

面向无人机航拍图像传输的信道弹性编解码方法

张中强¹, 邱奇^{1,2}, 苗雅杰^{1,2}, 黎斌¹, 王野¹, 石光明¹

(1. 鹏城实验室, 广东 深圳 518055; 2. 哈尔滨工业大学(深圳)信息学部, 广东 深圳 518055)

摘要: 无人机(UAV, unmanned aerial vehicle)是低空经济网络的核心组成部分, 实现无人机航拍图像高效传输在空地通信中至关重要。在无人机执行任务的过程中, 由于飞行区域的变化, 信道状态快速波动, 如信噪比(SNR, signal-to-noise ratio)和传输速率。为适应这些变化, 提出了一种面向无人机航拍图像传输的信道弹性编解码方法。该方法包含轻量化特征提取模块、特征增强模块、传输速率自适应模块和相应的解码模块。轻量化特征提取模块利用残差块的局部细节提取能力与轻量 Mamba 的全局上下文建模能力, 可有效地获取航拍图像丰富的语义特征。特征增强模块根据信道状态 SNR, 增强关键语义特征。传输速率自适应模块根据信道传输速率, 自适应地调整传输特征的数量。在航拍图像数据集(AID, aerial image dataset)上的实验结果表明, 该方法比现有方法具有更高的传输精度。

关键词: 无人机航拍图像传输; 信噪比自适应; 速率自适应; 语义通信

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.2096-3750.WLW25161

A channel elastic encoding and decoding method for unmanned aerial vehicle image transmission

Zhang Zhongqiang¹, Qiu Qi^{1,2}, Miao Yajie^{1,2}, Li Bin¹, Wang Ye¹, Shi Guangming¹

1. Pengcheng Laboratory, Shenzhen 518055, China

2. College of Informatics, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

Abstract: Unmanned aerial vehicle (UAV) is recognized as the core component of the low-altitude economy network. It is crucial to achieve efficient image transmission in air-to-ground communication. During mission execution, rapid channel fluctuations are induced by changes in the flight area, including variations in signal-to-noise ratio (SNR) and transmission rate. To adapt to these changes, a channel elastic encoding and decoding method for UAV image transmission was proposed. This method included a lightweight feature extraction module, a channel feature enhancement module, a transmission rate adaptive module, and symmetric decoding modules. The lightweight feature extraction module contained residual blocks and the mobile Mamba, which could quickly extract local detail features and global context. It could effectively obtain rich semantic features from aerial images. The feature enhancement module enhanced key seman-

收稿日期: 2025-10-22; 修回日期: 2026-05-12

通信作者: 邱奇, qq.qiu.qi@gmail.com

基金项目: 国家自然科学基金资助项目 (No. 62501333); 宁夏回族自治区自然科学基金资助项目 (No. 2026AAC030469); 甘肃省自然科学基金资助项目 (No. 26JRRN002); 移动信息网络国家科技重大专项 (No. 2024ZD1300700); 贵州省基础研究计划项目 (No. Qiankehejichu-MS[2026]269); 陕西省自然科学基金基础研究计划项目 (No. 2025JC-YBQN-924); 鹏城实验室重大攻关项目 (No. PCL2024A01, No. PCL2025A03)

Foundation Items: The National Natural Science Foundation of China (No. 62501333), the Natural Science Basic Research Project of the Ningxia Hui Autonomous Region (No. 2026AAC030469), the Natural Science Basic Research Project of Gansu Province (No. 26JRRN002), the National Science and Technology Major Project-Mobile Information Networks (No. 2024ZD1300700), the Basic Research Program of Guizhou Province (No. Qiankehejichu-MS[2026]269), the Natural Science Basic Research Project of Shaanxi Province (No. 2025JC-YBQN-924), the Peng Cheng Laboratory Major Project (No. PCL2024A01, No. PCL2025A03)

tic features according to the channel state SNR. The transmission rate adaptive module adjusted the size of transmitted features based on the channel transmission rate. Extensive experiments on the aerial image dataset (AID) demonstrate that the proposed method achieves higher transmission accuracy than the state-of-the-art methods.

Key words: UAV aerial image transmission, SNR adaptation, rate adaptation, semantic communication

0 引言

无人机在低空经济网络场景下具备广阔的应用前景,如货物快递、交通监控、应急响应。所有这些应用都需要超可靠和低时延的通信。其中,航拍图像快速回传是无人机的重要基础任务之一。

目前无人机无线通信方法大多采用信源信道分离编码方式。其中,信源编码通过去除源数据冗余来提高数据的有效性。常用的信源编码方法包括联合图像专家小组(JPEG, Joint Photographic Experts Group)、更好的可移植图形(BPG, Better Portable Graphics)和联合图像专家小组2000(JPEG2000, Joint Photographic Experts Group 2000)。信道编码采用错误检查或纠错编码技术来提高传输稳定性,例如,低密度奇偶校验码(LDPC, low density parity check code)、极化码和Turbo码。传统方法受低信噪比或高误码率的影响很大。这个问题被称为“悬崖效应”,通常存在于大多数无线通信系统中^[1]。现有的通信方式通常受到功耗和通信带宽的限制,导致传输速率低下。因此,如何快速有效地传输无人机航拍图像仍然是一个挑战。

为应对上述挑战,许多研究者提出了语义通信方法。Xiao等^[2]提出了一种基于语义认知通信网络的图像快速传输方法。Boursoulatze等^[3]提出了一种基于卷积神经网络的信源和信道联合编码(JSCC, joint source and channel coding)技术,在CIFAR10数据集上,实现了比基于分离方案更好的传输性能。Fu等^[4]提出了一种结合全局和任务相关文本信息的语义通信方法。Dai等^[5]提出了一种在线学习联合信源和信道编码方法。Kurka和Gündüz^[6]开发了一种基于自动编码器的具有信道输出反馈的信源信道联合编码方法。Xu等^[7]提出了基于自适应深度学习的图像传输JSCC架构。Yang^[8]提出了一种用于无线图像传输的自适应深度JSCC方案,该方案使用单个网络模型调整传输速率。

现有的语义通信方法在相对稳定的信道状态或传输速率下表现良好。然而,由于飞行区域的变化,

信道状态信噪比(SNR, signal-to-noise ratio)和传输速率出现快速波动,无人机难以长期保持高效传输^[9]。因此,亟须研究在不稳定的信道条件和传输速率下航拍图像的高效传输。本文提出了一种面向无人机航拍图像传输的信道弹性编解码方法,该方法包括轻量化特征提取模块、特征增强模块、传输速率自适应模块和相应的解码模块。该方法能够自适应信道状态信噪比和传输速率的变化,实现航拍图像的高效传输,本文的主要贡献如下。

1) 设计了一个轻量化特征提取模块,该模块包含两个残差块、两个词元(token)生成块、两个轻量Mamba块^[10]和一个自注意力层。残差块可以提取局部细节特征,轻量Mamba块可以捕获全局上下文信息,提取丰富的语义特征。

