

基于深度灵活编码策略与性能预测的无线语义图像传输系统

杨舒涵^{1,2,3}, 申滨^{1,3}, 黄晓舸^{1,3}

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 运城学院数学与信息技术学院, 山西 运城 044000;
3. 移动通信技术重庆市重点实验室, 重庆 400065)

摘要: 为了提升多样化的图像内容在复杂多变的无线信道条件下的自适应性和编码灵活性, 面向语义图像传输 (SIT) 系统提出一种深度灵活的信源信道联合编码 (DF-JSCC) 方案以及传输性能预测 (PreNet) 模型。具体而言, 在 DF-JSCC 中融入多尺度特征融合模块 (MS-FFM), 提高模型对细节特征的捕获能力。进一步引入特征增强模块 (FEM), 通过自适应地调整特征图的通道权重, 增强对重要特征的表达能力。此外, PreNet 通过在细节丰富的特征图中进行特征分布建模, 并利用多维特征融合和学习, 实现了对重构图像峰值信噪比 (PSNR) 的高质量预测。实验结果表明, 与现有最先进的 DeepJSCC 方案相比, DF-JSCC 在多样化通信场景中实现了至少 1.03 dB 的 PSNR 性能改进, 同时在相同的传输条件下, 能够节约 50% 以上的带宽资源消耗。PreNet 通过进一步降低 20% 的均方误差 (MSE), 增强了复杂环境中 SIT 的预测质量和鲁棒性。

关键词: 编码灵活性; 语义图像传输; 信源信道联合编码; 性能预测

中图分类号: TN911; TN919

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025094

Wireless semantic image transmission system based on deep flexible coding strategy and performance prediction

YANG Shuhan^{1,2,3}, SHEN Bin^{1,3}, HUANG Xiaoge^{1,3}

1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. School of Mathematics and Information Technology, Yuncheng University, Yuncheng 044000, China
3. Chongqing Key Laboratory of Mobile Communications Technology, Chongqing 400065, China

Abstract: To improve adaptivity and coding flexibility for diverse image content under complex and dynamic wireless channel conditions, a deeply flexible joint source-channel coding (DF-JSCC) scheme and a transmission performance prediction network (PreNet) for semantic image transmission (SIT) were proposed. Specifically, a multi-scale feature fusion module (MS-FFM) was incorporated into DF-JSCC to enhance detailed feature extraction. A feature enhancement module (FEM) was also designed to adaptively adjust channel weights in the feature maps to reinforce critical information. Additionally, PreNet was introduced to accurately predict the peak signal-to-noise ratio (PSNR) performance of reconstructed images by modeling feature distribution in detail-rich feature maps and utilizing multidimensional feature fusion and learning. Experimental results demonstrate that, compared with the existing state-of-the-art DeepJSCC method, DF-JSCC achieves at least 1.03 dB improvement in PSNR across various communication scenarios, while reducing bandwidth consumption by over 50% under equivalent transmission conditions. The PreNet effectively improves the prediction quality and robustness of the SIT in complex environments by reducing the mean squared error (MSE) by 20%.

Keywords: coding flexibility, semantic image transmission, joint source-channel coding, performance prediction

收稿日期: 2025-02-10; 修回日期: 2025-05-14

通信作者: 申滨, shenbin@cqupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62371082); 重庆市自然科学基金资助项目 (No.CSTB2023NSCQ-MSX0726)

Foundation Items: The National Natural Science Foundation of China (No.62371082), The Natural Science Foundation of Chongqing (No.CSTB2023NSCQ-MSX0726)

0 引言

传统通信核心关注通信符号在物理层面和技术层面传输的准确性和可靠性,对传输信息的语义和语用层面则关注较少^[1]。6G通信将助力“万物智联”时代的到来,带来大量的新兴智能应用,通信需求从语法层面转向更加智能的语义和语用层面^[1-2]。语义通信以“达意”为目标,能够有效提取语义信息并实现可靠压缩,大大降低数据冗余,提高通信节点协作效能^[1,3]。

作为智能应用产生的重要数据模态之一,图像传输在6G通信中扮演着至关重要的角色^[4-5]。语义通信为图像传输带来了全新的思路和方法,通过引入内容感知的压缩算法、容错与恢复策略以及上下文感知等技术,来进一步提升图像传输的智能化水平和应用价值^[6-8]。

