

doi:10.13756/j.gtxyj.2026.240107.

光通信系统与网络技术

余继科,孙晓川,杨硕晗,等.鲁棒语义传输:跨域协作的联合源信道编码[J].光通信研究,2026(1):240107.

Yu J K, Sun X C, Yang S H, et al. Robust Semantic Transmission: Joint Source Channel Coding for Cross-Domain Collaboration[J]. Study on Optical Communications, 2026(1): 240107.

## 鲁棒语义传输:跨域协作的联合源信道编码

余继科<sup>1,2</sup>,孙晓川<sup>1,2</sup>,杨硕晗<sup>1,2</sup>,李莹琦<sup>1,2</sup>

(1. 华北理工大学 人工智能学院,河北 唐山 063200; 2. 河北省工业智能重点感知实验室,河北 唐山 063200)

**摘要:**【目的】实现高质量、高可靠的信息传输是语义通信领域的重要目标。深度联合源信道编码(DeepJSCC)作为一种有效的语义通信方法,已经取得了一定的进展。然而,现有基于DeepJSCC的语义通信方法在低信噪比(SNR)环境下仍然面临信道干扰导致的语义失真问题,难以达到理想的语义传输质量,从而制约了通信的可靠性和准确性。为解决这一痛点,文章旨在设计一种新型的DeepJSCC框架,有效抑制信道噪声对语义信息的干扰,提高语义通信系统的鲁棒性。【方法】文章所提新型DeepJSCC框架融合了空间域和频域两种视角,实现了对语义信息全面、高效地表达和传输。在空间域,该框架对图像进行全局与局部语义特征的高效提取,确保语义信息在编码阶段得到完整的保留;在频域,则对频率成分进行精准识别,能够准确分辨对解码端任务影响最大的频率分量。从而充分增强核心语义频率分量的表达,同时抑制噪声频率,大幅减少信道噪声导致的语义失真。【结果】文章在公开数据集上评估了所提方法的性能表现,并将其与现有的先进语义通信方法进行对比。实验结果表明,与现有DeepJSCC方法相比,文章所提新框架能够在恶劣的通信环境(如低SNR)中显著提升语义信息的传输准确性,有效缓解语义失真对通信质量的影响,从而增加了语义通信系统的鲁棒性。【结论】文章所提新型DeepJSCC框架融合了空间域和频域的优势,通过创新的编码策略实现了高效的语义特征提取和核心语义频率分量增强,从而极大地提高了语义通信在恶劣环境下的鲁棒性,为语义通信系统的可靠性和高质量传输提供了新的解决方案。

**关键词:**深度联合源信道编码;语义通信;频域处理

中图分类号:TN929

文献标志码:A

## Robust Semantic Transmission: Joint Source Channel Coding for Cross-Domain Collaboration

YU Jike<sup>1,2</sup>, SUN Xiaochuan<sup>1,2</sup>, YANG Shuohan<sup>1,2</sup>, LI Yingqi<sup>1,2</sup>

(1. Artificial Intelligence Institute, North China University of Science and Technology, Tangshan 063200, China;

2. Hebei Key Laboratory of Industrial Intelligent Sensing, Tangshan 063200, China)

**Abstract:** 【Objective】 Realizing high quality and highly reliable information transmission is an important goal in the field of semantic communication. Deep Joint Source Channel Coding (DeepJSCC) has emerged as an effective method for semantic communication and has made significant progress. However, existing DeepJSCC-based semantic communication methods still face the problem of semantic distortion caused by channel interference in low Signal-to-Noise Ratio (SNR) environments, making it difficult to achieve the desired quality of semantic transmission, thereby limiting the reliability and accuracy of communication. To address this issue, this paper aims to design a novel DeepJSCC framework that effectively suppresses the interference of channel noise on semantic information, improving the robustness of semantic communication systems. 【Methods】 The proposed DeepJSCC framework integrates both spatial and frequency domain perspectives, enabling comprehensive and efficient representation and transmission of semantic information. Specifically, in the spatial domain, the framework efficiently extracts global and local semantic features from images, ensuring that semantic information is fully preserved during the encoding stage. In the frequency domain, it precisely identifies the frequency components, enabling accurate discrimination of the frequency components that have the most significant impact on the decoding task. Consequently, it enhances the expression of core semantic frequency components while suppressing the noise frequencies, significantly reducing the semantic distortion caused by channel noise. 【Results】 We evaluated the performance of the proposed method on public datasets and compared it with existing advanced semantic communication methods. The experimental results demonstrate that, compared to existing DeepJSCC methods, the proposed framework can significantly improve the accuracy of semantic information transmission in adverse communication environments (such as low SNR), effectively mitigating the impact of semantic distortion on communication quality, thereby increasing the robustness of semantic communication systems. 【Conclusion】 The proposed DeepJSCC framework integrates the advantages of both spatial and frequency domains. Through an innovative coding strategy, it achieves efficient semantic feature extraction and enhancement of

收稿日期:2024-06-04; 修回日期:2024-08-19; 纸质出版日期:2026-02-10

基金项目:国家自然科学基金青年科学基金资助项目(62401206)

作者简介:余继科(2000-),男,安徽阜阳人。硕士,主要研究方向为无线通信和深度学习。

孙晓川(1983-),男,山东莱阳人。副教授,博士,主要研究方向为无线流量预测和深度学习。

作者贡献声明:余继科与孙晓川贡献等同,为共同第一作者。

通信作者:李莹琦,副教授。E-mail:liyinqi@ncst.edu.cn

© Editorial Office of *Study on Optical Communications*. This is an open access article under the CC BY-NC-ND license.

core semantic frequency components, greatly improving the robustness of semantic communication in adverse environments. This method complements existing DeepJSCC methods rather than replacing them, providing a new solution for the reliability and high-quality transmission of semantic communication systems. Our work provides a new solution for the reliability and high-quality transmission of semantic communication systems.