2) 提出了一种特征增强模块,该模块包含内卷积操作和信噪比调整模块,可以根据信道状态SNR增强关键语义特征。

3) 设计了一个传输速率自适应模块,包含速率调整块和速率掩码块。该自适应模块可以根据传输速率自适应地调整传输特征的数量。

在航拍图像数据集(AID, aerial image dataset)^[11]上的实验结果表明,在不同的信道状态SNR和传输速率下,本文所提方法获得了比现有方法更高的传输精度。

1 相关研究工作

本节主要介绍了面向任务的通信方法,联合信源信道编解码方法,以及所提方法与现有方法的对比分析。

1.1 面向任务的通信方法

在物联网时代,通信对数据的传输起着重要作用。香农在其经典工作中给出了通信的基本定义^[12]。与传统通信方式不同,面向任务的通信是一种新颖的通信范式,它可以传递数据的语义信息^[13],而非原始数据。

当前,一些学者将面向任务的通信方法应用到文本和语音传输任务。Xie等^[14]提出了面向文本传

输的语义通信系统,该系统对信道变化具有更强的鲁棒性,并且能够在低信噪比条件下实现更高的文本相似度。Xie和Qin^[15]进一步提出了基于深度学习的轻量级分布式语义通信系统,用于低复杂度的文本传输。该系统设计了信道状态信息辅助训练模块,降低了信道衰落对传输的影响。Weng和Qin^[13]提出了一种面向语音信号的语义通信系统。Xie等^[16]进一步设计了面向任务的多模态多用户语义通信系统。该系统可以利用来自多个用户的多模态数据之间的相关性,对信道变化表现出比传统通信系统更强的鲁棒性。Jiang等^[17]提出了一种混合式自动重复请求方法来减少语义传输错误,该方法可以极大地减少文本传输所需的比特数。Kutay和Yener^[18]提出了一种文本语义量化和压缩方法,该方法利用句子嵌入和语义失真度量来保留语义特征。Zhou等^[19]开发了基于Transformer的语义通信系统,该系统可以通过自适应循环策略灵活地传输具有不同语义信息的文本。Bokaei等^[20]提出了一种利用联合信源信道编码和深度神经网络的极低时延可配置语音传输系统。上述工作主要面向文本和语音进行语义通信方法的设计。

在图像传输方面,Jankowski等^[21]提出了两种基于数字和模拟通信的替代方案,在数字方法中,该方案设计了一种基于深度神经网络辅助的面向检索的图像压缩方案,在模拟JSCC方法中,特征向量被直接映射到信道符号中。Hu等^[22]提出了一种鲁棒的端到端语义通信系统框架,并设计了特征重要性模块提高系统的鲁棒性。Xing等^[23]提出了一种基于具有动态模型聚合的联邦学习的深度语义通信系统。Luo等^[24]提出了面向一对多无线图像传输场景的自适应深度联合信源信道编码方案,信道增益自适应模块动态调整潜在空间内各个信道的显著性。Li等^[25]设计了一种从粗到细信道自适应JSCC框架,旨在处理无线信道的显著波动和快速变化。Yan等^[26]提出了一种语义自适应特征提取框架,通过允许用户根据其信道条件选择不同的子语义组合,显著提高了带宽效率。

一些研究者设计了多模态语义通信方法。Zhang等^[27]提出了一种统一的语义通信系统,可以利用多种模态的数据执行不同的任务,该系统可以针对不同的任务动态调整传输符号的数量。Jiang等^[28]提出了一个基于大语言深度学习的多模态语义通信框架,该框架可以将多模态语言模型和大语言模型集成到

语义通信系统中。

1.2 联合信源信道编码

近年来,基于深度学习的JSCC方法已经成为语义通信中的一个重要方向。Dai等^[29]提出了一种语义引导的信源信道编码方法,该方法同时利用数据语义分集和无线信道分集来提高整个系统的性能。Xu等^[30]提出了一种基于注意力深度学习的JSCC方法,该方法根据信道SNR动态调整信源编码中的压缩比和信道编码速率。Kurka和Gündüz^[31]提出了一种基于深度JSCC的灵活带宽图像传输方法。Zhang等^[32]提出了自适应深度编码框架,可在给定目标传输质量要求下实现灵活码率优化的预测。

Erdemir等^[33]提出了一种JSCC方案,该方案利用深度生成式模型进行无线图像传输。Yang和Kim^[8]提出了一种自适应速率控制的无线图像传输深度联合信源信道编码方案,该方案可以根据信道状况和图像内容动态控制传输速率。Tung等^[34]提出了一种基于输入字母表的端到端优化JSCC方法,来进行无线图像传输,该方法可以在不可预测的信道条件下保持图像质量。

总之,现有JSCC的算法主要考虑数据压缩和信道状态信噪比的影响,忽略了传输速率的影响。同时,现有算法通常在固定信噪比和单一噪声类型下进行图像传输。针对这些问题,本文提出了一种面向无人机航拍图像传输的信道弹性编解码方法,该方法可以根据信道状态信噪比和传输速率自适应调整语义特征,实现航拍图像的高效传输。本文方法与其他相关工作的贡献对比见表1。

2 信道弹性编解码方法

本节将详细阐述所提的信道弹性编解码框架,其架构如图1所示,包含轻量化特征提取模块、特征增强模块、传输速率自适应模块和相应的解码模块。Zhang等^[35]构建了一种DeepMA网络,编码器可将输入数据编码为相互正交的语义符号向量,解码器可从混合接收信号中检测并恢复目标数据。Jiang^[36]等提出了一种ASE-JSCC框架,可高效提取、筛选并压缩下游任务所需的语义特征。

2.1 总体架构

在这一部分,首先将航拍图像 X 输入轻量化特征提取模块。该模块包含一个卷积层、一个ReLU非线性激活层、两个残差块、两个token生成块、两

表1

本文方法与其他相关工作的贡献对比

方案类型	年度	参考文献	贡献
面向任务通信	2024	[20]	提出一种利用联合信源信道编码和深度神经网络的低时延可配置语音传输系统
	2024	[27]	提出一种在线学习的信源信道联合编码方法,利用深度学习的过拟合特性提升系统泛化能力
	2025	[28]	提出一种基于深度学习的多模态语义通信框架,将多模态模型与大语言模型集成至语义通信系统中
JSCC	2024	[35]	构建一种DeepMA网络,包含编解码器,编码器可将输入数据编码为相互正交的语义符号向量,解码器可从混合接收信号中检测并恢复目标数据
	2025	[36]	提出一种ASE-JSCC框架,包含语义提取模块与关键特征选择模块,可高效提取、筛选并压缩下游任务所需的语义特征
本文方法			设计一种信道增强模块:该模块包含内卷积操作和信噪比调整模块,可以根据信道状态SNR增强关键语义特征 设计一种传输速率自适应模块:该模块包含速率调整块和速率掩码块,可以根据传输速率自适应调整传输特征的数量

