultaneously, communication systems based on deep learning frameworks have been developed in recent years, and their performance has shown considerable promise in comparison to traditional communication systems. The authors in [14] suggested a communication system based on end-to-end learning and optimization of autoencoders, and their groundbreaking work has influenced the design of DL-based communication systems.

Thanks to the above enlightenment and the development of deep learning, semantic communication has become a breakthrough to improve communication efficiency again. Since the effectiveness of the source channel joint coding scheme for structured sources has been thoroughly established, the framework of source channel joint coding has currently taken the lead among semantic communication systems. For structured sources such as natural language processing, the method in [15] embeds the sentences into the semantic space to preserve the semantic information of sentences and then performs joint source and channel coding on these embeddings. Authors in [16] propose a deep learning based semantic communication system, DeepSC, for text transmission, aiming to maximize the system capacity and minimize the semantic errors by recovering the meaning of sentences rather than bit or symbol errors. To make the DeepSC affordable for Internet-of-Things devices, the authors in [17] propose a model compression algorithm, including network sparsification and quantization, to reduce the size of DL models by pruning the redundancy connections and quantizing the weights. Similarly, a DL-enabled semantic communication system for speech signals was first developed in [18], which can learn and extract speech signals and then recover them directly at the receiver from the received features.

For image structured sources, Deep JSCC [19] directly maps the image pixel values to the complex-valued channel input symbols rather than relying on explicit codes for error correction, which does not suffer from the "cliff effect". Authors in [20] use nonlinear transformation to extract the semantic features of the image source and provide auxiliary information for source channel coding, which learns the source latent representation and incorporates the entropy model as a priori on the latent representation to upgrade the Deep JSCC.

The present deep learning-based semantic model is targeted at certain structured sources. The receiving node must have the semantic models of all information sources during the actual communication process, which surely increases the node's load and power consumption. Xie and Qin proposed a distributed semantic communication system including three steps in [17]: Model Initialization/Update, Model Broadcasting, Semantic Features Upload. The debate in this work about model training and deployment in IoT nodes has prompted us to consider the current state of AI models in the semantic communication paradigm. Special sources and special channels are limitations in all proposed semantic AI models in semantic communication systems. As a result, in the face of numerous sources and channels in the physical world, future semantic communication systems will rely on the deployment of a huge number of AI models. Each node in the IoT Network is unable to deploy such a large number of AI models. Based on the explanation above, the transmission of the AI model across nodes will be a critical component of the semantic communication system. However, the propagation of the AI model will lead to the consumption of communication resources because the AI model has enormous parameters. Inspired by the work of the Scaled Video Coding (SVC) [21] and the image semantic coding method [22], the semantic AI model is composed of many AI slice-models, which can complete semantic extraction and semantic restoration according to semantic requirements.

In this paper, we consider a new communication paradigm and design a layer-based semantic coding communication system for image signals, named LSCI, which can use different semantic slice-models, named SeSM, to realize image semantic extraction, transmission and image reconstruction according to the requirements of varying semantics, accuracy and channel conditions. The main contributions of this paper are summarized as follows:

- An innovative semantic communication paradigm driven by artificial intelligence (AI) models is proposed. The semantic communication system is designed based on the Shannon-based physical layer, and the AI models are integrated into the transmitter and receiver. Model propagation is the main feature of the proposed paradigm.
- A task-driven semantic metric, semantic service quality (SS), is developed from the perspective of semantics to quantify the performance of the semantic communication system, based on the practical meaning of semantics.
- A layer-based semantic communication system for image (LSCI) for verification is designed. To achieve image semantic reduction and restoration, the basic model is trained through a multi-scale generation adversarial model. Furthermore, semantic slice-models (SeSM) are developed and designed for the first time to adapt to semantic transmission tasks in various circumstances, with different accuracy, channel situations, and transmission goals.
- Based on extensive experiments and simulations, the compression performance of the proposed image semantic communication system is better than the traditional JPEG and JPEG2000 methods. At the semantic level, semantically controllable image restoration is realized according to the receiver's requirements. We evaluate the proposed system on Cityscapes and Open Images datasets, resulting in averaged 10% and 2% bit rate reduction over JPEG and JPEG2000, respectively. In comparison to LDPC, the proposed channel coding scheme can averagely save 2dB and 5dB in AWGN channel and Rayleigh fading channel, respectively.

This paper is arranged as follows: Section 2 introduces the proposed semantic communication paradigm and the performance metric. Section 3 details the proposed layer-based semantic communication system for the image. Numerical results are presented in section 4. Finally, conclusions about our work are drawn in section 5.

II. Semantic Communication Paradigm

The architecture of the semantic communication paradigm is depicted in Fig. 1. The semantic communication paradigm