1) 语义图像传输中的预处理编码

在无线通信系统中,传统图像传输中的信源与信道独立编码方式可能导致额外的带宽需求和系统复杂性的增加,且难以实现最优性能,如采用高效图像压缩格式(BPG)的信源编码以及低密度奇偶校验码(LDPC)的信道编码^[9]。此外,传统无线图像传输中普遍存在“悬崖效应”,即在数字通信中,当信噪比(SNR, signal-to-noise ratio)下降到某个临界点时,图像的传输质量呈断崖式下降。

为应对上述问题,部分研究者逐步探索了语义图像传输(SIT, semantic image transmission)技术,通过引入图像语义信息,有效提升了传输效率与质量。Agustsson等^[10]首次基于生成对抗网络(GAN, generative adversarial network)提出了全分辨率图像压缩系统,通过对编解码器的联合训练,在极低比特率下获取了优越的视觉质量和存储效率。Huang等^[11]提出了一种基于GAN的图像语义编码框架,采用由粗到精的编码模型和感知优化技术,在极低比特率下实现了高质量的图像重建和语义信息的有效保留。此外,Dong等^[12]基于人工智能(AI, artificial intelligence)提出层级语义通信系统,通过使用语义切片模型,从语义角度优化了系统性能。

在图像语义的数字化鲁棒传输中,Hu等^[13-14]研究了SIT系统对抗语义噪声的鲁棒性。该系统采用向量量化的变分自编码器实现语义特征提取,并通过引入掩码策略降低噪声对系统的影响。Fu等^[15]基于深度学习(DL, deep learning)与向量量

化提出了SIT系统,通过引入多尺度语义嵌入空间实现特征量化和语义共享。

总体来看,上述SIT研究方法普遍通过将图像预处理为特征向量来降低数据维度和系统复杂度,从而提升传输效率以及捕获图像的高级语义特征,有助于在传输过程中保留重要的语义信息。然而,以上研究方法存在一些不足之处。一方面,特征的向量表示虽能有效降低数据维度,但也可能导致部分细节特征的丢失,这种损失可能会严重影响图像的重构质量,尤其是在对细节要求较高的应用场景中,如医学影像或遥感图像处理等。另一方面,这些预处理方式通常依赖于预训练的深度学习模型,而这些模型的局限性和偏差可能限制了SIT技术在实际应用中的效果^[16]。例如,预训练模型的泛化能力可能不足以处理特定领域的图像数据,或者在面对未训练的信道条件时,系统的性能可能会显著下降。此外,模型的计算复杂性和对硬件资源的需求也可能成为实际应用中的瓶颈,特别是在资源受限的环境中。因此,为了推动SIT技术的进一步发展,亟须探索更加有效的特征提取与编码方法,以在降低数据维度的同时保留更多的图像细节。此外,开发更加鲁棒且泛化能力更强的模型,有望进一步提升SIT技术的实用性和可靠性。

2) 语义图像传输中的信源信道联合编码

利用机器学习模型直接提取、编码和传输图像特征,能够弥补向量化编码等图像预处理方法的一些缺陷。这类方法能够自适应地选择最相关的特征,并有效处理多样化的图像内容,从而为SIT提供更大的灵活性与更优的传输性能^[17]。

Bourtsoulatze等^[18]首次提出了基于DL的信源信道联合编码(DeepJSCC, DL-based joint source-channel coding)技术用于高分辨率无线图像传输。该技术直接将图像像素值映射到复值信道输入符号,通过对编解码器的联合训练,有效缓解了数字通信中的“悬崖效应”。Kurka等^[19]在JSCC架构的基础上引入了信道输出反馈机制,提出了DeepJSCC-f(feedback-based DeepJSCC)模型。该方案通过在多个资源块上分阶段传输图像特征,有效提升了接收端的重构质量。

为了进一步提高JSCC对SNR的自适应能力,Xu等^[20]结合了注意力机制与DeepJSCC,提出了AD-JSCC(attention DL-based JSCC)方案。Song等^[21]