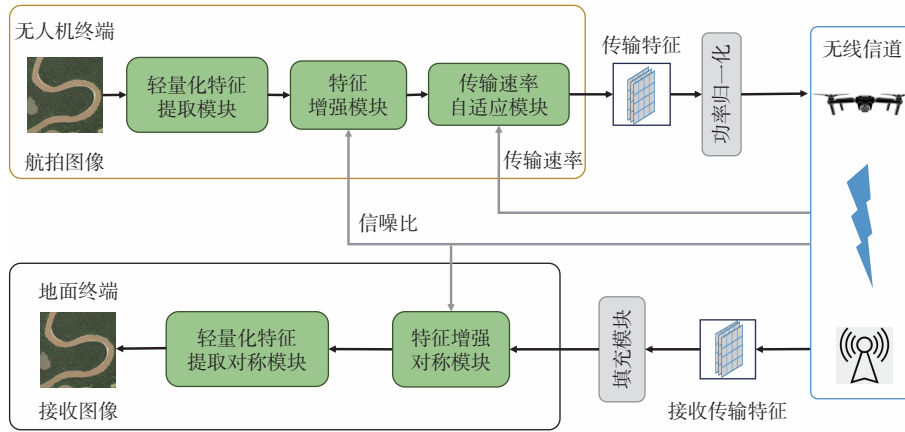


图1 信道弹性编解码架构

个轻量Mamba块和一个自注意力层。其中,残差块用于捕获局部空间细节特征,而轻量Mamba块可以捕获长距离依赖信息。此外,特征增强模块可以利用内卷积操作和SNR调整模块来增强关键语义特征。为了适应传输速率的变化,本文设计了传输速率自适应模块,动态调整传输特征 F_t 的大小。形式上,编码过程可以概括为 $F_t = \text{Encoder}(X; \theta_e)$,其中, θ_e 表示编码模块的所有参数。

在通过无线信道发射 F_t 之前,使用功率归一化操作确保平均发射功率被约束到 P_t ,如式(1)所示。

$$F_t^{\text{norm}} = \sqrt{\frac{P_t}{\mathbb{E}[\|F_t\|_2^2]}} F_t \quad (1)$$

其中, $\mathbb{E}[\cdot]$ 表示对 F_t 所有元素取数学期望, $\|\cdot\|_2$ 是欧几里得范数。

在航拍图像传输过程中,为提高该方法的适应性,引入加性高斯白噪声(AWGN, additive white Gaussian noise)和瑞利衰落干扰来模拟空地信道状态变化。

当在无线信道上直接发送归一化模拟传输特征 F_t^{norm} 时,AWGN被添加到传输特征 F_t^{norm} 中,表示为

$$\hat{F}_t^{\text{norm}} \leftarrow F_t^{\text{norm}} + \text{AWGN}_{\text{noise}} \quad (2)$$

其中, $\text{AWGN}_{\text{noise}} \sim \mathcal{N}(0, P/r)$ 分布, P 和 r 分别表示符号功率和线性标度的SNR, $r = 10^{\frac{\text{SNR}}{10}}$ 。

瑞利衰落模型可以很好地模拟空地信道的时变和空间特性。在该模型中,信道系数 h 由复高斯随机变量组成,其维度与传输特征 F_t^{norm} 相同。随后,将传输特征 F_t^{norm} 乘以信道系数 h ,并叠加高斯白噪声 G ,即可得到接收信号,可表示为

$$\hat{F}_t^{\text{norm}} \leftarrow h \cdot F_t^{\text{norm}} + G \quad (3)$$

其中, h 中各信道系数均为复数值,其实部和虚部分别服从零均值、等方差的独立同分布高斯随机变量。

对应的解码模块在地面端运行,与无人机端同时训练。所提的解码器和编码器具有对称架构,包含填充模块^[32]、特征增强对称模块、轻量化特征提取对称模块。联合信源信道编解码可以提高信道的抗噪声能力和传输速率适应性。解码模块的过程也

可以表示为

$$X' = \text{Decoder}(\hat{F}_t^{\text{norm}}; \theta_d) \quad (4)$$

其中, θ_d 表示解码模块的所有参数。

2.2 轻量化特征提取模块

由于无人机的功耗和计算能力有限, 本文设计了一个轻量化的特征提取模块, 其结构如图2所示, 包含一个卷积层、一个 ReLU 非线性激活层、两个残差块、两个 token 生成块、两个轻量 Mamba 块和一个自注意力层。初始卷积和 ReLU 层用来提取嵌入特征 F^{emb} , 表示为

$$F^{\text{emb}} = \text{ReLU}(\text{Conv}_{\text{ini}}(X)) \quad (5)$$

其中, X 表示输入航拍图像, Conv_{ini} 表示具有 3×3 的初始卷积层。

此外, 残差块可以提取包含纹理和颜色的局部细节特征。残差块由3个卷积层和3个 ReLU 非线性激活层组成, 可以相应表示为

$$\begin{aligned} F^1 &= \text{ReLU}(\text{Conv}_1(X)) \\ F^2 &= \text{ReLU}(\text{Conv}_2(X)) \\ F^3 &= \text{ReLU}(\text{Conv}_3(X)) \end{aligned} \quad (6)$$

其中, Conv_1 、 Conv_2 、 Conv_3 分别表示3个具有 1×1 的卷积层。第1残差块的输出可以表示为

$$F^{\text{res1}} = F^{\text{emb}} + F^3 \quad (7)$$

类似于第1残差块的操作, 可以获得第2残差块输出 F^{res2} 。

残差块主要提取局部细节特征, 难以对全局图像建模。为此, 本文设计了两个 token 生成块、两个轻量 Mamba 块和一个自注意力层, 从 F^{res2} 中提取长距离依赖特征。轻量 Mamba 块包含两个归一化层、一个一维卷积层、一个非线性激活层和一个 S6 模型^[10]。

本文仅采用水平正向与水平逆向生成两个定向 token 序列, 水平正向和水平逆向 token 序列可以分

别表示为 T_{hf} 和 T_{hb} 。在轻量 Mamba 中, 以 T_{hf} 为例, 首先通过归一化层、一维卷积层和激活函数获得中间特征 t_{hf} , 可以表示为

$$t_{\text{hf}} = \text{SiLU}(\text{Conv}(\text{Norm}(T_{\text{hf}}))) \quad (8)$$

其中, $\text{SiLU}(\cdot)$ 表示 Sigmoid 线性单元激活函数。

此外, 可以通过线性层、S6 模型和归一化层获得 Mamba 特征, 这个过程可以表示为

$$F_{\text{hf}} = \text{Norm}(S6(\text{Linear}(t_{\text{hf}}))) \quad (9)$$

其中, $S6(\cdot)$ 表示选择性状态空间。

为了进一步增强捕捉全局上下文和远程空间依赖性的能力, 本文在两个轻量 Mamba 块之间采用了自注意力机制, 可表示为

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

其中, Q 、 K 、 V 分别表示查询、键、值矩阵, 由输入 token 特征 F_{hf} 与可学习的映射矩阵相乘得到, d_k 表示特征维度。

类似于第1个 token 生成和轻量 Mamba 块的操作, 可以获得第2个轻量 Mamba 块输出 $F \in \mathbf{R}^{C \times H \times W}$, 其中, C 表示通道数, H 表示特征高度, W 表示特征宽度。