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

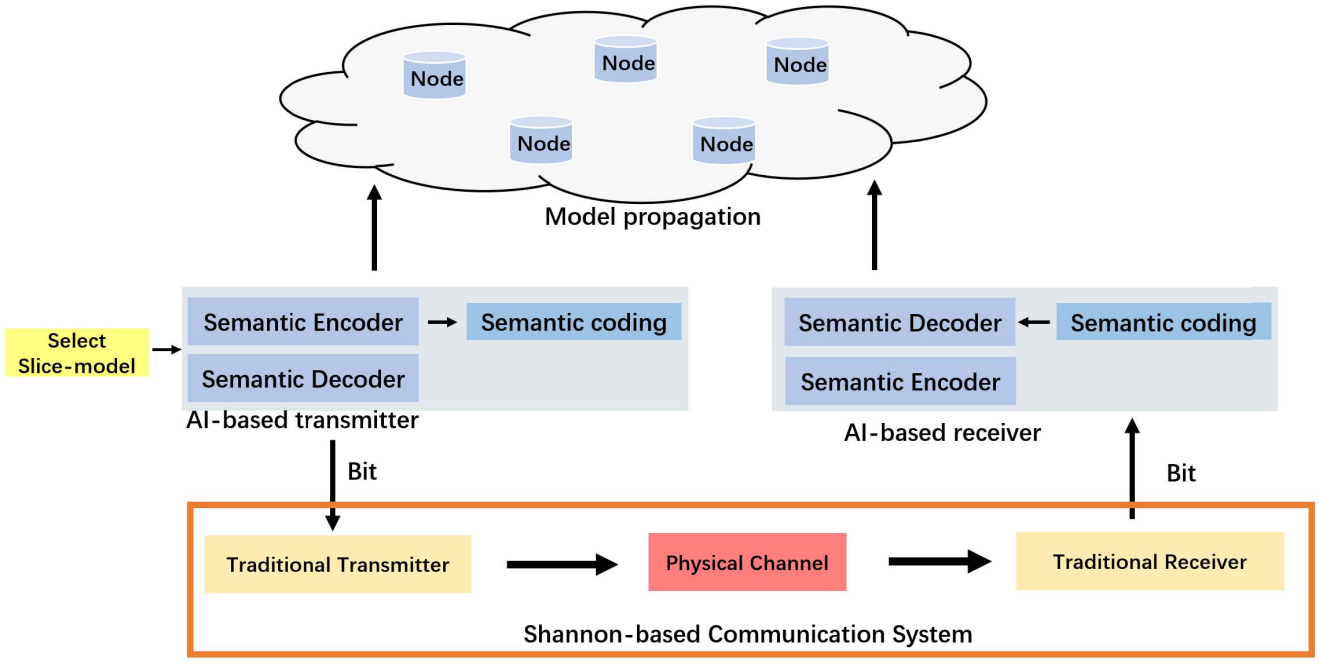


Fig. 1. The architecture of the innovative semantic communication paradigm. AI-based transmitters, receivers, and Shannon-based communication systems make up the system architecture. The semantic model, which is represented by purple blocks, and the semantic coding, which is represented by blue blocks, are transmitted as bitstreams across the physical channel. The source is recovered by the receiver using the received semantic decoder. Model propagation is an important aspect of the paradigm. Specific semantic slice-models (SeSM) will be propagated based on each node's semantic needs.

is based on the classic Shannon communication system and has no effect on the Shannon physical layer's individual modules. We embed the AI models into the transmitter and receiver of the physical layer to perform effective semantic communication, inspired by the existing semantic communication architecture proposed in [16], [17], and [18]. The AI models, on the other hand, are constrained by three factors: a specific source, a specific scene, and a specific task. For example, in the automatic driving scene, the image-based target detection model is only applicable to the image source, vehicle driving scene and target detection task. The promotion of a semantic communication system will unavoidably bring many AI models into the real physical world, where there are multiple signal sources, sceneries, and tasks. In the semantic communication system, the survival form of AI models with large parameters has become a significant barrier for the semantic communication paradigm.

As shown in Fig. 1, the AI models do not just exist in nodes in our proposed paradigm, the propagation of the AI models is the main feature of the semantic communication system. If all nodes deploy the required AI models, the burden on nodes will be tremendous. The trade-off between AI model survival time and propagation frequency at the node is an essential topic that will be tackled at the upper layer but not in this study. Instead of deploying AI models in advance at all nodes, AI model propagation will become a key element of semantic communication systems in the future.

The end-to-end optimization system has the characteristics of accessible training and higher performance in the existing artificial intelligence models of semantic communication systems. However, dividing the end-to-end AI model into sections to fulfill their specialized goals on their own is tough. For example, the image compression system in [23] is an end-to-end optimized AI model. Numerous AI models with identical parameters with the exception of the number of channels in the last layer must be trained in order to achieve images compression at various bit rates, which significantly raises the training cost and consumes more resources during model deployment. As a result, the end-to-end AI model is incompatible with the model propagation paradigm. We propose the concept of semantic slice-models (SeSM) to address the problem of high communication resource consumption induced by large parameter transfer of AI models during model propagation. Inspired by the coarse-to-fine image semantic coding model for multimedia semantic communication system proposed in [22], the model can be designed hierarchically. According to the requirements of the model performance, channel situation and transmission goals, an appropriate slice-model can be chosen to enhance the output of the basic model. It is obvious that the resource consumption of model propagation will be efficiently decreased if the model can be communicated hierarchically according to needs. For example, when a node executes a target detection task, its accuracy rate is 95%. When you wish to raise the node's target detection performance to 98%, you don't have to send the entire model; instead, send semantic slice-models based on the accuracy criteria to improve the accuracy, which can significantly cut model transmission traffic. A layer-based image semantic communication system for verification will be described in detail next section.

Due to the existence of model propagation, the proposed communication system can receive and integrate the AI models from other nodes based on various criteria such as service requirements and quality. Therefore, in terms of performance evaluation, the transmission of the semantic layer should pay attention to whether the information recovered by semantics

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

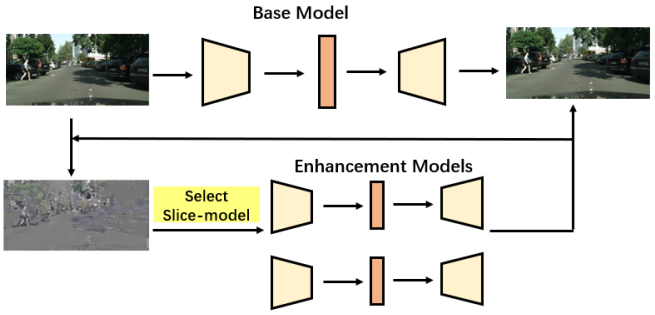


Fig. 2. Layer-based image semantic communication system consists of basic model and enhancement models, in which each enhancement model can be regarded as a semantic slice-model (SeSM) that can control accuracy and semantics.

can meet the expectations of subsequent tasks. A general quality metric of semantic service is given by,

$$SS = \frac{ST(\hat{S})}{ST(S)}, \tag{1}$$

where S denotes the unprocessed information at the transmitter, $\hat{S}$ represents the information recovered through semantics at the receiver and $ST(\cdot)$ represents the performance of the source in performing subsequent tasks. We suggest using $Sigmoid$ and other functions to map the result of $ST(\cdot)$ to $[0, 1]$, and the value of $Sigmoid(ST(S))$ should be close to 1. The indicator for selecting subsequent tasks should be that the better the performance, the higher the indicator. If SS is equal to 1, the semantic information is completely recovered. If SS is equal to 0, it fails to recover.

III. Layer-Based Image Semantic Communication System

The proposed layer-based semantic communication system for images, LSCI, will be described in detail in this section. The architecture of the image semantic compression system is inspired by [22], which is composed of the base model and enhancement models, as illustrated in Fig. 2. The base model is a semantic compression system based on multi-scale GAN technology. The role of enhancement models includes the improvement of accuracy and semantic control. Each layer of enhancement models can be regarded as the semantic slice-models (SeSM) mentioned in this paper. In addition, this section will also introduce the channel coding slice-models suitable for resisting different channel noises.