则提出了一种新的语义自适应模型 AMJSCC (JSCC with adaptive model) 用于 SIT。AMJSCC 通过联合考虑动态信道条件与系统计算资源, 自适应地选择适当的 JSCC 模型和增强模型。为了在不同信道条件和图像内容下实现自适应的深度编码, Zhang 等^[22]在 ADJSCC 的基础上引入了掩码策略, 提出了用于可变码长编码的 DeepJSCC-V (variable code length-enabled DeepJSCC) 方案, 通过简单的掩码操作与特定压缩比, 实现了灵活的码率调整, 在不同 SNR 和图像内容条件下表现出优越的重构性能和适应性。DeepJSCC-V^[22]方案是本文视野范围内最先进的深度 JSCC 方案, 为未来 SIT 的研究提供了新的方向和方法。

尽管以上 5 种 JSCC 方案较好地缓解了传统数字通信中的“悬崖效应”, 并在图像重建质量方面取得了显著进展, 但大多数现有研究仅在加性白高斯噪声 (AWGN, additive white gaussian noise) 信道和二进制对称信道 (BSC, binary symmetric channel) 信道上展开, 仍存在一定的局限性。DeepJSCC^[18]、DeepJSCC-f^[19]以及 AMJSCC^[21]仅在单一 SNR 条件下训练模型, 可能导致系统在特定 SNR 条件下表现较好, 而应对 SNR 变化的自适应能力不足。ADJSCC^[20]虽通过在特定 SNR 区间内训练模型提高了系统鲁棒性, 但其研究仅限于 AWGN 信道, 忽略了实际通信中的复杂信道环境, 同时未充分考虑不同图像内容的自适应编码问题。DeepJSCC-V^[22]通过掩码策略实现了可变码长编码, 降低了带宽资源消耗, 获得了与 ADJSCC 相近的图像重构性能, 但是在某些 SNR 条件下与传统的 BPG+LDPC 编码相比, 仍未表现出显著优势, 其编码能力有待进一步提升。

为了解决现有 SIT 方案中的多种问题, 本文提出了一种深度灵活的信源信道联合编码 (DF-JSCC, deeply flexible joint source-channel coding) 模型以及传输性能预测网络 (PreNet, prediction network) 模型, 在多变的信道环境和图像传输内容条件下实现更加灵活且适应性更强的码率控制以及性能预测, 以进一步提高图像重构质量和优化传输策略。本文的具体贡献如下。

1) 提出了对于无线信道条件和系统压缩比 (CR, compression ratio) 深度灵活的可变码长编码 DF-JSCC 模型, 引入了多尺度特征融合模块 (MS-

FFM, multi-scale feature fusion module), 通过不同尺度的卷积操作, 使模型能够捕获到更加丰富的语义特征, 以提高对细节特征的捕获能力, 从而提升图像的重构质量和对抗多变信道的鲁棒性。

2) 在 DF-JSCC 中进一步融入了特征增强模块 (FEM, feature enhancement module), 以在多种 SNR 条件下自适应地调整图像语义特征图的通道权重, 从而增强特征表示能力。

3) 提出了基于 DF-JSCC 模型的图像重构质量预测模型 PreNet。该模型能够根据图像传输内容、SNR 和系统 CR, 实现对图像传输性能指标峰值信噪比 (PSNR, peak signal-to-noise ratio) 的高质量预测, 从而进一步指导 DF-JSCC 模型灵活地调整码长, 以适应多样化的信道环境和图像传输内容。

本文对所提出的 DF-JSCC 模型以及 PreNet 模型在 CIFAR10 和 ImageNet 数据集上进行训练, 并在 CIFAR100 和 Kodak 数据集上进行性能测试。实验结果表明, DF-JSCC 对多变信道条件和图像内容具有灵活的自适应性。与 DeepJSCC-V^[22]相比, DF-JSCC 实现了显著的 PSNR 性能优势以及更有效的带宽资源利用效率。此外, PreNet 能够对重构图像实现高质量的 PSNR 性能预测, 证明了其在 SIT 中的潜力。