**Key words:** DeepJSCC; semantic communication; frequency domain processing

## 0 引言

传统数字通信技术面临巨大的传输速率限制,难以满足现代通信的需求<sup>[1]</sup>。语义通信被视为突破该瓶颈的创新解决方案<sup>[2]</sup>。有别于传统传输原始数据的方式,语义通信在发送端仅传输核心语义信息,在接收端依赖先验知识对语义信息进行重构。这种创新传输方式在带宽受限的环境中能显著提高数据传输效率和质量,特别是对于高语义密度的图像数据<sup>[3]</sup>。

联合源信道编码(Joint Source Channel Coding, JSCC)解决了信息论中信源信道分离编码为渐近最优的问题,是实现图像语义通信的有效手段。然而,传统的 JSCC 策略往往依赖于对源数据的概率建模,忽视了数据中潜在的语义属性。随着深度学习技术在图像识别、语义解析和图像重建等领域展现出的卓越能力<sup>[4]</sup>,现代研究已经逐步将深度学习技术引入 JSCC,开发深度学习驱动的深度 JSCC (DeepJSCC)模型,深入挖掘和利用数据中潜在的语义信息,推动语义通信的发展<sup>[5-8]</sup>。然而,现有的语义通信方案在低信噪比(Signal-to-Noise Ratio, SNR)环境下仍存在语义失真问题,难以满足通信系统对高可靠性传输的严格要求。

为了缓解语义通信中信道噪声带来的语义失真问题,本文提出了一种混合域的 DeepJSCC 框架。该框架在语义编解码过程中对图像进行空间域和频域的联合处理,能够有效去除噪声频率,显著减少由语义失真所导致的不利影响。

## 1 系统模型

本文考虑了基于 DeepJSCC 的语义通信框架,该框架在结构上由编码器  $E_s$ 、解码器  $D_s$  和物理信道组成。源端的语义编码器负责提取图像中的语义信息并将其映射到低维语义空间。具体地,给定输入原始图像  $x \in \mathbb{R}^{C \times H \times W}$ ,语义信息  $S$  的编码过程可表示为

$$S = E_s(x, r, R), \quad (1)$$

式中: $C$ 、 $H$  和  $W$  为图像的通道、高度和宽度; $r$  为反映信道质量的 SNR; $R$  为输入数据的压缩比(Com-

pression Ratio, CR),反映带宽限制。经过编码后, $S$  在受发送端功率限制的物理信道上传输,传输过程可以表示为

$$\tilde{S} = hS + n, \quad (2)$$

式中: $\tilde{S}$  为经过物理信道传输后在接收端获得的信号表示; $h \in \mathbb{C}$  为复值信道增益系数, $\mathbb{C}$  为复数集合; $n$  为平均功率为  $\sigma^2$  的加性高斯白噪声(Additive White Gaussian Noise, AWGN)。最后,在解码端,根据接收到的语义信息重构出与原始图像高度相似的重建图像  $\hat{x}$ ,可表示为

$$\hat{x} = D_s(\tilde{S}, r). \quad (3)$$

该语义通信系统的目标为最小化原始图像  $x$  与重建图像之间的平均失真。两者之间的失真使用均方误差(Mean Square Error, MSE)  $\mathcal{L}$  进行评估,可表示为

$$\mathcal{L} = \frac{1}{N} \sum_{k=1}^N d(x_k, \hat{x}_k), \quad (4)$$

式中: $x_k$  和  $\hat{x}_k$  分别为第  $k$  张原始图像和相应的重构图像; $d(x_k, \hat{x}_k)$  为 MSE 分布; $N$  为样本个数。通过损失函数的设计,实现了整个语义通信系统的端到端优化训练。

## 2 混合域的鲁棒语义通信模型

### 2.1 混合域语义编码器

结构上语义编码器包含 3 个阶段的语义提取,每阶段包含 3 个串联的功能模块,包括用于进行 SNR 适应的 SNR 调制模块(SNR-aware Modulation Block, SMB)、频域语义提取模块(Frequency Domain Semantic Extraction Module, FDSEM)以及空间域的全局和局部语义增强(Global and Local Semantic Enhancement, GLSE)模块,语义通信系统模型总体结构如图 1 所示。以第 1 阶段语义提取为例,具体如下所示。