A. Base Model

The basic model realizes image semantic compression and reconstruction at a very low bit rate, as shown in Fig. 3. The basic model consists of an encoder, quantizer and generator, which extracts the basic semantic features w of the image x at the transmitter,

$$w = E(x). \tag{2}$$

The nearest neighbor quantization operation is performed on the extracted semantic features w,

$$w_i^q = argmin_j \parallel w_i - c_j \parallel, \tag{3}$$

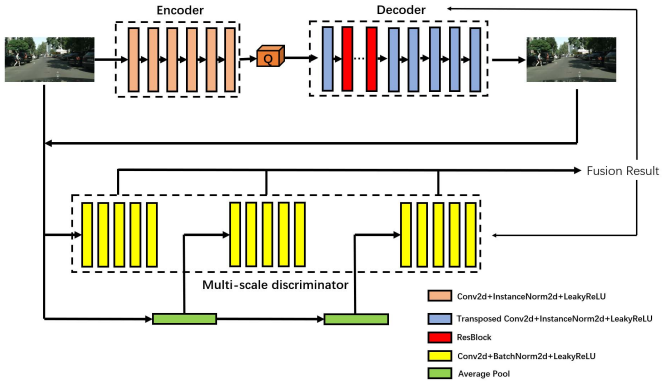


Fig. 3. The network structure and framework of the basic model. The semantic image compression system is made up of the encoder, decoder, and quantizer. The images with three different down-sampling ratios are judged three times by the multi-scale discriminator and trained against the encoder and decoder.

where the set of quantization centers is,

$$C = \{c_0, c_1, \ldots, c_j, \ldots, c_l\}. \tag{4}$$

The quantitative back-propagation gradient calculation method is consistent with [22]. The process of quantifying semantic features can be considered as the new concept of the semantic constellation, and the boundary of semantic entropy is preliminarily given, but more semantic information theory will not be discussed here. From the finite coding length $dim(w^q)$ and the length of the quantization center set $l + 1$, the boundary of semantic entropy is,

$$H(w^q) \leq dim(w^q) \log_2(l + 1). \tag{5}$$

At the receiver, the source image is restored by the generator,

$$x' = G(w^q). \tag{6}$$

The image compression and reconstruction process of the basic model is consistent with the image coding and recovery process of the autoencoder. Given the dataset X, the encoder E maps the data set X to the specific distribution P_w, and the generator G approximates the mapped w from the specific distribution P_w to the distribution P_x of X. Therefore, the optimization objective of the encoder E and the generator G is given by,

$$\min_{E,G} \lambda \mathbb{E}[d(x, G(w^q))], \tag{7}$$

where $d(\cdot)$ represents the distance between the source image x and the restored image x', including texture and perception distance, which does not depend on the adversarial loss generated by the discriminator D, L2, SSIM [24] and LPIPS are used [25] to measure,

$$d(x, x') = \parallel x - x' \parallel_2 + \alpha \cdot (1 - SSIM(x, x')) + \beta \cdot LPIPS(x, x'). \tag{8}$$

The SSIM is defined as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{9}$$

where μ_x, σ_x and σ_{xy} are the mean, stantdard deviation and cross correlation between the two patches x, y, respectively.

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

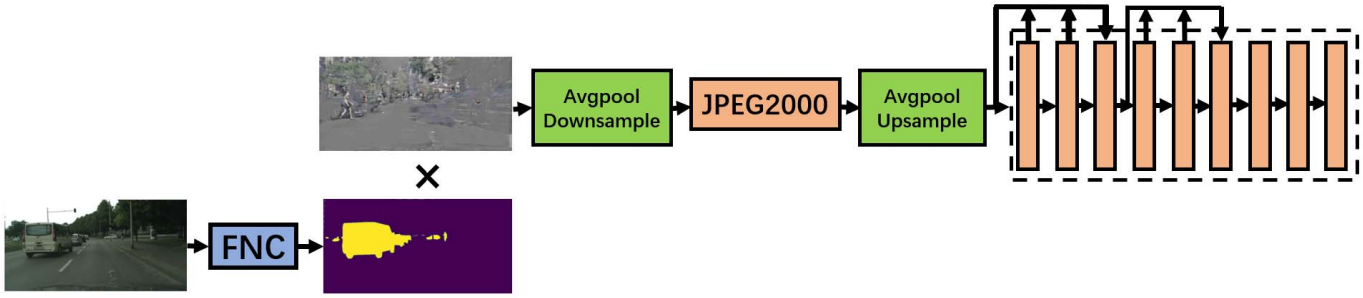


Fig. 4. A pre-trained semantic segmentation model, downsampling module, and super-resolution model are all part of the enhancement slice-models architecture. The resultant mask is multiplied by the residual image, and FNC represents the semantic segmentation model. To achieve the control of specific semantics and accuracy, the pooling layer and the accompanying super-resolution model are chosen based on the accuracy requirements.

C_1 and C_2 terms can avoid instability when the means and variances are close to zero. The LPIPS is defined as follows:

$$LPIPS(x, x') = \sum_{l=1}^{L} \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \parallel \omega_l \odot (x_{hw}^l - x_{hw}^{'l}) \parallel_2^2, \tag{10}$$

where L represents the number of the feature stacks, H, W represents the height and width of the feature map, respectively. $\odot$ denotes the the operation of scaling activation's channel wise by vector ω_l.

The discriminator D is a key aspect of the GAN network, but the system will not use it after training, therefore the discriminator's complexity will not effect the system's operating efficiency. The multi-scale discriminator is used, as illustrated in the system structure. The down sampling layers provide three images with various resolutions, which are then fed into the three discriminators, yielding the fusion decision results. The optimization objective of the discriminator is to be able to distinguish the dataset X and the recovered images, given by,

$$\underset{D}{max}\mathbb{E}[f(D(x))] + \mathbb{E}[g(D(x'))], \tag{11}$$

g where f and g are scalar functions, which help to minimize the divergence between the probability distribution of restored images x' and the probability distribution p_x of X. We use WGAN-GP [26] method,

$$f(y) = \sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij}, \tag{12}$$

$$g(y) = -\sum_{i=1}^{M} \sum_{j=1}^{N} y_{ij}, \tag{13}$$

where y represents the image patch result of $M \times N$ size obtained by the multi-scale discriminator. The encoder, generator and discriminator are jointly optimized, and the optimization objectives is as follows,

$$\underset{E,G}{min}\ \underset{D}{max}\mathbb{E}[f(D(x))] + \mathbb{E}[g(D(x'))] + \lambda\mathbb{E}[d(x, G(w^q))]. \tag{14}$$

B. Enhancement Slice-Models

Textures and details in regulated semantic areas of an image can be enhanced using enhancement slice-models, as demonstrated in Fig. 4. The basic model compresses an image at a very low bit rate, which is a lossy compression process. Inspired by the work of the Scaled Video Coding (SVC) [21] and the image semantic coding method [22], it is a delicate idea to enhance the accuracy of image restoration by using the residual parts between the reconstructed image and the original image,

$$r = x - x'. \tag{15}$$

The more information the residual part is used, the more pronounced the accuracy enhancement and the more transmission bandwidth is consumed. We control the relationship between enhanced accuracy and transmission in two ways: the selection of semantic area masks and the down sampling of residual images.