2.3 特征增强模块

特征增强模块 (FEM, feature enhancement module) 可根据信道 SNR, 自适应增强特征的重要性。特征增强模块结构如图3所示, 由两个内卷积层、两个 SNR 调整块、一个线性层和一个 Softmax 层组成。

首先, 引入内卷积运算。传统的卷积运算沿着单个图像滑动窗口, 该图像在空间域上共享卷积核权重, 这限制了传统卷积对随空间变化的噪声的适

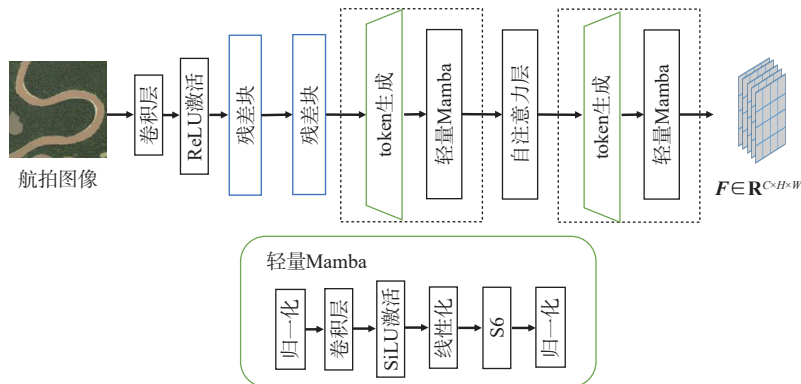


图2 轻量化特征提取模块结构

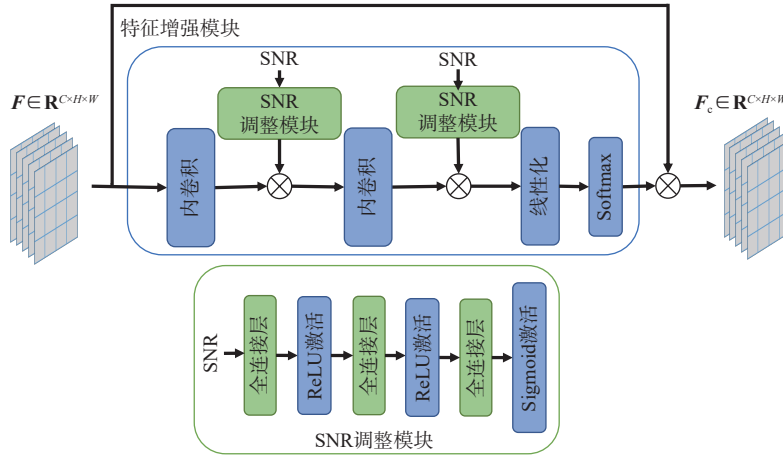


图3 特征增强模块结构

应能力。而内卷积操作沿着信道滑动窗口，解决了传统卷积在处理变化的信道噪声时的局限性。具体地，内卷积核 $H_{i,j} \in \mathbf{R}^{K \times K \times 1}$ 直接由输入特征 $X_{i,j}$ 上的核生成函数 α 产生，可以表示为

$$H_{i,j} = \alpha(X_{i,j}) = W_1 \sigma(W_0 X_{i,j}) \quad (11)$$

其中， $K \times K$ 表示内卷积核的空间大小， $X_{i,j}$ 表示位于相应坐标 (i,j) 处的特征。 σ 表示归一化和 ReLU 非线性激活函数，可以增强模型的表示能力。 $W_1 \in \mathbf{R}^{(K \times K \times 1) \times C}$ 和 $W_0 \in \mathbf{R}^{C \times C}$ 表示两个线性变换。内卷积核 $H_{i,j}$ 具有与 $X_{i,j}$ 相同的空间大小。从理论上讲，内卷积层允许进行特定空间的重加权，意味着模型可以自适应地为不同的空间位置分配不同的权重，这可以在更宽的空间范围内总结上下文信息。

SNR 调整模块由全连接层和激活层构成，该模块有效地增强了对信道状态 SNR 的感知能力，在无需高计算成本的情况下，确保了在动态信道条件下的稳定性能，具体如式(12)所示。

$$\begin{aligned} s^1 &= \text{ReLU}(W_s^1 \cdot \text{SNR} + b_s^1), \\ s^2 &= \text{ReLU}(W_s^2 \cdot s^1 + b_s^2), \\ s^3 &= \text{ReLU}(W_s^3 \cdot s^2 + b_s^3), \\ s &= \text{Sigmoid}(W_s^4 \cdot s^3 + b_s^4) \end{aligned} \quad (12)$$

其中， $\text{ReLU}(\cdot)$ 和 $\text{Sigmoid}(\cdot)$ 分别表示线性激活函数和非线性激活函数， W_s^i 和 b_s^i ， $i \in \{1, 2, 3, 4\}$ ，分别表示全连接层的权重和偏差， s 表示 SNR 调整模块的输出。

第1个 SNR 调整模块的输出 C 维向量与由第1层内卷积生成的特征通道相乘，之后第2个模块进行类似操作，可表示为

$$\begin{aligned} f_{c1} &= s \otimes \text{Involution}^1(F), \\ f_{c2} &= s \otimes \text{Involution}^2(f_{c1}) \end{aligned} \quad (13)$$

其中， $\otimes$ 表示逐元素乘法， $\text{Involution}(\cdot)$ 表示内卷积， f_{c1} 和 f_{c2} 分别表示第1次和第2次逐元素乘法的输出。

此外，线性层和 Softmax 层用于信道加权参数 W_c ，如式(14)所示。

$$W_c = \text{Softmax}(\text{Linear}(f_{c2})) \quad (14)$$

因此，可以将通道增强特征的输出表示为

$$F_c = W_c \otimes f_{c2} \quad (15)$$

2.4 传输速率自适应模块

在基于无人机的空对地航拍图像传输过程中，信道状态信噪比和传输速率都是时变的。为了更好地实时适应传输速率的变化，本文提出了一种有效的传输速率自适应模块（TRAM, transmission rate adaptive module）。该模块采用了一种可学习的掩码策略，可以基于传输速率 R 重新调整由 FEM 生成的增强特征，从而自动调节模型对传输速率的自适应能力，确保模型可以在任意传输速率下灵活运行。本文方法可以在无人机空对地通信系统中实现高效可靠的航拍图像传输。