① SMB:该模块的作用是使模型能够适应各种 SNR 条件,并且能够根据图像的语义内容和信道 SNR 信息动态调整所传输的语义特征表示,其结构如图 2 所示。具体的,给定输入图像  $x \in \mathbb{R}^{3 \times H \times W}$ ,首先将 SNR 信息  $\varphi$  与输入进行拼接,得到包含 SNR 信息的图像语义特征  $s$ ,可表示为

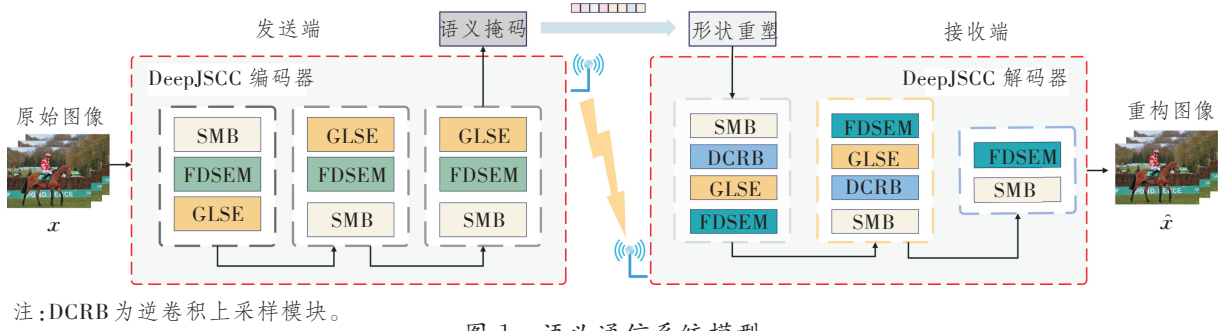


图 1 语义通信系统模型

Figure 1 The semantic communication system model

$$s = \text{concat}[\phi, p(x)] \in \mathbb{R}^{257 \times H \times W}, \quad (5)$$

式中： $\text{concat}$  为拼接操作； $p(\cdot)$  为自适应池化层，用于获取图像的全局语义表示。然后，通过压缩激励块 (Squeeze-and-Excitation Block, SE Block)，对拼接后的语义特征进行重标定，通过学习每个通道的重要性来实现 SNR 对语义特征的调整。具体来说，调制过程可表示为

$$s' = \delta(\text{Conv}(\text{SE}(s))) \otimes x, \quad (6)$$

式中： $s'$  为经过 SNR 自适应调制后的语义特征； $\delta$  为 Sigmoid 激活函数； $\text{Conv}(\cdot)$  为  $1 \times 1$  的卷积 (Conv) 层，用于调整特征图的通道数至目标输出通道数； $\text{SE}(\cdot)$  为 SE Block 函数； $\otimes$  表示逐元素相乘。通过这种方式，使得 SNR 信息跟图像语义内容融合，共同调整特征表示，从而在各种 SNR 条件下都能保持良好的性能。

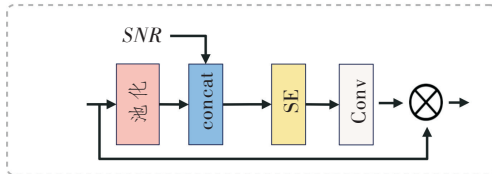


图 2 SNR 适应模块的网络结构

Figure 2 Network architecture for SNR adaptation

② FDSEM: 频域为图像中的语义信息提供了另一种表征视角。不同的频率分量对应着图像中不同的结构和纹理信息。低频分量主要反映图像的整体轮廓和平滑区域，而高频分量则对应于边缘等细节特征。通过频域分析，模型可以聚焦于反映图像本质语义属性的主要频率成分，同时忽略掉无关的高频噪声，从而提取出更加紧凑有效的特征表示，提高语义通信模型对信道噪声的鲁棒性。FDSEM 结构如图 3 所示，包含频域注意力和频率选择两个模块。具体的，拼接后的语义特征  $s'$  首先经过频域注意力模块，通过  $1 \times 1$  卷积与  $3 \times 3$  深度可分离卷积 (Depthwise Separable Convolution, Dconv) 得到  $F_q, F_k$  和  $F_v$ 。然后，将快速傅里叶变换 (Fast Fourier Transform, FFT)

应用于特征  $F_q$  和  $F_k$ ，并估计  $F_q$  和  $F_k$  在频域的相关性  $A$ ，从频域中识别和强调任务相关的语义特征，从而提高语义的表达能力。相关性  $A$  可表示为

$$A = F^{-1}(F(F_q) \overline{F(F_k)}), \quad (7)$$

式中： $F^{-1}(\cdot)$  为逆 FFT (IFFT)； $F(\cdot)$  为 FFT； $\overline{F(\cdot)}$  为共轭转置运算。最后，通过以下方法聚合最终的语义特征  $s_p$ ，可表示为