1) Semantic Area Masks: People always concentrate on different semantic domains for various contexts and tasks. For example, we focus more on the foreground than the background in video call scenarios. As a result, residual images from non-interested regions do not need to be encoded or transmitted. To segment the image, an existing semantics segmentation model is applied and the proper semantic mask is chosen to multiply the residuals according to the upper requirements,

$$M = FNC(x'), \tag{16}$$

where FNC denotes the pretrained semantics segmentation model.

2) Downsampling: Adaptive average pooling is used to reduce the resolution of residual images, and JPEG2000, an existing standard image compression technology, is used for re-compression.

$$rc = Ds(M \cdot r), \tag{17}$$

where Ds denote downsampling and JPEG2000 compression operation. The receiver adaptively sampled the received residual image to the original resolution, and used the matching super-resolution model to restore the original residual model,

$$rc' = Ds^{-1}(rc), \tag{18}$$

where Ds^{-1} denotes the operation of restoring the residual image. As shown in Fig. 4, the super-resolution model uses the idea of features reuse of DenseNet [27] to achieve the purpose of parameter reduction. The super-resolution model's main goal is to compensate for the mistake introduced by

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

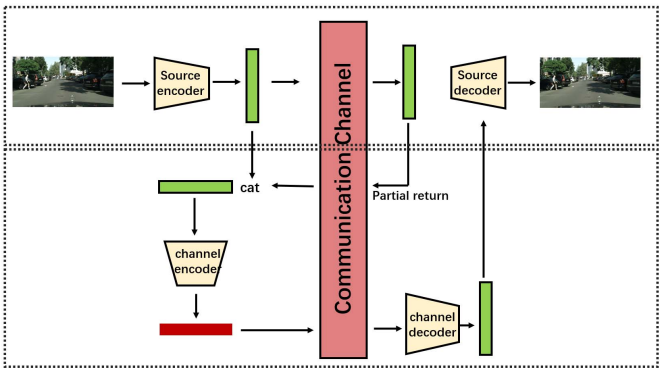


Fig. 5. Channel coding slice-models architecture includes channel slice-encoder and channel slice-decoder.

average pooling and JPEG2000 quantization. The optimization goal is as follows,

$$\min_{Sr} \parallel M \cdot r - rc' \parallel_2, \tag{19}$$

where Sr denotes the super-resolution model.

The receiver superimposes the enhanced information on the basic information to restore the semantically enhanced image,

$$x'_e = x' + rc' \tag{20}$$

C. Channel Coding Slice-Models

The influence of the channel is an essential component of any semantic communication system, and most existing semantic communication systems use the joint source-channel coding (JSCC) [19], [20] method to resist channel noise, which can effectively solve the "cliff effect" phenomenon. However, each trained semantic communication system is only applicable to the specific training channel, making it difficult to deploy semantic communication systems in the future scene. And it is also not suitable for the paradigm of semantic propagation. The channel coding slice-models are trained separately without updating the parameters of the semantic source compression model in order to enable the semantic communication system to operate with flexibility in combination with the channel model trained offline in various situations. Inspired by the idea of layered feedback in [28], the channel coding slice-models use the feedback signal to alleviate the influence of channel noise in wireless transmission, which is shown in Fig. 5. The first upper half of the architecture is the source coding part, introduced in the basic model and enhancement slice-models, which do not need to be updated. The lower part is the channel coding part, mainly composed of channel slice-encoder and channel slice-decoder.

The input image is represented as x, w^q is obtained after source coding, and y is encoded by the channel encoder. Considering Rayleigh fading additive Gaussian noise channel,

$$z = h \cdot y + n, \tag{21}$$

where z represents the complex vector corresponding to the channel, h is the channel gain satisfying Rayleigh distribution that remains constant during transmission, and n represents the independent identically distributed Gaussian noise. When we consider additive-white-Gaussian-noise-channel (AWGN), let $h = 1$. Considering the existence of noise feedback, the received part of the returned signal, which contains source signal features and noise, is spliced with semantic features,

$$wf = cat[w^q, z], \tag{22}$$

where cat is the concatenation operation. The wf with feedback signal flows via the autoencoder-based channel coding module. Channel slice-encoder Ce can add redundancy to resist noise by using channel characteristics from feedback signal. The receiver recovers the source semantic information $\hat{w^q}$ via the channel slice-decoder Cd. However, the optimization goal of Ce and Cd is not only to minimize the distance of the semantic feature but also to minimize the distance between the reconstructed images and the source images,

$$\min_{Ce,Cd} \parallel w^q - \hat{w^q} \parallel_2 + \gamma d(x, x'). \tag{23}$$

The difference between this objective function and JSCC is that it only optimizes the channel coding part and fixes the parameters of source coding. Still, the channel coding slice-models also take semantic noise into account.

IV. Experiments and Numerical Results

In this section, the dataset, the relevant training settings, and the structure of each network are introduced in detail. A series of experiments are implemented to evaluate the performance of image semantic transmission under the new paradigm of the semantic communication system in various scenarios. The performance of LSCI's image semantic transmission under the influence of different compression rates and channels is discussed. In addition, the influence of compressed SeSM on LSCI performance is discussed.

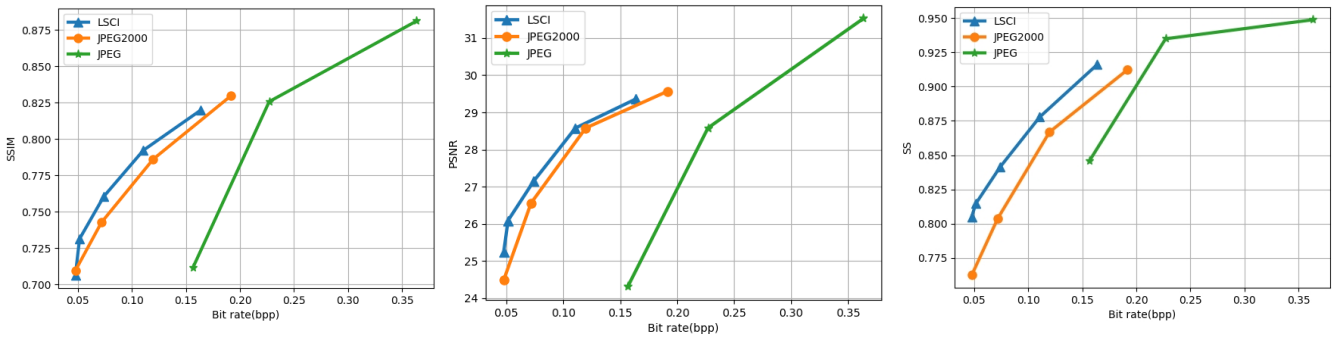
A. Setting

The layer-based semantic coding communication system for image, LSCI, is trained on the Cityscapes dataset [29] which has a training set of 2975 images, a valid set of 500 images and a test set of 1525 images. We additionally employ 188k images from the Open Images dataset, which contains a larger variety of natural images to train and evaluate the LSCI. In this paper, the images on the Cityscapes dataset and the Open Images dataset are downscaled to 512×1024 and 512×512, respectively. For loss function settings, $\alpha = 0.01$, $\beta = 0.001, \lambda = 10, \gamma = 0.01$. The models in this paper are optimized iteratively by Adam optimizer. Only one transmitter and one receiver are assumed in our experiment.