传输速率自适应模块结构如图4所示，TRAM 包含3个速率调整块、4个具有 1×1 核的卷积层和1个速率掩码块。速率调整块由3个全连接层组成，可以将信道带宽比（CBR, channel bandwidth ratio）转换为 C 维向量 r 。这个过程可以表示为

$$\begin{aligned} r^1 &= \text{ReLU}(W_r^1 \cdot \text{CBR} + b_r^1), \\ r^2 &= \text{ReLU}(W_r^2 \cdot r^1 + b_r^2), \\ r^3 &= \text{Sigmoid}(W_r^3 \cdot r^2 + b_r^3) \end{aligned} \quad (16)$$

其中， W_r^i 和 b_r^i ， $i \in \{1, 2, 3\}$ ，分别表示速率调整块

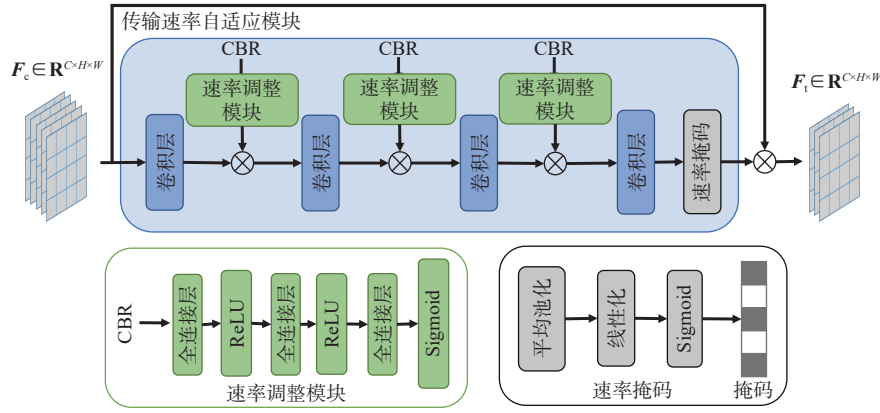


图4 传输速率自适应模块结构

的全连接层的权重和偏差。

模块通过3个速率调整块和3个卷积层，可以获得中间特征 f_t 。此外，速率掩码被设计成生成掩码向量 m ，表示为

$$\begin{aligned} a &= \text{AveragePooling}(f_t), \\ m &= \text{Sigmoid}(\text{Linear}(a)) \end{aligned} \quad (17)$$

因此，传输速率自适应模块的输出可以被表示为

$$F_t = m \otimes F_c \quad (18)$$

其中， F_t 表示进入信道前的传输特征。

2.5 损失函数

本文所提方法以端到端的方式进行训练，双重损失函数由重建损失和感知损失组成，平衡重建时的保真度和潜在空间正则化。

1) 重建损失

为了实现高保真图像重建，本文采用重建图像 $\hat{x}$ 和原始图像 x 之间的均方误差作为损失

$$L_{\text{rec}} = \|X' - X\|_2^2 \quad (19)$$

该损失函数惩罚逐像素的差异，确保解码图像与原始图像尽可能相似。

2) 感知损失

为了保证解码图像和原始图像在感知层面上尽可能相似，本文采用如下感知损失

$$L_{\text{per}} = \frac{1}{M \times H \times W} \|\hat{F}_t^{\text{norm}} - F_t^{\text{norm}}\|_2^2 \quad (20)$$

总体训练目标是两个组成部分的加权组合

$$L = L_{\text{rec}} + \beta \cdot L_{\text{per}} \quad (21)$$

其中， $\beta (\beta > 0)$ 是控制感知损失比例的超参数。

3 实验测试

本节通过对比实验，与现有方法进行性能比

较，如 JPEG+LDPC、Deep JSCC^[3]、VAE^[37]、JSCCformer-f^[38]，以验证本文所提方法在AID上的性能。

AID包括30个类别，总共10 000张图像。该数据集包含许多场景类别，如机场、桥梁等。每个类别有220~420张大小为600 × 600像素的图像。

此外，在 JPEG+LDPC、Deep JSCC、VAE 和本文所提方法之间进行了广泛的对比实验。

JPEG+LDPC：该方法应用 JPEG^[39]编解码器进行压缩，结合 LDPC^[40]进行信道编码。

DeepJSCC：该方法通过两个联合训练的卷积神经网络对编码器和解码器进行参数优化。

VAE：该方法采用变分自动编码器模型。

JSCCformer-f：该方法将信道反馈引入信源信道联合编解码过程中，协同优化多个模块间的传输过程。

在本文中，为评估接收图像的质量，采用了3项评估指标：峰值信噪比（PSNR, peak signal-to-noise ratio）、多尺度结构相似性指数度量（MS-SSIM, multi-scale structural similarity index measure）^[41]以及基于学习的感知图像块相似性（LPIPS, learned perceptual image patch similarity）^[42]。其中，PSNR和MS-SSIM的数值越高，表示图像重建质量越优，LPIPS的数值越低，表示感知失真越小，方法的性能越好。

PSNR是图像质量评估中被广泛采用的像素级度量指标，定义为

$$\text{PSNR}(I_g, I_r) = 10 \lg \left(\frac{P_{\text{max}}}{\text{MSE}(I_g, I_r)} \right) \quad (22)$$

其中， I_g 、 I_r 分别表示地面真实图像与接收图像， P_{max}

为最大像素值。 $MSE(\cdot)$ 表示均方误差, 计算式为

$$MSE(I_g, I_r) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (I_g^{ij} - I_r^{ij})^2 \quad (23)$$

其中, N 表示图像单边像素数, I_g^{ij} 和 I_r^{ij} 分别表示地面真实图像和接收图像在像素位置 (i, j) 处的像素值。

SSIM是衡量两幅图像在亮度、对比度和结构上相似度的指标。SSIM评估指标更贴近人类视觉系统的感知, 能够准确地反映图像质量。SSIM定义为

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (24)$$

其中, μ_x 和 μ_y 分别是图像 x 和 y 在局部窗口内的平均值, 表示图像的亮度水平; σ_x^2 和 σ_y^2 分别是图像 x 和 y 在局部窗口内的方差, 表示图像的对比度; σ_{xy} 是图像 x 和 y 在局部窗口内的协方差, 表示图像的结构相似性。

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2 \quad (25)$$

其中, 通常 $K_1 = 0.01$, $K_2 = 0.03$, L 是像素值的动态范围。本文采用的MS-SSIM根据不同尺度下SSIM的结果, 对图像质量进行评估。MS-SSIM能够同时考虑图像的全局和局部结构信息, 提高评估的准确性和鲁棒性。

此外, 为更准确地衡量图像间的语义相似性, 本文引入LPIPS指标。LPIPS基于深度神经网络提取多层级语义特征, 并将图像划分为多个局部块。随后, 计算地面真实图像与接收图像对应特征块之间的欧几里得距离, 以量化语义相似度。最后, 对各块的相似性得分进行加权平均, 得到整体LPIPS分数, 其表达式为

$$LPIPS(I_g, I_r) = \sum_l \frac{1}{H_l W_l} \sum_{i,j} \|w_l \odot (F_g^l - F_r^l)\|_2^2 \quad (26)$$