$$s_p = \text{GDN}(A)F_v + s', \quad (8)$$

式中， $\text{GDN}(\cdot)$  为基于局部分裂归一化的归一化变换函数。

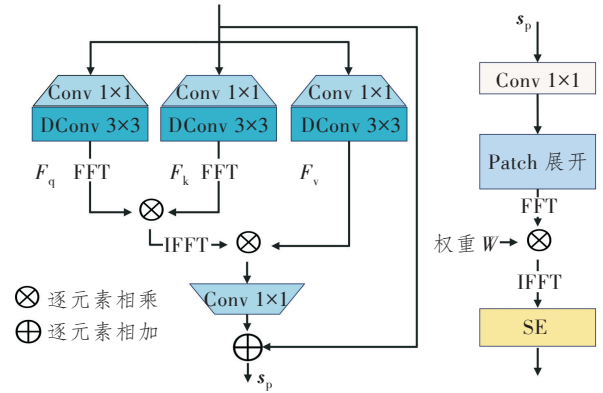


图 3 FDSEM 的网络结构

Figure 3 Network structure of the FDSEM

进一步， $s_p$  通过频率选择模块自适应确定应保留哪些关键语义信息。该模块在联合图像专家组 (Joint Photographic Experts Group, JPEG) 压缩算法的激励下，引入了一个可学习的量化矩阵  $W$ ，并通过 JPEG 压缩的逆方法来学习它，以此优化频域特征的保留策略，确保关键语义信息的保留，去除冗余信息。其具体实现为

$$\begin{cases} X_1 = \text{Conv}(s_p) \\ X_1^f = F(P(X_1)) \\ X_2 = F^{-1}(WX_1^f) \\ S_p = G(P^{-1}(X_2)) + s_p \end{cases}, \quad (9)$$

式中： $X_1$  为对输入特征进行卷积后的特征表示； $X_1^f$

为频域变换后的特征;  $X_2$  为频域加权后转回空域的特征;  $S_p$  为经频域选择与残差连接后的最终特征;  $P(\cdot)$  和  $P^{-1}(\cdot)$  分别为 JPEG 压缩方法中的 patch 展开和折叠操作;  $G(\cdot)$  为高斯误差线性单元函数 (Gaussian Error Linear Unit, GELU)。

③ GLSE 模块: 其结构如图 4 所示。首先将频域提取后的语义特征  $S_p$  分为全局语义特征  $S_{global}$  和局部语义特征  $S_{local}$ , 两者分别经过全局分支和局部分支, 最后再将两者的输出进行融合, 实现更加细致和全面的语义表达。全局分支采用注意力机制建模长程语义信息。给定输入特征  $S_{global}$ , 通过 FC 层对其进行线性映射得到  $Q_g, K_g$  和  $V_g$ 。由于注意力特征输出的空间尺度由  $Q_g$  决定, 若对  $Q_g$  进行下采样将会降低输出分辨率并引入额外插值误差, 而对  $K_g, V_g$  下采样可在保持空间输出尺度不变的同时显著减少计算量, 因此, 本文仅对  $K_g, V_g$  进行下采样, 表示为

$$\begin{cases} Q_g, K_g, V_g = FC(S_{global}) \\ S_{global} = Attention(Q_g, Pool(K_g), Pool(V_g)) \end{cases}, \quad (10)$$

式中:  $Pool(\cdot)$  为池化函数;  $FC(\cdot)$  为全连接函数 (FC);  $Q_g, K_g$  和  $V_g$  分别为通过 FC 进行线性变换后得到的查询、键和值矩阵;  $Attention$  为相关性计算函数, 用于聚合全局信息。

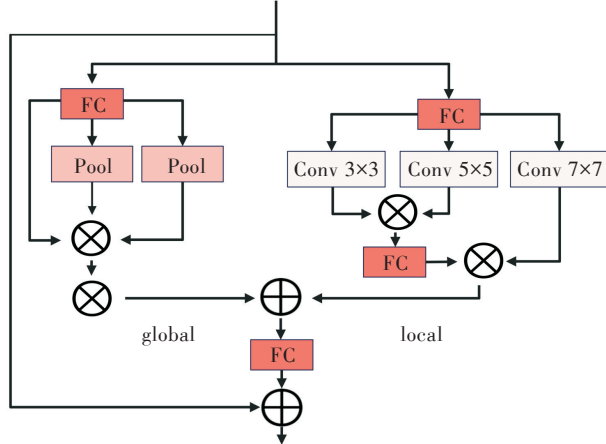


图 4 GLSE 模块网络结构

Figure 4 The network architecture of GLSE module

局部分支则是处理更细粒度的语义特征。其首先通过一个 FC 对局部特征进行扩展, 然后通过 Dconv 对不同的查询、键和值进行语义提取, 具体为

$$\begin{cases} Q, K, V = FC(S_p^{local}) \\ Q_1, K_1, V_1 = Dconv(Q, K, V) \\ A_{tm} = FC(\gamma(FC(Q_1 \otimes K_1))) \\ S_{local} = A_{tm} \otimes V_1 \end{cases}, \quad (11)$$