1 SIT 系统模型

图 1 展示了深度灵活的 SIT 系统模型。该模型由发送端、接收端、物理信道、重构质量评估以及预测等关键部分组成。发送端通过语义提取、编码、掩码等操作将原始图像转化为重要语义特征。接收端通过对语义特征进行填充、解码等操作实现语义图像重构。重构质量评估负责衡量重构图像对原始图像的相似程度, 通过反馈评估结果以辅助优化图像重构算法。重构质量预测根据图像传输内容、信道条件以及系统 CR 对重构图像进行性能预测。

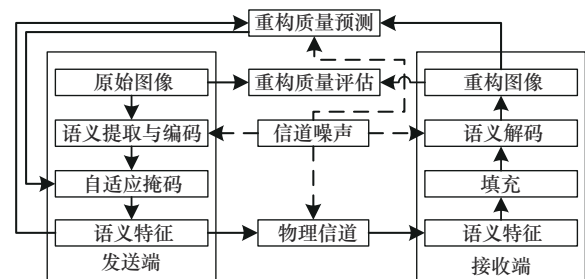


图1 SIT 系统模型

1.1 SIT 实现过程

如图 1 所示, SIT 系统模型的发送端包括语义提取与编码模块以及自适应掩码模块。设 U 维的原始图像为 $m \in \mathbb{R}^U$, $U = H \times W \times C$, 其中, H 表示图像 m 的高度, W 表示宽度, C 表示通道数。语义提取与编码模块采用卷积神经网络 (CNN, convolutional neural network) 作为编码器, 负责从 m 中提取主要的语义特征, 并通过融入 FEM 和 MS-FFM 增强细节特征表达。自适应掩码模块通过考虑语义重要性程度、数据冗余度及系统计算能力等因素, 对语义特征实现灵活压缩, 从而得到与图像传输任务高度相关的语义特征。语义编码的过程表示为

$$\mathbf{f} = E(m, \theta, \tau) \quad (1)$$

其中, $E(\cdot)$ 表示语义编码过程; $\mathbf{f} \in \mathbb{C}^{K_{\max}}$ 表示 K_{\max} 维的复数向量, 是对 m 编码后的输出特征, K_{\max} 表示允许的最大编码长度, 即 $K_{\max} = \frac{U}{2}$; θ 表示语义编码器的参数集; τ 表示信道 SNR。自适应掩码过程表示为

$$\mathbf{z} = \mathbf{f} \odot \boldsymbol{\varepsilon} \quad (2)$$

其中, $\mathbf{z} \in \mathbb{C}^{K_{\max}}$ 表示掩码后的 K_{\max} 维语义特征向量, $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_i, \dots, \varepsilon_{K_{\max}}]^T$ 表示二进制掩码向量, \odot 表示逐元素相乘。元素 ε_i 定义为

$$\varepsilon_i = \begin{cases} 1, & i \leq K \\ 0, & \text{其他} \end{cases} \quad (3)$$

当 $\varepsilon_i = 1$ 时, 将会发送 \mathbf{f} 的第 i 个向量。式(2)的作用是选择 \mathbf{f} 的前 K 个向量作为掩码特征 $\mathbf{x} \in \mathbb{C}^K$, 经物理信道传输到接收端。将语义编码与掩码过程统一表示为

$$\mathbf{x} = E(m, \theta, \tau, R) \quad (4)$$

其中, $R = \frac{K}{U}$ 表示系统压缩比, 即 \mathbf{x} 与 m 的维度比值。根据 K_{\max} 可知, 系统的最大压缩比为 0.5。

发送端将 \mathbf{x} 经物理信道传输到接收端。为了以端到端的方式优化通信系统, 必须允许物理信道能够反向传播, 因此将其融入整个神经网络, 并建模为不可训练层。信号的传输过程可以表示为

$$\mathbf{y} = \mathbf{h} \odot \mathbf{x} + \mathbf{n} \quad (5)$$

其中, $\mathbf{y} \in \mathbb{C}^{K \times 1}$ 表示接收的语义特征向量; $\mathbf{h} \in \mathbb{C}^{K \times 1}$ 表示信道增益矩阵; $\mathbf{n} \in \mathbb{C}^{K \times 1}$ 是 AWGN, 服从圆对称复高斯分布, 即 $\mathbf{n} \sim \text{CN}(0, \sigma^2 \mathbf{I})$, σ^2 表示噪声功率, $\mathbf{I} \in \mathbb{C}^{K \times K}$ 表示单位矩阵。