其中, H_l 、 W_l 是第 l 个图像块特征的高度和宽度; $\odot$ 表示元素乘积; w_l 为不同图像块的权重; F_g^l 、 F_r^l 分别是地面真实图像和接收图像的第 l 级特征图。

3.1 实验设置

本文将训练集与测试集按8:2的比例划分, 并对模型进行3次独立训练和评估, 对比各项性能

指标的平均值。在每个训练轮次, 在训练集上使用Adam随机优化算法更新模型参数, 批量大小为64, 学习率为 10^{-4} 。

3.2 实验结果

本节在AID上对比评估了不同方法的训练与测试结果。在训练中引入了不同信噪比AWGN和瑞利衰落信道干扰, 模拟空地传输中的复杂通信场景。

不同方法在AID上的性能比较见表2, 在压缩比为6、SNR为1的AID上, 本文方法在分类精度方面优于JPEG、Deep JSCC、VAE及JSCCformer-f方法。值得注意的是, 在PSNR指标上, 本文方法较传统JPEG+LDPC方法在AWGN信道下提升6.05 dB, 在瑞利衰落信道下提升5.43 dB, 表明所提方法具有更强的特征提取能力。同时, 相较于当前的JSCCformer-f方法, 本文方法在PSNR上分别提升0.71 dB和0.93 dB, 进一步验证了其在特征增强与传输速率自适应能力方面的优越性。此外, 本文方法在MS-SSIM和LPIPS两项人类视觉相似性度量中均取得更优表现。以上结果表明, 在AID上本文方法不仅在像素级图像重建质量上表现良好, 而且在图像结构相似性方面具有显著优势。

3.3 消融实验

为验证本文所提方法中特征增强模块和传输速率自适应模块的有效性, 本文在AID上进行消融实验, 不同方法在AID上的性能比较见表3。实验条件统一设定为CBR=0.166 7, SNR=1 dB。为系统地评估各模块, 设计了3组对照方法: “本文方法缺少FEM” “本文方法缺少TRAM” 和 “同时缺少FEM和TRAM”。

由表3可知, FEM和TRAM在航拍图像传输过程中均发挥重要作用。值得注意的是, FEM对于增强关键语义特征非常重要, 当从本文方法移除FEM模块时, 传输性能显著下降。

3.4 可视化结果

为了进一步验证本文所提方法的有效性, 本文

表2 不同方法在AID上的性能比较

方法	AWGN信道			瑞利衰落信道		
	PSNR/dB	MS-SSIM	LPIPS	PSNR/dB	MS-SSIM	LPIPS
JPEG+LDPC	24.90	0.830 5	0.546 5	24.45	0.809 5	0.567 7
Deep JSCC	22.94	0.809 3	0.572 5	19.51	0.544 3	0.796 3
VAE	23.07	0.825 9	0.585 5	23.07	0.825 9	0.585 5
JSCCformer-f	30.24	0.938 0	0.192 0	28.95	0.915 0	0.225 0
本文	30.95	0.941 8	0.186 7	29.88	0.926 8	0.204 5

表 3

不同方法在 AID 上的性能比较(消融实验)

方法	AWGN 信道			瑞利衰落信道		
	PSNR/dB	MS-SSIM	LPIPS	PSNR/dB	MS-SSIM	LPIPS
本文方法缺少 FEM	29.600	0.911 4	0.179 8	28.983	0.910 3	0.239 1
本文方法缺少 TRAM	29.619	0.915 8	0.197 4	29.651	0.915 8	0.185 7
同时缺少 FEM 和 TRAM	28.780	0.902 4	0.200 2	28.553	0.895 2	0.210 2
本文方法	30.950	0.941 8	0.186 7	29.880	0.926 8	0.204 5

选取包含山地、桥梁、池塘、工业区、学校操场等典型场景的航拍图像作为输入测试样本，所有原始图像均源自 AID。实验条件设为 CBR=0.167 和 SNR=1 dB，不同方法在 AID 上的可视化结果如图 5 所示。图 5 中前 3 行是 AWGN 信道下不同方法在图像重建质量上的对比，最后两行是瑞利衰落信道下的对比。前两列显示输入原始图像和原始图像块，后 5 列显示在不同方法下图像块重建后的结果。从图 5 可知，本文方法在恢复图像细节与结构完整性方面显著优于对比方法；相较之下，JSCCformer-f 方法在重建图像时存在模糊和失真现象；本文方法能有效地保留纹理边缘与空间结构，重建图像更真实。

3.5 信道状态 SNR 的影响

本节主要分析信道状态参数 SNR 对模型性能的影响。SNR 直接影响语义特征的增强表示能力，进而影响图像在传输过程中的重建精度。为此，本文在 SNR 取值为[-1, 1, 3, 5, 7, 9] dB 的多个信道条件下，系统地比较了本文方法与现有方法的传输性能。

信道状态 SNR 对 PSNR 的影响如图 6 所示，在 AID 上分别展示了 AWGN 信道和瑞利衰落信道下各方法的 PSNR 曲线。实验结果表明，本文方法在

所有 SNR 水平下均取得最优结果，具有更强的鲁棒性。当 SNR 较低 (SNR=1 dB) 时，本文方法仍显著优于其他方法，实现了更高的传输精度。

3.6 传输速率的影响

本节重点分析了传输速率对模型性能的影响。在 SNR=7 dB 的条件下，系统地比较了不同方法在传输速率 CBR=[0.010 4, 0.020 8, 0.041 7, 0.083 3, 0.166 7]时的图像传输性能。

传输速率对 PSNR 的影响如图 7 所示，在 AID 上，本文所提方法在 AWGN 信道和瑞利衰落信道下均展现出最优的 PSNR 表现。实验结果表明，无论在何种信道环境中，本文方法均体现出优异的传输速率自适应能力。

3.7 模型复杂度分析

本节重点分析了本文方法与现有方法在模型参数量、推理时延和计算量上的差别。所有测量都在单台 NVIDIA RTX 3090 GPU 上进行，并将 CBR 设置为 0.041 7，保证相同的参数设置，不同方法在 AID 上的复杂度比较见表 4。