B. Baseline and Evaluation Metrics

The standard image compression technologies JPEG and JPEG2000 serve as the benchmark for the image compression comparison. An end-to-end compression framework based on convolutional neural networks, ComCNN\RecCNN, is also used as the benchmark [30]. LDPC channel coding and the JSCC are used as the baseline for performance comparison under wireless transmission channels. The degree of visual

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.



(a) SSIM (b) PSNR (c) SS

Fig. 6. The performance of the reconstructed image of JPEG, JPEG2000, and our model LSCI with varying bit rates in terms of perception and semantic distortion. The distortion performance is measured by the metrics (a) SSIM and (b) PSNR, whereas the semantic distortion performance is measured by the metrics (c) SS.

distortion between images I and K is measured using SSIM and PSNR.

The PSNR is defined as follows:

$$PSNR(x, x') = 10 \cdot \log_{10}(\frac{MAX^2}{MSE}), \tag{24}$$

where MSE is the mean squared-error between the reference image x and the reconstructed image x', and the MAX is the maximum possible value of the image pixels, which equal to 255 in our experiments.

In particular, the SS metric proposed in Section II is used to measure the degree of image semantic distortion. The driving task is the semantic segmentation task, and MIoU is used as the performance standard.

C. Model Structure

1) Base Model: The basic model structure design refers to the structure of [23]. The encoder is made up of seven convolution layers, as indicated in Table I, which extract semantic features while compressing redundant information in space. Six deconvolution layers and nine residual blocks make up the generator. Using the idea of PatchGAN [31], the multi-scale discriminator is a full convolution structure, and the structures of the three discriminators are the same.

2) Enhancement Slice-Models: Table II and Fig. 4 demonstrate the network structure of the super-resolution model in the enhanced slice-model. The first three layers' output results are spliced according to the channel dimension. Similarly, the fourth, fifth, and sixth layers' output results are spliced according to the channel dimension and serve as the sixth layer's input.

3) Channel Coding Slice-Models: The channel coding slice-model is shown in Table III. It is worth noting that the last channel number C is used to control the channel coding efficiency. We return half of the semantic features in the experiment, and the channel coding efficiency is set at $1/2$. Before going through the channel, the signal must also be treated with a reshaping operation and energy normalization.

D. Rate Perception Performance

Fig. 6 compares the performance of standard image compression technologies JPEG, JPEG2000 and LSCI in the cityscapes test set using the metrics of SSIM, PSNR and SS. The distortion performance in SSIM and PSNR, respectively, is shown in Fig. 6 (a) and (b). It is obvious that the performance values of SSIM and PSNR of LSCI are slightly higher than that of JPEG2000 and much higher than that of JPEG from Fig. 6 (a) and (b). The PSNR value of LSCI is 5 dB higher than that of JPEG at the same bit rate, but it has no discernible advantage over JPEG2000. But it is enough to show that LSCI can meet the existing compression performance standards. In addition, LSCI can realize image compression with a very low bit rate compared with JPEG.

Fig. 6 (c) shows the performance of task-driven semantic metric SS under different compression bit rates. In terms of semantic performance evaluation metric, LSCI has apparent advantages. The SS value of LSCI is around 0.3 points greater than that of JPEG2000 at the same bit rate. The goal of

TABLE I
BASE MODEL

Model	Structure
Encoder	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=128
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=256
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=512
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=1024
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=1024
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=4
Generator	ConvT, 3×3, kernel=(3,3), stride=(2,2), feature=1024
	ResBlock(1024, 1024) × 9
	ConvT, 3×3, kernel=(3,3), stride=(2,2), feature=512
	ConvT, 3×3, kernel=(3,3), stride=(2,2), feature=256
	ConvT, 3×3, kernel=(3,3), stride=(2,2), feature=128
	ConvT, 3×3, kernel=(3,3), stride=(2,2), feature=64
	ConvT, 3×3, kernel=(3,3), stride=(2,2), feature=3
Discriminator	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=128
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=256
	Conv2d, 3×3, kernel=(3,3), stride=(2,2), feature=512
	Conv2d, 3×3, kernel=(1,1), stride=(1,1), feature=1

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

Fig. 7. Visual example of images produced by LSCI along with the corresponding results for JPEG and JPEG2000 on Cityscapes. The three images are compressed and reconstructed at similar bit rates. LSCI has a similar metric value as JPEG2000, but the image is more realistic and the details are more visible.

TABLE II
SUPER RESOLUTION MODEL

Model	Structure
Super resolution model	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=32
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=32
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=32
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=128
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=3

TABLE III
CHANNEL CODING SLICE-MODEL

Model	Structure
Channel Slice Encoder	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=128
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=2C
Channel Slice Decoder	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=128
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=64
	Conv2d, 3×3, kernel=(3,3), stride=(1,1), feature=C

minimizing perceptual loss is added to the optimization target of the LSCI basic model, which impacts the decision distance of the deep learning model to some extent and enhances semantic retention in the semantic task with deep learning as the core. The blurring phenomenon caused by JPEG and JPEG2000 at low bit rate and the blocking artifact caused by block coding greatly mislead the sensitive task based on deep learning, resulting in a low semantic performance value.

Fig. 7 shows the visual examples produced by LSCI along with the JPEG and JPEG2000 with roughly 0.15 bit rate. At the low bit rate, the image compressed by JPEG not only distorts the details but also completely distorts the color and background of the road. The PSNR and SSIM values of the image compressed by JPEG2000 are similar to those of the LSCI, but after close inspection of the tree and the car logo, the image reconstructed by the LSCI is clearer, and the texture is more realistic, demonstrating that traditional image metrics can't fully demonstrate reconstruction performance. The semantic metric is a novel perspective to reflect the performance of image reconstruction under the constraint that the PSNR and SSIM values are similar.