在接收端, 首先对 \mathbf{y} 进行 0 填充, 以确保语义特征向量的维度恢复至 K_{\max} , 填充长度为 $K_{\max} - K$ 。然后, 将填充后的特征向量输入解码器进行图像重构, 表示为

$$\hat{m} = D(\mathbf{y}, \varphi, \tau, R) \quad (6)$$

其中, $D(\cdot)$ 表示语义解码过程, $\hat{m} \in \mathbb{R}^U$ 表示重构图像, φ 表示语义解码器的参数集。

在通信过程中, SNR 的变化会对图像重构质量产生明显的影响。假设物理信道的噪声为 AWGN, 将 SNR 表示为

$$\tau = 10 \lg \left(\frac{\mathbb{E} |h_i|^2 P}{\sigma^2} \right) \quad (7)$$

其中, P 表示 \mathbf{x} 的平均功率, h_i 为传输 \mathbf{x} 的第 i 个特征向量的信道增益。 \mathbf{x} 需要满足功率约束条件 $\frac{1}{K} \mathbb{E} [\mathbf{x}^* \mathbf{x}] \leq P$, \mathbf{x}^* 表示 \mathbf{x} 的共轭转置。

1.2 语义图像重构质量评估与预测

图像重构质量评估通过量化重构图像与原始图像的差异, 辅助系统优化图像重构算法, 以提升重构效果并增强重构图像与原始图像的相似性。

本文采用 PSNR 作为图像重构质量度量指标, PSNR 越大, 表示重构图像与原始图像之间的差异越小, 重构质量越高。PSNR 通过计算原始图像与重构图像之间的均方误差 (MSE, mean squared error), 并结合图像的最大像素值来确定, 表示为

$$\text{PSNR} = 10 \lg \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (8)$$

其中, MAX 表示图像像素的最大可能值, 对于 8 位的灰度图像, MAX 值为 255。MSE 的计算式为

$$\text{MSE} = \frac{1}{HW} \sum_{j=0}^{H-1} \sum_{k=0}^{W-1} [m(j, k) - \hat{m}(j, k)]^2 \quad (9)$$

其中, $m(j,k)$ 和 $\hat{m}(j,k)$ 分别表示 m 和 \hat{m} 在 (j,k) 处的像素值。采用 MSE 来度量重构图像与原始图像的差异程度, 损失函数可以表示为

$$\mathcal{L}_{0,\phi}(m, \hat{m}; \tau, R) = \frac{1}{V} \sum_{i=1}^V d(m, \hat{m}) \quad (10)$$

其中, V 表示图像样本数量, $d(m, \hat{m}) = \frac{1}{n} \|m - \hat{m}\|^2$ 表示均方误差。

图像重构质量预测以语义编码和掩码后的语义特征向量为图像传输内容参照目标, 结合 SNR 以及系统 CR 对重构图像的 PSNR 性能进行预测, 以进一步指导系统在复杂通信环境中对 CR 的优化。可将性能预测建模为

$$\overline{\text{PSNR}} = S(x, \phi, \tau, R) \quad (11)$$

其中, $S(\cdot)$ 表示 PSNR 性能预测函数, $\overline{\text{PSNR}}$ 表示重构图像的性能预测值, ϕ 表示预测模型的参数。采用 MSE 作为预测误差的损失函数, 具体表示为

$$\mathcal{L}_{\phi}(x; \tau, R) = (\text{PSNR} - \overline{\text{PSNR}})^2 \quad (12)$$

2 DF-JSCC 模型及 PreNet 预测网络

为了提高复杂多变环境中的图像重构质量, 降低传输资源消耗, 本文提出一种新的 DeepJSCC 模型, 称为 DF-JSCC, 通过深度灵活的信源信道联合编码, 实现可变码长传输。同时, 为了预测特定信道和系统 CR 条件下的图像重构质量, 基于 DF-JSCC 模型提出了 PreNet 模型, 通过预测结果反馈, 实现对图像重构算法和 CR 的优化, 以进一步节约传输带宽资源。