与 JSCCformer-f 方法相比，本文所提方法在轻量化部署方面表现出显著优势。本文方法将计算

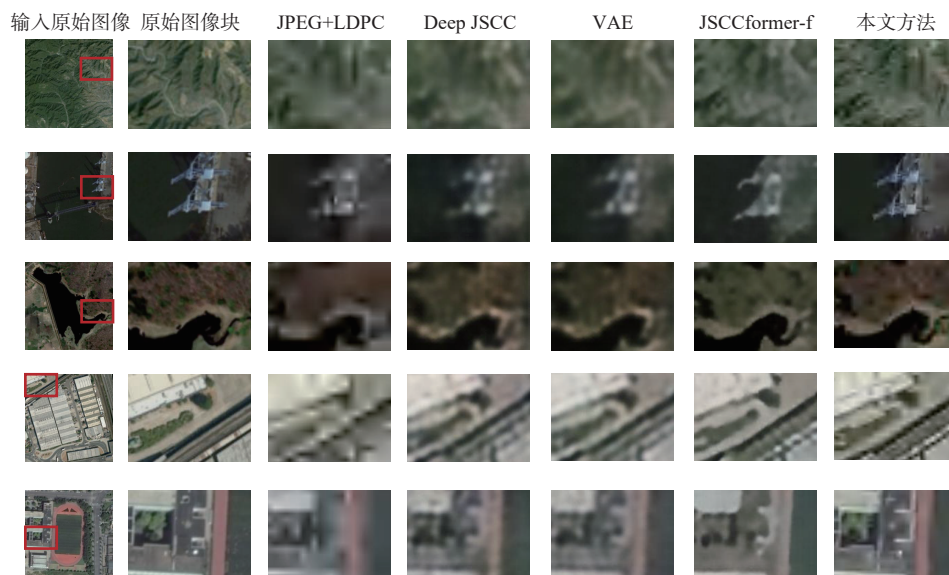


图 5 不同方法在 AID 上的可视化结果

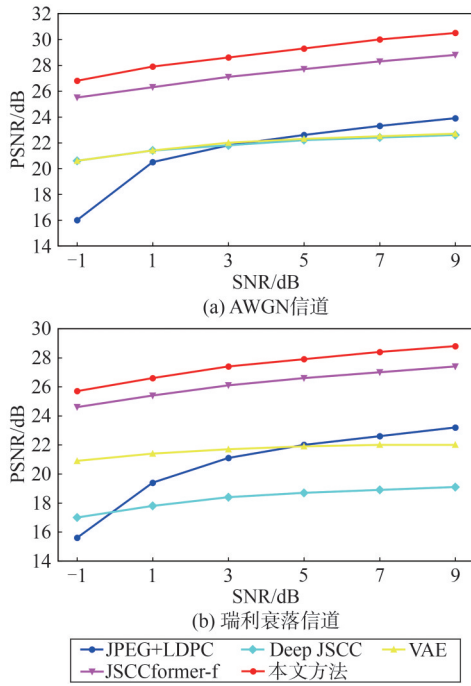


图6 信道状态 SNR 对 PSNR 的影响

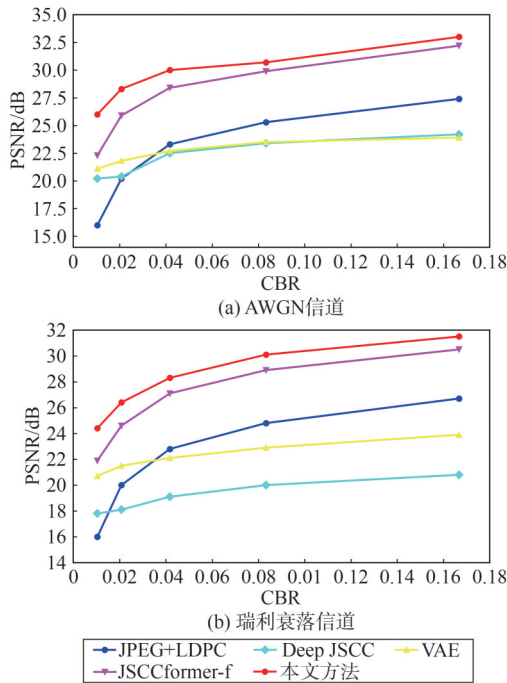


图7 传输速率对 PSNR 的影响

量降低至 3.05 G FLOPs, 约为 JSCCformer-f 方法的 30%, 并实现了 11.35 ms 的推理时延, 显著低于 JSCCformer-f 方法的 123.48 ms 高推理时延。尽管与 Deep JSCC 和 VAE 相比, 本文方法在参数量和计算量上有适度增加, 但由此带来的性能提升显著。如表 2 所示, 在 AWGN 信道下的 AID 上, 本文方法实现了 30.95 dB 的 PSNR, 比 Deep JSCC (22.94 dB)

表4 不同方法在 AID 上的复杂度比较

方法	模型参数量	FLOPs	推理时延/ms
JPEG+LDPC	—	—	54.40 ms
Deep JSCC	0.16×10^6	0.51×10^9	0.68 ms
VAE	0.20×10^6	0.53×10^9	1.13 ms
JSCCformer-f	13.07×10^6	10.12×10^9	123.48 ms
本文方法缺少 FEM	1.25×10^6	2.62×10^9	8.92 ms
本文方法缺少 TRAM	1.43×10^6	288.00×10^9	10.13 ms
同时缺少 FEM 和 TRAM	1.57×10^6	2.76×10^9	10.97 ms
本文方法	1.75×10^6	3.05×10^9	11.35 ms

和 VAE (23.07 dB) 高出约 8 dB。

4 结束语

本文提出了一种面向无人机航拍图像传输的信道弹性编解码方法, 该方法包含轻量化特征提取模块、特征增强模块、传输速率自适应模块和相应的解码模块。轻量化特征提取模块可以捕获图像局部细节特征和全局上下文信息, 提取丰富的语义特征信息。特征增强模块可以根据信道状态 SNR 增强关键语义特征。传输速率自适应模块可以根据传输速率自适应地调整传输特征的数量。在 AID 上的实验结果表明, 本文所提方法在 PSNR、MS-SSIM 与 LPIPS 这 3 项关键指标上均优于现有方法, 表现出更强的语义保真能力与信道自适应性。由于本文方法对 SNR 和 CBR 的状态输入存在敏感性, 若 SNR 或 CBR 的反馈时延较大, 性能可能下降。未来, 将进一步研究如何提升在信道状态信息不完美、反馈时延等场景中方法的鲁棒性, 以及如何将方法扩展到视频传输等通信场景中, 提高模型的适用性。

参考文献:

- [1] Kokalj-Filipović S, Soljanin E. Suppressing the cliff effect in video reproduction quality[J]. Bell Labs Technical Journal, 2012, 16(4): 171-185.
- [2] Xiao Y, Sun Z J, Shi G M, et al. Imitation learning-based implicit semantic-aware communication networks: multi-layer representation and collaborative reasoning[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(3): 639-658.
- [3] Bourtsoulatz E, Burth Kurka D, Gündüz D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 567-579.
- [4] Fu Y Z, Cheng W C, Zhang W. Content-aware semantic communication for goal-oriented wireless communications[C]//Proceedings of the IEEE INFOCOM 2023-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2023: 1-6.