In addition, we present a visual example of semantic control in Fig. 8. Semantically adjustable detail reconstruction can be achieved by using a mask to increase the region of interest according to the semantic criteria. Many crucial scenes, such as automated driving and video calls, require semantically regulated image reconstruction. Because the texture of numerous unimportant places does not play a role in the vehicle's judgment of the next action in the automatic driving scene, this part of the semantic features cannot be sent.

Example validation images for a side-by-side comparison of our method with JPEG2000 and ComCNN/RecCNN for images from the Open Images data set can be found in Fig. 9. The bit rates of the two image compression examples are about 0.85 and 0.77 respectively, and the SSIM of the three schemes are comparable. Overall, the color distortion of

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

(a) Enhance foreground (b) Enhance background

Fig. 8. A visual representation of semantic control. (a) and (b) are examples of using the enhancement slice-model to improve the foreground and background, respectively. In (a), the features of the vehicle and people are visible, but the background is rather blurry. The foreground in (b) is blurred as the background, such as houses and lights, becomes relatively distinct.

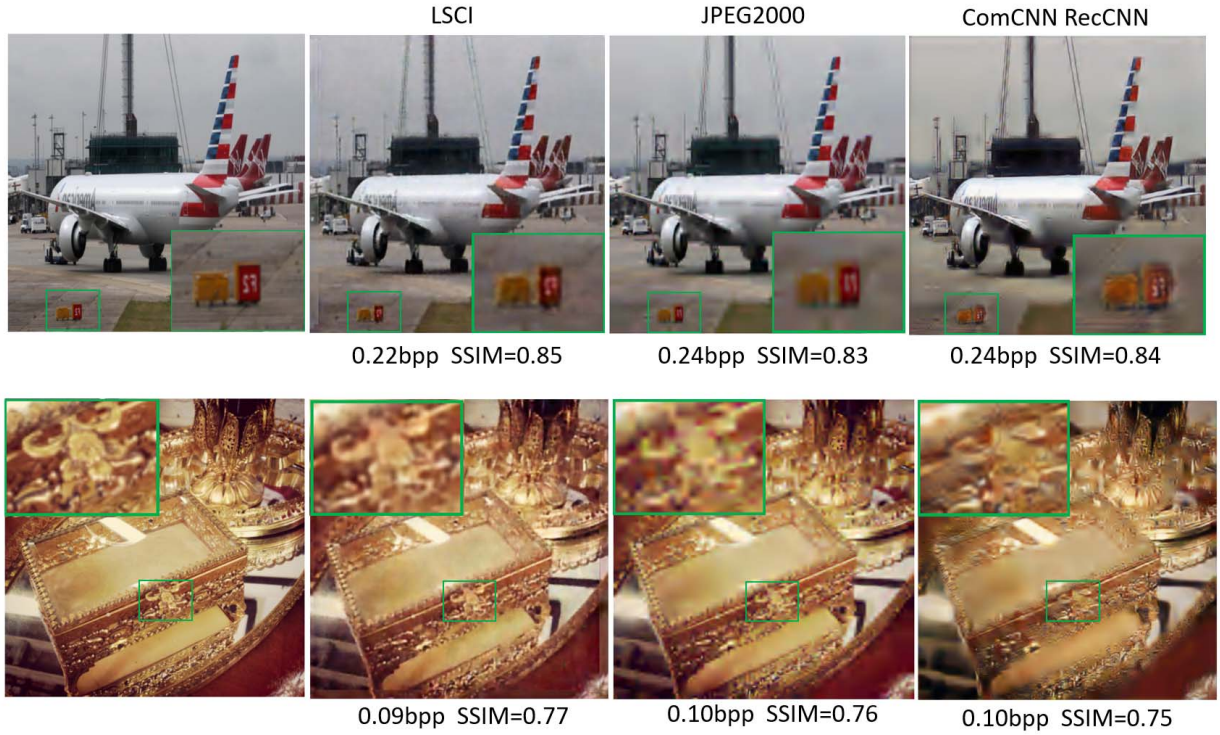


Fig. 9. Visual examples of two images from Openimage produced by the LSCI compared to JPEG2000 and ComCNN/RecCNN.

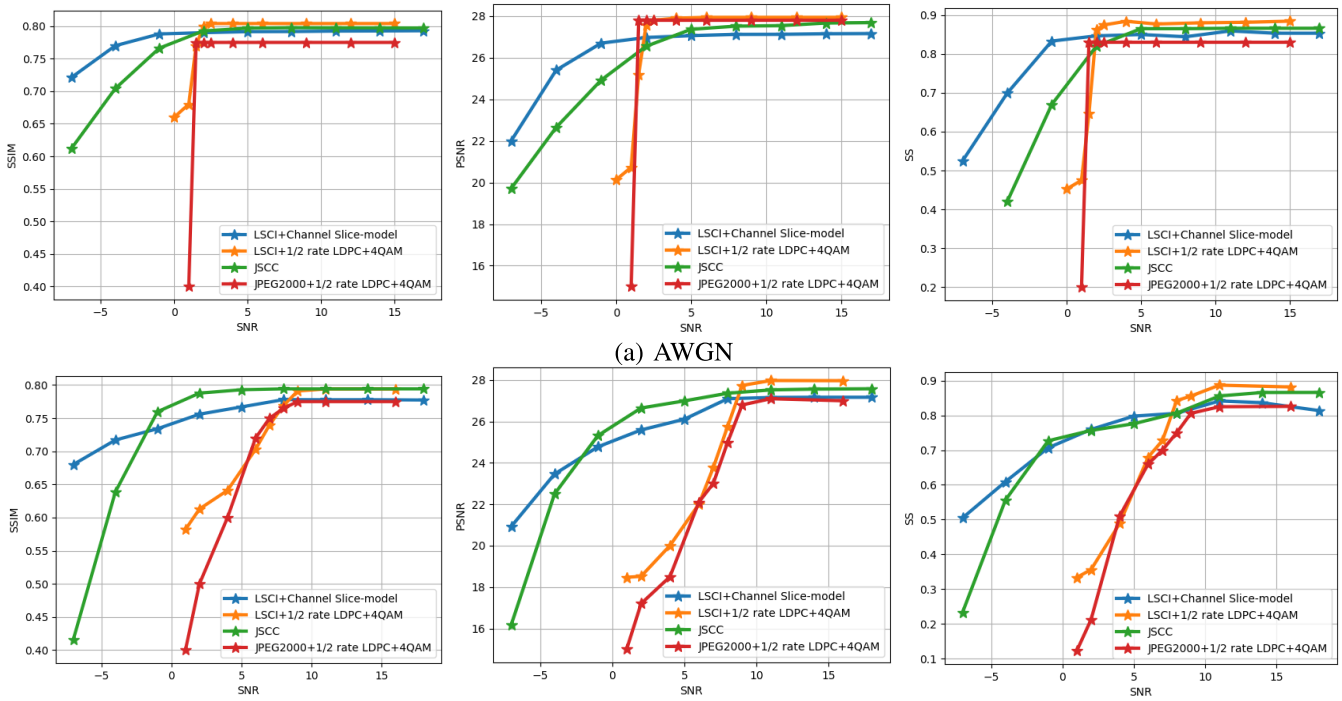
ComCNN/RecCNN is more obvious, while LSCI can retain a more similar color. Zoom in on the local details, the oil pipe in the lower left corner of the first group of images and the lock of the second group of images, and the edge details of LSCI are clearer.