2.1 DF-JSCC 模型

DF-JSCC 模型采用端到端的对称结构, 如图 2 所示。发送端主要包含编码和掩码模块。掩码模块遵循与 DeepJSCC-V^[22] 中的自适应掩码过程相同的设计原理, 通过在不同 CR 下对模型的自动化训练, 来灵活调整编码过程中的 CR。编解码器基于 CNN 设计, 编码器的输入为原始图像的像素值, 输出则是信道传输的符号。

DF-JSCC 模型的编码器由卷积层和 FEM 交替组成, 将原始图像作为输入, CNN 负责学习和提取全局特征, FEM 用于提取细节特征。此外, 在第一层卷积层之后, 引入 MS-FFM, 通过从不同尺度捕获图像特征且进行融合, 来增强特征表达能力, 以提升系统处理多样化图像的表现力和适应性。

1) 全局特征提取

在传统的神经网络模型中, 增加网络深度能够增强特征提取能力。然而, 当网络深度增加到一定程度时, 可能会出现梯度消失或梯度爆炸等问题。此外, 在模型训练过程中, 网络结构过深可能导致系统性能退化等问题。

为了避免出现上述问题, 本文设计的 CNN 由 5 个残差块 (ResBlock, residual block) 构成, 在每个 ResBlock 中通过采用广义除法归一化 (GDN) 非线性归一化操作优化输入和中间层特征, 从而抑制特定通道的过度激活。GDN 可以根据数据的特性自适应调整归一化过程, 以缓解梯度消失或梯度爆炸等问题。同时, 采用 PReLU 激活函数, 通过自主学习参数来适应不同的数据分布和特征。CNN 的参数设置如表 1 所示。

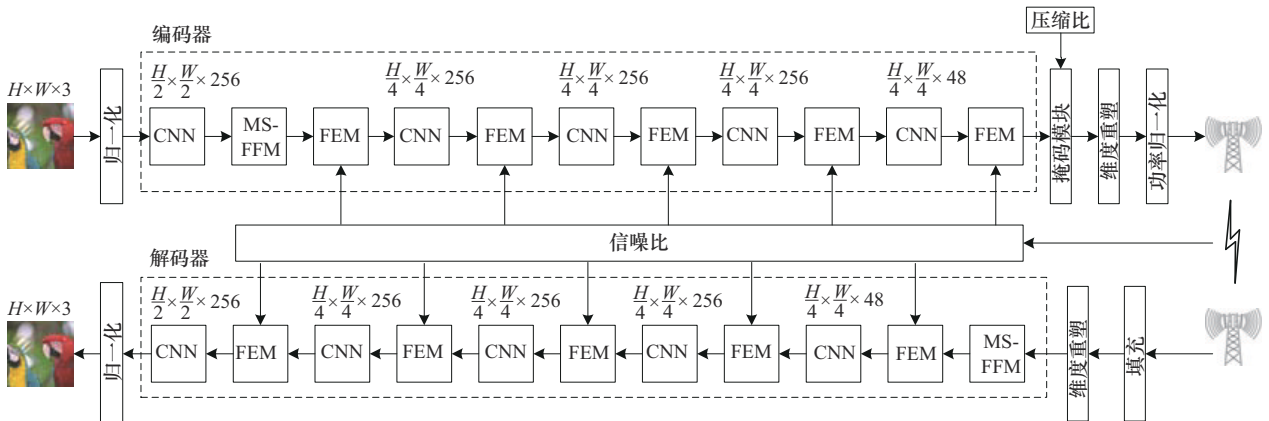


图 2 DF-JSCC 模型结构

表1 CNN的参数设置

名称	参数设置	输出
Conv1	GDN, PReLU	$\frac{H}{2} \times \frac{W}{2} \times 256$
	$\begin{bmatrix} 5 \times 5, 256, 2 \\ 1 \times 1, 256, 1 \\ 1 \times 1, 256, 1 \end{bmatrix}$	
Conv2	GDN, PReLU	$\frac{H}{4} \times \frac{W}{4} \times 256$
	$\begin{bmatrix} 5 \times 5, 256, 2 \\ 1 \times 1, 256, 1 \\ 1 \times 1, 256, 1 \end{bmatrix}$	
Conv3、Conv4	GDN, PReLU	$\frac{H}{4} \times \frac{W}{4} \times 256$
Conv5	GDN, PReLU	$\frac{H}{4} \times \frac{W}{4} \times 48$
	$\begin{bmatrix} 5 \times 5, 48, 1 \\ 1 \times 1, 48, 1 \\ 1 \times 1, 48, 1 \end{bmatrix}$	