式中:  $\gamma$  为 Swish 激活函数;  $A_{tm}$  为计算得到的注意力权重;  $Q_1, K_1$  和  $V_1$  为由局部特征通过全连接与深度可分离卷积提取得到的查询、键和值特征。最后, 将全局语义信息  $S_{global}$  和局部语义信息  $S_{local}$  进行合并, 得到一个经过全局和局部增强的完整语义  $S_c$ , 表示为

$$S_c = FC(Concat(S_{global}, S_{local})). \quad (12)$$

第 1 阶段属于浅层语义提取, 不足以完整表征图像语义。因此, 浅层语义特征  $S_c \in R^{256 \times H \times W}$  会再次进行两阶段的语义提取。在第 2 阶段,  $S_c$  会首先经过一个下采样层调整其特征图尺寸, 使其更加紧凑高效, 而后再次经过 SMB、FDSEM 和 GLSE 模块得到更丰富的语义表示  $S_{c2} \in R^{256 \times \frac{H}{2} \times \frac{W}{2}}$ 。最后,  $S_{c2}$  会再次经过与第 2 阶段相同的流程进行最后的语义提取, 得到更深层次的语义表示  $S_e \in R^{48 \times \frac{H}{4} \times \frac{W}{4}}$ 。

为了适应传输需求的变化,  $S_e$  将通过一个掩码过程自适应调整待传输语义码的长度, 以节省带宽资源。掩码生成过程与文献[9]类似。首先, 初始化一个全 0 的掩码向量, 其长度等于语义码的最大长度, 然后根据给定 CR, 将前 CR 比例的向量置 1, 其余元素置 0, 从而生成二值掩码矩阵  $m$ 。将  $m$  与  $S_e$  逐元素相乘, 然后将  $S_e$  中为 0 的部分直接丢弃, 则完成了压缩过程, 压缩后的语义码将通过信道传输, 得到噪声干扰后的语义码  $\tilde{S}$ 。

## 2.2 语义解码器

接收端的语义解码器负责精准地解码接收到的语义信息流, 重建与原始图像语义内容高度一致的视觉表征。语义解码器的网络结构与编码端类似, 同样包含 SMB、FDSEM 和 GLSE 模块, 确保其能够精准识别对语义表达至关重要的频率成分, 从而最大限度地消除信道噪声造成的语义失真, 使图像细节得以高保真还原。此外, 设计了 DCRB, 其网络结构如图 5 所示, 其中, 通道注意力机制 (Channel Attention, CA) 用于自适应建模各通道特征的重要性权重。该模块对经过信道传输的语义信息进行高质量的空间维度还原, 使视觉内容在语义和细节层面均能实现最优重建。由于解码器与编码器的网络结构设计相似, 故在此不再赘述。

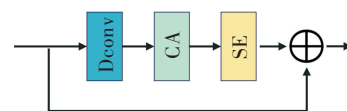


图 5 DCRB 网络结构

Figure 5 The network architecture of DCRB

### 3 实验分析

#### 3.1 实验设置

① 数据集：本实验采用 Cifar10 数据集评估所提算法的性能。Cifar10 包含 60 000 张  $32 \times 32$  的彩色图像，其中 50 000 张用于训练，另外 10 000 张用于测试。

② 评价指标：为了与现有的一些图像语义通信方法进行对比，本文采用图像传输领域广泛使用的评价指标，即峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)、结构相似指数 (Structural Similarity Index, SSIM) 和学习感知图像块相似度 (Learned Perceptual Image Patch Similarity, LPIPS)。这些指标的计算方法为

$$\begin{cases} \text{PSNR} = 10 \log_{10} \frac{\text{MAX}^2}{\text{MSE}} \text{ (dB)} \\ \text{SSIM} (x, \hat{x}) = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \\ \text{LPIPS} = d(x, \hat{x}) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|\omega_l \odot (\hat{y}_{hw}^l - y_{hw}^l)\|_2^2 \end{cases}, \quad (13)$$

式中： $\text{MAX}$  为图像像素的最大可能值； $\text{MSE}$  为原始图像与重建图像之间的 MSE； $\mu_x$  和  $\mu_{\hat{x}}$  分别为原始图像和重建图像的像素均值； $\sigma_{x\hat{x}}$  为原始图像和重建图像之间的 MSE； $\sigma_x^2$  和  $\sigma_{\hat{x}}^2$  分别为原始图像和重建图像之间的像素方差； $c_1$  和  $c_2$  分别为保持稳定性的常数； $\|\cdot\|_2$  为向量的欧几里得范数； $l$  为特征层的索引； $W_l$  为第  $l$  个特征层的宽度； $H_l$  为第  $l$  个特征层的高度； $h$  和  $w$  为空间位置索引； $y$  和  $y_0$  表示重建图像与原始图像提取的特征表示。

③ 基准模型：为了展现所提模型的卓越性能，本文将其与当前最先进的图像语义通信模型进行了全面对比分析，包括 WITT<sup>[10]</sup>、SemViT<sup>[9]</sup> 和 Deep-

JSCC-V<sup>[11]</sup>方法。此外，为了进一步验证基于语义提取的图像传输策略的有效性，也将所提方法与传统数字通信方案进行了对比，包括采用高效图像 (Better Portable Graphics, BPG) 编码、低密度奇偶校验码 (Low-Density Parity-Check, LDPC) 信道编码以及正交幅度调制 (Quadrature Amplitude Modulation, QAM) 的实际系统，以及基于 BPG 与信道容量的理论性能上限 (BPG+Capacity) 方法。