- [5] Dai J C, Wang S X, Yang K, et al. Toward adaptive semantic communications: efficient data transmission via online learned nonlinear transform source-channel coding[J]. *IEEE Journal on Selected Areas in Communications*, 2023, 41(8): 2609-2627.
- [6] Kurka D B, Gündüz D. DeepJSCC-f: deep joint source-channel coding of images with feedback[J]. *IEEE Journal on Selected Areas in Information Theory*, 2020, 1(1): 178-193.
- [7] Xu J L, Tung T Y, Ai B, et al. Deep joint source-channel coding for semantic communications[J]. *IEEE Communications Magazine*, 2023, 61(11): 42-48.
- [8] Yang M Y, Kim H S. Deep joint source-channel coding for wireless image transmission with adaptive rate control[C]//*Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Piscataway: IEEE Press, 2022: 5193-5197.
- [9] Yang K, Wang S X, Dai J C, et al. SwinJSCC: taming swin transformer for deep joint source-channel coding[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2025, 11(1): 90-104.
- [10] Liu Y, Tian Y, Zhao Y, et al. Vmamba: visual state space model[J]. *Advances in Neural Information Processing Systems*, 2024, 37: 103031-103063.
- [11] Xia G S, Hu J W, Hu F, et al. AID: a benchmark data set for performance evaluation of aerial scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3965-3981.
- [12] Shannon C E. A mathematical theory of communication[J]. *Bell System Technical Journal*, 1948, 27(3): 379-423.
- [13] Weng Z Z, Qin Z J. Semantic communication systems for speech transmission[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(8): 2434-2444.
- [14] Xie H Q, Qin Z J, Li G Y, et al. Deep learning enabled semantic communication systems[J]. *IEEE Transactions on Signal Processing*, 2021, 69: 2663-2675.
- [15] Xie H Q, Qin Z J. A lite distributed semantic communication system for Internet of things[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(1): 142-153.
- [16] Xie H Q, Qin Z J, Li G Y. Task-oriented multi-user semantic communications for VQA[J]. *IEEE Wireless Communications Letters*, 2022, 11(3): 553-557.
- [17] Jiang P W, Wen C K, Jin S, et al. Deep source-channel coding for sentence semantic transmission with HARQ[J]. *IEEE Transactions on Communications*, 2022, 70(8): 5225-5240.
- [18] Kutay E, Yener A. Semantic text compression for classification[C]//*Proceedings of the 2023 IEEE International Conference on Communications Workshops (ICC Workshops)*. Piscataway: IEEE Press, 2023: 1368-1373.
- [19] Zhou Q Y, Li R P, Zhao Z F, et al. Semantic communication with adaptive universal transformer[J]. *IEEE Wireless Communications Letters*, 2022, 11(3): 453-457.
- [20] Bokaei M, Jensen J, Doclo S, et al. Channel-configurable deep wireless speech transmission[C]//*Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC)*. Piscataway: IEEE Press, 2024: 1-6.
- [21] Jankowski M, Gündüz D, Mikołajczyk K. Wireless image retrieval at the edge[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(1): 89-100.
- [22] Hu Q Y, Zhang G Y, Qin Z J, et al. Robust semantic communications with masked VQ-VAE enabled codebook[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(12): 8707-8722.
- [23] Xing H L, Zhang H L, Wang X H, et al. A multi-user deep semantic communication system based on federated learning with dynamic model aggregation[C]//*Proceedings of the 2023 IEEE International Conference on Communications Workshops (ICC Workshops)*. Piscataway: IEEE Press, 2023: 1612-1616.
- [24] Luo L, He Z Y, Wu J J, et al. Adaptive deep joint source-channel coding for one-to-many wireless image transmission[J]. *IEEE Transactions on Broadcasting*, 2025, 71(3): 914-929.
- [25] Li H, Zhang G, Zhou K, et al. Coarse-to-fine: a dual-phase channel-adaptive method for wireless image transmission[J]. *IEEE Transactions on Wireless Communications*, 2026, 25: 2607-2623.
- [26] Yan Y, Li L, Zhang X, et al. SAFE: semantic adaptive feature extraction with rate control for 6G wireless communications[C]//*Proceedings of the 2024 IEEE Globecom Workshops (GC Wkshps)*. Piscataway: IEEE Press, 2024: 1-6.
- [27] Zhang G Y, Hu Q Y, Qin Z J, et al. A unified multi-task semantic communication system for multimodal data[J]. *IEEE Transactions on Communications*, 2024, 72(7): 4101-4116.
- [28] Jiang F B, Dong L, Peng Y B, et al. Large AI model empowered multimodal semantic communications[J]. *IEEE Communications Magazine*, 2025, 63(1): 76-82.
- [29] Dai J C, Zhang P, Niu K, et al. Communication beyond transmitting bits: semantics-guided source and channel coding[J]. *IEEE Wireless Communications*, 2023, 30(4): 170-177.
- [30] Xu J L, Ai B, Chen W, et al. Wireless image transmission using deep source channel coding with attention modules[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(4): 2315-2328.
- [31] Kurka D B, Gündüz D. Bandwidth-agile image transmission with deep joint source-channel coding[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(12): 8081-8095.
- [32] Zhang W Y, Zhang H J, 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.
- [33] 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.
- [34] Tung T Y, Kurka D B, Jankowski M, et al. DeepJSCC-Q: constellation constrained deep joint source-channel coding[J]. *IEEE Journal on Selected Areas in Information Theory*, 2022, 3(4): 720-731.
- [35] Zhang W Y, Bai K Y, Zeadally S, et al. DeepMA: end-to-end deep

multiple access for wireless image transmission in semantic communication[J]. IEEE Transactions on Cognitive Communications and Networking, 2024, 10(2): 387-402.

- [36] Jiang Y, Xie K, Ouyang Y D, et al. High quality compression and transmission of remote sensing images based on semantic communication[J]. IEEE Transactions on Sustainable Computing, 2025, 10(5): 843-857.
- [37] Estiri A H, Sabramooz M R, Banaei A, et al. A variational auto-encoder approach for image transmission in noisy channel[C]//Proceedings of the 2020 10th International Symposium on Telecommunications (IST). Piscataway: IEEE Press, 2020: 227-233.
- [38] Wu H T, Shao Y L, Ozfatura E, et al. Transformer-aided wireless image transmission with channel feedback[J]. IEEE Transactions on Wireless Communications, 2024, 23(9): 11904-11919.
- [39] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview[J]. IEEE Transactions on Consumer Electronics, 2000, 46(4): 1103-1127.
- [40] Richardson T, Kudekar S. Design of low-density parity check codes for 5G new radio[J]. IEEE Communications Magazine, 2018, 56(3): 28-34.
- [41] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. Piscataway: IEEE Press, 2003: 1398-1402.
- [42] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 586-595.

[作者简介]



张中强(1993-), 男, 鹏城实验室助理研究员, 主要研究方向为语义通信、卫星通信、水声通信。



邱奇(2003-), 男, 哈尔滨工业大学(深圳)信息学部博士生, 主要研究方向为语义通信。



苗雅杰(2003-), 男, 哈尔滨工业大学(深圳)信息学部博士生, 主要研究方向为语义通信。



黎斌(1986-), 男, 鹏城实验室工程师, 主要研究方向为5G通信技术、无人机平台通信系统及其工程应用。



王野(1983-), 男, 鹏城实验室副研究员, 主要研究方向为卫星通信和移动互联网。



石光明(1965-), 男, 鹏城实验室副主任, 主要研究方向为人工智能、语义通信和人机交互。