假设输入的RGB图像尺寸为 $H \times W \times 3$ 。前2个ResBlock分别使用256个大小为 5×5 、步长为2的卷积核进行下采样，分别得到尺寸为 $\frac{H}{2} \times \frac{W}{2} \times 256$ 、 $\frac{H}{4} \times \frac{W}{4} \times 256$ 的输出特征。第3个和第4个ResBlock分别使用相同数量和大小、步长为1的卷积核进行特征提取，分别得到尺寸为 $\frac{H}{4} \times \frac{W}{4} \times 256$ 的输出特征。这一系列的卷积操作等价于信源编码过程。然后，第5个ResBlock使用48个大小为 5×5 、步长为1的卷积核进行输出通道调整，得到尺寸为 $\frac{H}{4} \times \frac{W}{4} \times 48$ 的输出特征，这一过程等价于信道编码。

2) 多尺度特征融合

金字塔池化是一种特征提取技术，通过对特征图进行多尺度的池化操作来捕获不同尺度的上下文信息，从而增强模型对不同对象大小和尺度的鲁棒性^[23]。受金字塔池化思想的启发，本文在特征提取中引入了MS-FFM，以增强模型对图像语义特征的深度理解，提高图像重构质量，设计原理如图3所示。

MS-FFM由3层组成。第一层的特征图 F_1 为CNN中首个ResBlock的编码输出，大小为 $\frac{H}{2} \times \frac{W}{2} \times 256$ 。然后，将第一层的输出作为第二层的输入，利用128个大小为 7×7 、步长为1的卷积核进

行卷积操作，得到输出特征图 F_2 ，大小为 $\frac{H}{2} \times \frac{W}{2} \times 128$ 。接着，将第二层的输出作为第三层的输入，利用128个大小为 5×5 、步长为1的卷积核进行卷积操作，得到输出特征图 F_3 ，大小为 $\frac{H}{2} \times \frac{W}{2} \times 128$ 。将特征图 F_1 、 F_2 和 F_3 在通道维度上进行拼接，得到大小为 $\frac{H}{2} \times \frac{W}{2} \times 512$ 的特征图 F_4 。最后，对 F_4 进行一维卷积操作，调整其通道数为原始特征图大小，从而得到最终的特征图 F 。通过这一系列不同尺度的卷积操作，可以融合来自多层图像的上下文特征，增强特征表达的丰富性，从而获得全局图像特征。

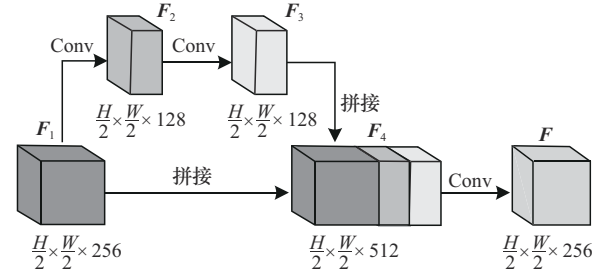


图3 MS-FFM原理

3) 细节特征增强

本文结合通道注意力和空间注意力算法，提出了一种新的FEM，设计原理如图4所示。

FEM以DF-JSCC模型中前一层MS-FFM或CNN的输出特征图和SNR作为输入，其主要思想是根据输入特征图中每个通道的特征向量和SNR来判断每个通道的重要程度以及信道对通道特征表达的影响程度。