④ 实验配置：实验环境：在 Linux 操作系统的基础上，配备 24 GB 显存的 NVIDIA GeForce GTX 3090 图形处理单元 (Graphics Processing Unit, GPU)。软件方面的设置：选择 Anaconda3 作为主要的开发环境，搭载了广泛使用的深度学习框架 Pytorch。训练参数：epoch 数为 200，批大小为 64，学习率为 0.000 1，采用 Adam 优化器和 MSE 损失。

#### 3.2 实验结果分析

图 6 所示为所提图像语义通信方法在各种动态通信环境中的传输质量评估。实验涵盖了一系列 SNR 和信道带宽比 (Channel Bandwidth Ratio, CBR)，其中 SNR 区间为  $[-3, 0, 3, 6, 9, 12, 15, 18, 21, 24]$  dB，CBR 的取值范围设定为  $[1/12, 1/6, 1/4, 1/3]$ ，并在 AWGN 信道和瑞丽 (Rayleigh) 衰落信道两种信道条件下进行了验证。由图可知，随着 SNR 和 CBR 的增大，图像的传输质量显著提高，表明信道条件对通信传输质量有着直接且显著的影响。在低带宽和高噪声的恶劣环境下，所传输的图像仍能保持 25 dB 以上的 PSNR 性能和 0.9 以上的 SSIM 性能，保持了良好的传输质量，体现了其在实际通信环境中处理噪声的卓越能力。

图 7 所示为  $R=1/6$  时，所提图像语义通信模型与若干基准模型的传输质量对比。由实验结果可

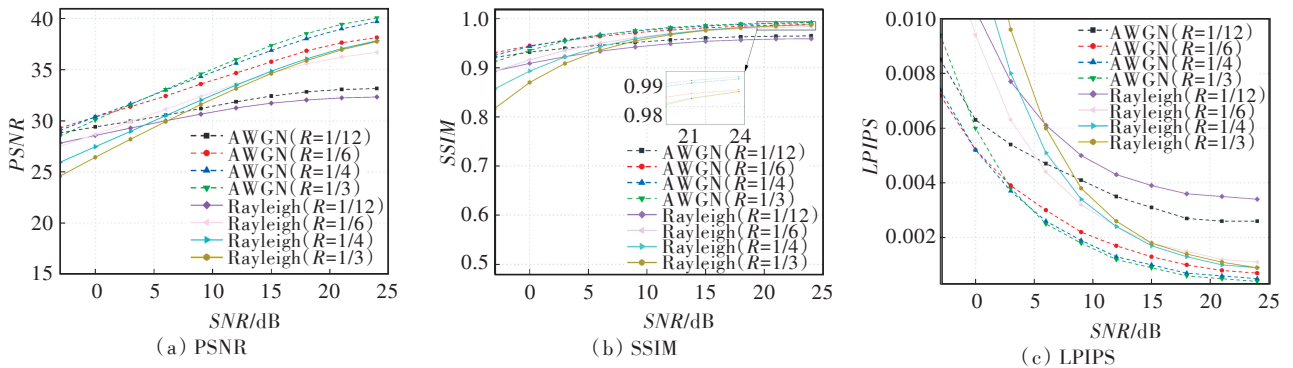
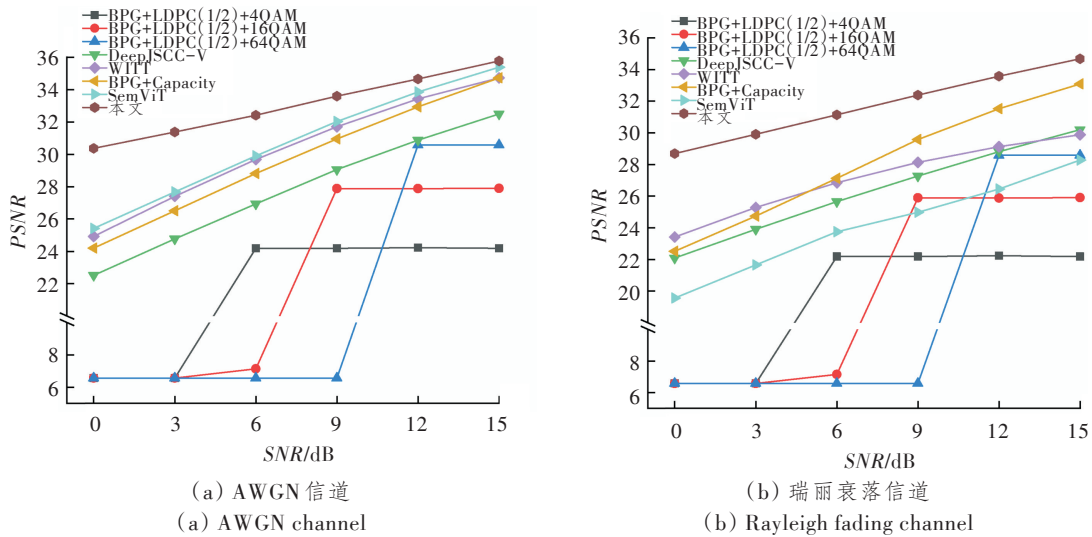


图 6 所提模型在各种通信环境下的传输质量评估

Figure 6 Transmission quality assessment of the proposed model in various communication environments