E. Transmission Performance

We compare the performance of LSCI with channel coding slice-model, LDPC channel coding using the LSCI, JSCC and JPEG2000 with 1/2 rate LDPC coding in 4QAM, which are all set at a bit rate of roughly 0.1 to simulate semantic wireless transmission in AWGN and Rayleigh fading channels. It should be noted that the image bandwidth compression ratio of the JSCC scheme is set to 5, a value that is significantly lower than that of the other methods.

Under high signal-to-noise ratio (SNR) conditions, the performance of the LSCI system using LDPC channel coding and the digital transmission scheme(JPEG2000+LDPC+4QAM) are steady and slightly higher than that of the LSCI system using channel coding slice-models (Fig. III-C). However, with the decrease in signal-to-noise ratio, the performance of the

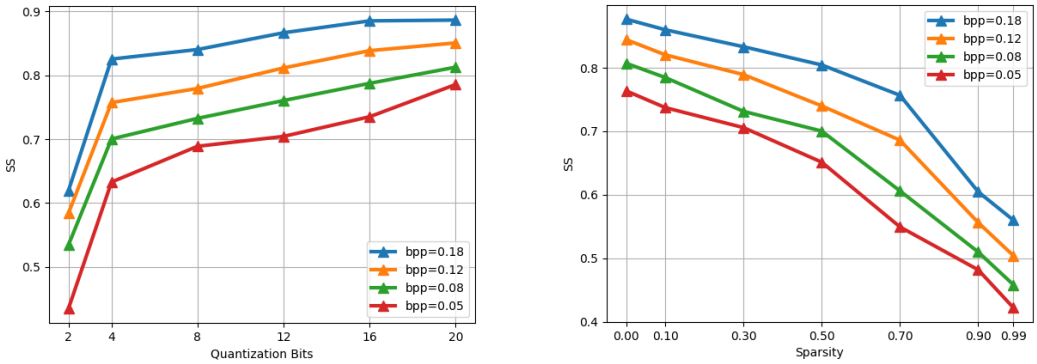
Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.



(a) AWGN

(b) Rayleigh Fading

Performance comparison of LSCI with the channel coding Slice-model, LSCI with LDPC, JSCC and JPEG2000 with LDPC coding in 4QAM in AWGN channel and Rayleigh fading channel. Channel coding slice-model effectively alleviates the cliff effect of LDPC channel coding.



(a) SeSM Quantization

(b) SeSM Pruning

Fig. 10. Effects of quantification and pruning of different enhancement Slice-models on system performance. The model compression method described in [17] is applied. SeSM with different sizes are obtained by model compression, and the model's performance varies with the size of the model parameters. Nodes in the network can choose SeSM with different volumes according to their requirements.

system using LDPC has declined rapidly, which is called the "cliff effect". The "cliff effect" occurs at roughly 3dB and 6dB in the AWGN and Rayleigh fading channels, respectively. We believe that the location of symbol errors is random and is not related to the image semantic features behind the bitstream. As a result, the fading is visible once a mistake happens.

The performance of JSCC scheme effectively alleviates the cliff effect. With the decline of SNR, there is no rapid decline in performance but a slow decline. And it can be found that the performance of LSCI declines more slowly and has stronger ability to resist low signal-to-noise ratio noise. The semantic distance between the reconstructed image and the source image is taken into account by channel slice-models in addition to decreasing the error rate of each symbol as the optimization goal. Therefore, the model will be able to learn how to preserve semantic features from semantic noise. In addition, the channel coding slice-model structure is similar to that of JSCC channel coding but with more parameters, so its performance will be better under the condition of large data training.

F. Model Compression

We believe that, in addition to picking different semantic features based on performance requirements, semantic slice-models with different parameter sizes can also be selected for propagation in the paradigm of the semantic communication system of model propagation. As a result, the impact of the enhancement slice-models' size on performance is explored in this section.

The quantization and pruning procedures described in [17] are utilized to compress the model. As shown in Fig. 10,

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
RUNNING TIME OF COMPARED SOURCE CODING METHODS TESTED ON A $256 \times 256 \times 3$ IMAGE

	CPU	GPU
LSCI	2.705s	0.255s
ComCNN RecCNN	1.646s	0.119s
JSCC	0.250s	0.023s
JPEG2000	0.090s	-
JPEG	0.011s	-

TABLE V
RUNNING TIME OF COMPARED CHANNEL CODING METHODS TESTED ON A $256 \times 256 \times 3$ IMAGE

	CPU	GPU
Channel coding slice-model	0.00977s	0.00177s
LDPC(AWGN:0dB)	22.3s	-
LDPC(AWGN:5dB)	0.119s	-
LDPC(AWGN:10dB)	0.0776s	-

not surprisingly, the fewer quantization bits, the higher the sparsity, and the weaker the model performance. It is worth noting that as the quantization bits are reduced from 20 to 4, performance gradually deteriorates. When the number of quantization bits is reduced to two, the model's performance appears to suffer. A comparable situation is depicted in Fig. 10 (b). When the sparsity is more than 70%, the model shows abrupt fading.

G. Complexity Analysis

A brief discussion of the computational complexity of the proposed LSCI is provided in this section. The running time of all baselines when dealing with a $256 \times 256 \times 3$ image in CPU and GPU are shown in Table IV and Table V. Table IV shows the running time of different source compression schemes. The main operations of the schemes based on deep learning are the 2D convolutions/transpose convolutions whose computational cost of a single convolutional layer is $\mathrm{F} \times \mathrm{F} \times \mathrm{D} \times \mathrm{K} \times \mathrm{W} \times \mathrm{H}$, where F is the filter size, K is the number of filters, D is the number of input channels and $\mathrm{W} \times \mathrm{H}$ is the size of the feature map [19]. Thus LSCI consumes more time to process images due to the deeper network structure.