知,本文所提方法在各种 SNR 条件下的图像传输质量均表现出最佳的传输性能。同时,从单一信道上的对比结果可知,随着 SNR 的降低,各种基准模型与本文所提模型的传输性能差异逐渐增大。在 AWGN 信道下,次优的 SemViT 模型在 SNR=15 dB 时,与本文的性能差异为 0.4 dB,而在 SNR=0 dB 时,性能之间差异增加至 4.95 dB,其他模型的差异则更为明显。这说明本文的框架能够在恶劣的噪声环境下更好地抵御信道干扰,实现鲁棒的语义信息传输。进一步,从两种信道条件之间的传输质量对

比中可以看出,一些在 AWGN 信道上传输质量较优的基准模型(如 WITT 和 SemViT),在瑞丽信道中的 PSNR 却急剧下降。相比之下,本文所提框架在瑞丽信道条件下依然表现出稳定的传输质量,充分证明了其在复杂多变信道环境中的鲁棒性。值得一提的是,在采用 BPG+LDPC+QAM 等传统方法进行图像传输时,由于“悬崖效应”,当信道 SNR 较低时,传统方案的图像传输性能会急剧下降,在接收端甚至无法解码出原始图像数据。



注:Capacity 为信道容量。

图 7 R=1/6 时所提模型与各种基准模型的传输质量对比

Figure 7 Comparison of transmission quality of the proposed model with various benchmark models when R=1/6

为了进一步证明本文所提模型能够应对恶劣的信道环境,本文对模型在 SNR<0 dB 时的传输性能进行了对比分析,实验结果如表 1 所示。由表可知,本文所提模型具备性能优势,此模型 PSNR 始终保持在 20 dB 以上,在极低 SNR 下也没有表现出大幅

的性能下降。其他模型则难以有效处理噪声影响,在高噪声环境下难以实现可靠的语义传输,且随着噪声的降低,与此模型的性能差异也在加大,从而进一步证明了本文所提模型的鲁棒性。

表 1 SNR<0 dB 时的 PSNR 性能对比

Table 1 Comparison of PSNR performance for SNR<0 dB

| SNR/dB | BPG+LDPC+QAM | DeepJSCC-V | BPG+Capacity | WITT     | SemViT   | 本文       |
|--------|--------------|------------|--------------|----------|----------|----------|
| -3     | 6.57         | 20.351 0   | 21.991 7     | 22.430 0 | 23.125 3 | 25.866 7 |
| -6     | 6.57         | 18.683 6   | 20.061 0     | 21.735 6 | 20.870 2 | 25.447 2 |
| -9     | 6.57         | 17.522 7   | 19.606 5     | 19.042 8 | 19.240 4 | 24.993 2 |
| -12    | 6.57         | 16.489 9   | 19.103 2     | 17.581 3 | 17.635 3 | 24.410 9 |
| -15    | 6.57         | 15.542 8   | 18.605 2     | 16.368 8 | 16.372 3 | 23.621 4 |
| -18    | 6.57         | 14.919 4   | 18.029 7     | 15.386 9 | 15.362 0 | 22.818 3 |

图 8 所示为 R=1/6 且 SNR=0 dB 的条件下,本文所提模型与各种基准模型的图像传输效果可视化对比。由图可知,本文提模型重构出的图像与原图相似度最高,展现出了在低 SNR 环境下对噪声

干扰的出色处理能力。相比之下,WITT 模型在处理噪声方面存在一定局限性,某些图像中的物体信息由于噪声影响而无法明确重现,尤其是在物体轮廓和背景较为模糊时。至于 DeepJSCC-V 模型,在

低 SNR 的环境下表现更为不佳,其重建出的图像普遍存在模糊,失真较为严重。而 BPG+LDPC+QAM 的传统图像传输方法在低 SNR 下则不能还原图像。这些可视化结果直观地展现了本文所提模型的有效性和其在图像语义通信中处理复杂信道条件的潜在应用价值。

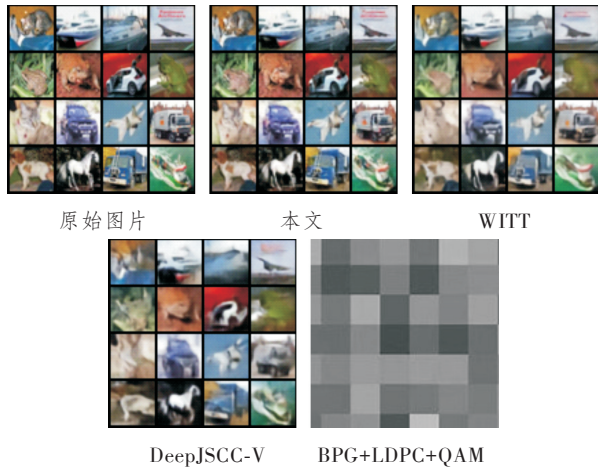


图 8 在  $R=1/6, SNR=0$  dB 情况下所提模型与各种基准模型的传输质量可视化比较

Figure 8 Visualised comparison of transmission quality of the proposed model with various benchmark models at  $R = 1/6$  and  $SNR = 0$  dB

上述实验从传输质量方面证明了方案的有效性。然而,除了性能之外,实施成本和计算复杂度对于该方法的实际应用同样重要。表 2 所示为本文所提方法与其他几种基准模型的存储开销对比。PSNR 和 SSIM 性能是在 SNR 范围为  $[0, 3, 6, 9, 12, 15, 18]$  dB 和  $R$  范围为  $[1/12, 1/6, 1/4, 1/3]$  条件下的平均性能,涵盖了 28 种通信条件的设置,更能全面反映本文所提方法的性能表现。

表 2 模型存储开销对比

Table 2 Comparison of model storage overhead

| 模型         | 参数量/<br>Mbit | 大小/<br>MB    | PSNR/<br>dB     | SSIM           |
|------------|--------------|--------------|-----------------|----------------|
| WITT       | 13.700       | 52.27        | 31.352 9        | 0.946 5        |
| SemViT     | 13.800       | 52.70        | 31.235 2        | 0.942 1        |
| DeepJSCC-V | 11.700       | 48.67        | 28.824 8        | 0.893 4        |
| 本文         | <b>3.760</b> | <b>40.00</b> | <b>33.374 1</b> | <b>0.964 8</b> |

由表可知,在保证传输质量的同时,本文所提的框架在计算复杂度方面也展现出了优异的性能。具体来说,相较于传输质量次优的 WITT 模型,所提方法平均 PSNR 提升 2.02 dB,但模型大小却降低了 12 MB,参数量则减少了一半以上。这充分说明,本文所提框架不仅在性能方面取得了显著提升,

在降低计算复杂度和减小资源占用等实际应用要求上也有着卓越的优势。

## 4 结束语

本文提出了一种融合频域和空间域分析的图像语义通信方法。通过充分的实验证明,与当前主流的图像语义通信方法相比,其在图像重建任务上表现优异,在各种复杂通信环境下都能实现良好的传输质量。尤其是在低 SNR 情况下,能够很好地消除噪声干扰带来的语义失真问题,显著提高传输图像的可辨识度和整体质量。

然而,当前的语义通信技术旨在支持各种智能任务的通信需求。本文所提方法尽管在图像重建方面取得了显著成绩,但在其他智能任务上的适用性和性能尚未进行深入探究。因此,我们对未来的研究方向进行了规划,旨在进一步完善和优化所提语义通信模型,使其不仅限于图像重建,而是能够广泛适用于各类下游任务,实现更加灵活和智能化的通信功能。

### 参考文献:

- [1] 朱一鸣,林显浩,迟楠,等. 面向5G/6G移动前传网络的关键技术[J]. 光通信研究, 2025(3): 240078.  
Zhu Y M, Lin X H, Chi N, et al. Key Technologies for 5G/6G Mobile Fronthaul Network [J]. Study on Optical Communications, 2025(3): 240078.
- [2] Zhu T, Peng B, Liang J, et al. How to Evaluate Semantic Communications for Images with ViTScore Metric [J]. IEEE Transactions on Cognitive Communications and Networking, 2024, 10(5): 1744–1758.
- [3] 张聿远,赵海涛,魏急波,等. 可扩展分层译码的语义通信系统[J]. 通信学报, 2023, 44(8): 49–60.  
Zhang Y Y, Zhao H T, Wei J B, et al. Extensible Hierarchical Codec Semantic Communication System [J]. Journal on Communications, 2023, 44(8): 49–60.
- [4] 刘凯,孙鹏,童世博,等. 语义驱动的颜色恒常决策算法[J]. 电讯技术, 2024, 64(4): 537–545.  
Liu K, Sun P, Tong S B, et al. A Semantic-Driven Color Constancy Decision Algorithm [J]. Telecommunication Engineering, 2024, 64(4): 537–545.
- [5] Erdemir E, Tung T Y, Dragotti P L, et al. Generative Joint Source-Channel Coding for Semantic Image Transmission [J]. IEEE Journal on Selected Areas in Communications, 2023, 41(8): 2645–2657.
- [6] Huang D, Tao X, Gao F, et al. Deep Learning-based Image Semantic Coding for Semantic Com-

- munications [C]//2021 IEEE Global Communications Conference (GLOBECOM). Madrid, Spain: IEEE, 2021: 9685667.
- [7] Nguyen L X, Tun Y L, Tun Y K, et al. Swin Transformer-based Dynamic Semantic Communication for Multi-User with Different Computing Capacity [J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 8957–8972.
- [8] Wu T, Chen Z, He D, et al. CDDM: Channel Denoising Diffusion Models for Wireless Semantic Communications[J]. IEEE Transactions on Wireless Communications, 2024, 23(9): 11168–11183.
- [9] Yoo H, Dai L, Kim S, et al. On the Role of ViT and CNN in Semantic Communications: Analysis and Prototype Validation[J]. IEEE Access, 2023, 11: 71528–71541.
- [10] Yang K, Wang S, Dai J, et al. WITT: a Wireless Image Transmission Transformer for Semantic Communications [C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, Greece:IEEE, 2023: 10094735.
- [11] Zhang W, Zhang H, Ma H, et al. Predictive and Adaptive Deep Coding for Wireless Image Transmission in Semantic Communication [J]. IEEE Transactions on Wireless Communications, 2023, 22(8): 5486–5501.