Table V shows the running time of different channel coding schemes. LDPC coding schemes require different iterations for different channel conditions, so the lower the signal-to-noise ratio, the longer the running time. Obviously, the channel coding slice-model has the same running time for different noises. As a comparison, the runtime of our proposed channel coding slice-model significantly outperforms the LDPC schemes.

V. CONCLUSION

In this paper, we propose a novel communication paradigm driven by artificial intelligence (AI) models, which is mainly characterized by the propagation of the model. The reasons and functions of model propagation as this paradigm's main feature are described in detail. Besides, the concept of semantic slice-models (SeSM) is proposed to reduce the communication consumption caused by model propagation and enable flexible model-resembling under the different requirements of the model performance, channel situation and transmission goals. Specifically, a layer-based semantic communication system for image (LSCI) is designed to verify the feasibility of the new paradigm. In addition, a task-driven semantic metric, semantic service quality (SS), is proposed from the perspective of semantics to measure the performance of the semantic communication system.

Experiments on the Cityscapes and Openimages datasets reveal that our proposed semantic communication system, LSCI, and SeSM, can perform effective image compression and reconstruction at a low bit rate. Enhancement slice-models can effectively improve the accuracy of the image semantic region of interest, and the channel coding slice-models can effectively alleviate the cliff effect of low SNR.

REFERENCES

[1] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 96–102, Jun. 2021.
[2] P. Zhang et al., "Toward wisdom-evolutionary and primitive-concise 6G: A new paradigm of semantic communication networks," *Engineering*, vol. 8, pp. 60–73, Jan. 2021.
[3] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[5] J. D. M.-W. C. Kenton and L. K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
[6] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in translation: Contextualized word vectors," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6297–6308.
[7] J. Sarzynska-Wawer et al., "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Res.*, vol. 304, pp. 114–135, Oct. 2021.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 91–99.
[10] F. Isensee, P. F. Jäger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," 2019, *arXiv:1904.08128*.
[11] I. Goodfellow et al., "Generative adversarial nets," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Nov. 2020.
[12] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.
[13] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 30–42, Jan. 2011.
[14] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2016, pp. 223–228.
[15] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2326–2330.
[16] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.
[17] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, Jan. 2021.
[18] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.
[19] E. Bourtsoulatze, D. B. Kurka, and D. Gunduz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.

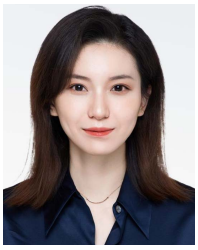
[20] J. Dai et al., "Nonlinear transform source-channel coding for semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2300–2316, Aug. 2022.
[21] I. E. Richardson, *H. 264 MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*. Hoboken, NJ, USA: Wiley, 2004.
[22] D. Huang, X. Tao, F. Gao, and J. Lu, "Deep learning-based image semantic coding for semantic communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.
[23] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool, "Generative adversarial networks for extreme learned image compression," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 221–231.
[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[25] 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/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
[26] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[27] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869*.
[28] D. B. Kurka and D. Gunduz, "DeepJSCC-F: Deep joint source-channel coding of images with feedback," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 178–193, Dec. 2020.
[29] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3213–3223.
[30] F. Jiang, W. Tao, S. Liu, J. Ren, X. Guo, and D. Zhao, "An end-to-end compression framework based on convolutional neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 3007–3018, Oct. 2017.
[31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

Chen Dong received the B.S. degree in electronic information sciences and technology from the University of Science and Technology of China, Hefei, China, in 2004, the M.Eng. degree in pattern recognition and automatic equipment from the University of Chinese Academy of Sciences, Beijing, China, in 2007, and the Ph.D. degree from the University of Southampton, U.K., in 2014. After a post-doctoral researcher experience in Southampton, he used to work in Huawei Device Company Ltd., China. Since 2020, he has been working with the Beijing University of Posts and Telecommunications (BUPT). His research interests include applied mathematics, relay systems, channel modeling, and cross-layer optimization. He was a recipient of the Scholarship under the U.K.–China Scholarships for Excellence Programme and the Best Paper Award at the IEEE VTC 2014.

Haotai Liang received the B.E. degree from the School of Electronic Information Engineering, Shenzhen University (SZU), Shenzhen, China, in June 2022. He is currently pursuing the M.E. degree in information and telecommunications engineering with the Beijing University of Posts and Telecommunications (BUPT).

Xiaodong Xu (Senior Member, IEEE) received the B.S. degree in information and communication engineering and the master's degree in communication and information system from Shandong University in 2001 and 2004, respectively, and the Ph.D. degree in circuit and system from the Beijing University of Posts and Telecommunications (BUPT) in 2007. He is currently a Professor at BUPT and a Research Fellow of the Peng Cheng Laboratory. He has coauthored nine books and more than 120 journals and conference papers. He is also the inventor or co-inventor of 51 granted patents. His research interests cover semantic communications, intellicise communication systems, moving networks, and mobile edge computing and caching.

Shujun Han (Member, IEEE) received the bachelor's degree in communication engineering and the master's degree in electronics and communication engineering from Zhengzhou University, Zhengzhou, China, in 2013 and 2016, respectively, and the Ph.D. degree in information and telecommunications engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2020. From 2018 to 2019, she visited the 5G Innovation Centre (5GIC), University of Surrey, U.K. She is currently a Post-Doctoral Researcher at BUPT. Her research interests are in the next-generation mobile communication systems, with a current emphasis on intelligent and security-driven edge-device collaboration in massive ultra-reliable low latency communications.



Bizhu Wang received the B.Eng. degree in telecommunications engineering from the Beijing University of Posts and Telecommunications in 2016 and the Ph.D. degree in electronic engineering from the Queen Mary University of London in 2020. In 2021, she joined the Beijing University of Posts and Telecommunications as a Lecturer. Her current research interests include next generation networks, semantic communication, intent-based networking, machine learning, and in-built security and privacy protection.

Ping Zhang (Fellow, IEEE) received the Ph.D. degree in circuits and systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 1990. He is currently a Professor with the Beijing University of Posts and Telecommunications, the Director of the State Key Laboratory of Networking and Switching Technology, the Director of the Department of Broadband Communication, Peng Cheng Laboratory, a member of IMT-2020 (5G) Experts Panel, and a member of Experts Panel for China's 6G development. He served as the Chief Scientist of National Basic Research Program (973 Program), an Expert with the Information Technology Division of National High-Tech Research and Develoment Program (863 Program), and a member of Consultant Committee on International Cooperation of National Natural Science Foundation of China. His research interests mainly focus on wireless communication. He is an Academician of the Chinese Academy of Engineering.

Authorized licensed use limited to: JILIN UNIVERSITY. Downloaded on November 14,2024 at 03:14:26 UTC from IEEE Xplore. Restrictions apply.