FEM通过特征提取、拼接、激活、加权等操作实现。特征提取是通过对输入特征图分别进行全局平均池化和最大池化操作来捕获每个通道的特征信息。平均池化可以提取整体的特征信息，抑制异常值的影响，而最大池化能够突出显著特征。假设FEM模块的输入特征图为 $F = [F'_1, F'_2, \dots, F'_i, \dots, F'_{C'}]$ ， F 的维度可标记为 $U' = H' \times W' \times C'$ ， H' 、 W' 和 C' 分别表示 F 的高度、宽度和通道数。相应地，平均池化和最大池化特征图可分别标记为 $Q = [Q_1, Q_2, \dots, Q_i, \dots, Q_{C'}]^T$ 和 $T = [T_1, T_2, \dots, T_i, \dots, T_{C'}]^T$ ， $Q, T \in \mathbb{R}^{1 \times C'}$ 。 Q_i 以及 T_i 分别表示第 i 个通道的平均

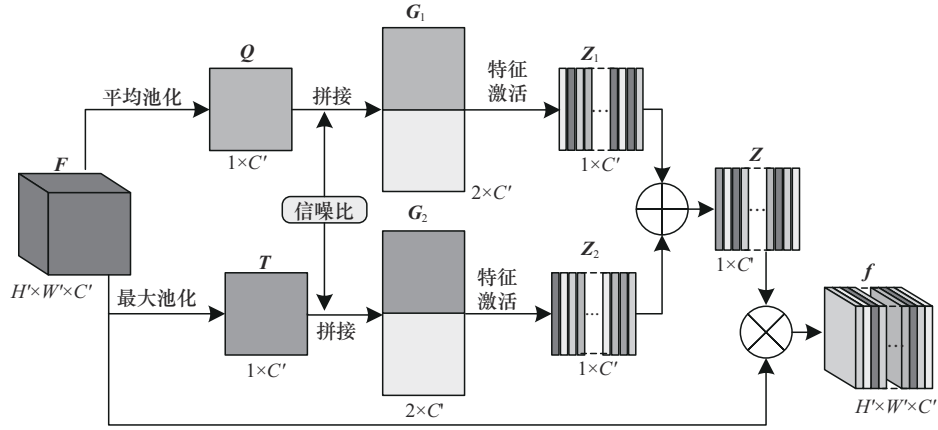


图4 FEM原理

池化特征和最大池化特征, 计算式分别为

$$Q_i = P_{ave}(F'_i) = \frac{1}{H' \times W'} \sum_{j=1}^{H'} \sum_{k=1}^{W'} F'_i(j,k) \quad (13)$$

$$T_i = P_{max}(F'_i) = \max_{j=1,2,\dots,H'} \max_{k=1,2,\dots,W'} F'_i(j,k) \quad (14)$$

其中, $P_{ave}(\cdot)$ 表示平均池化过程, $F'_i(j,k)$ 表示在位置 (j,k) 处的像素值, $P_{max}(\cdot)$ 表示最大池化过程。

为了提高特征对 SNR 的适应能力, 将池化结果 Q 以及 T 在空间维度上分别与 SNR τ 拼接, 得到特征向量 G_1 和 G_2 , 具体表示为

$$G_1 = \text{Concat}(Q, \tau) \quad (15)$$

$$G_2 = \text{Concat}(T, \tau) \quad (16)$$

对 G_1 和 G_2 分别进行一维卷积操作, 将输出空间维度调整为 1。然后, 对输出特征进行 ReLU 激活、一维卷积操作, 实现特征权值映射, 得到重要性权值 Z_1 和 Z_2 。将 Z_1 和 Z_2 逐元素相加, 并通过

Sigmoid 函数进行归一化, 从而得到所有通道最终的特征权值向量 Z , 表示为

$$Z = \lambda(Z_1 + Z_2) = \lambda(A) = \frac{1}{1 + e^{-A}} \quad (17)$$

最后, 将 Z 与输入特征 F 相乘, 得到加权后的特征图 $f = Z \times F$, 其维度与输入特征 F 相同。

2.2 PreNet 预测网络

为了充分发挥质量预测在 SIT 系统中的优势, 本文设计了基于 DF-JSCC 的预测模型 PreNet, 工作原理如图 5 所示。PreNet 以发送端编码器处理后的图像特征图 x 、信道 SNR、系统 CR 为输入, 通过特征提取与自动学习, 实现对重构图像 PSNR 性能的高质量预测。

PreNet 由特征分布建模、特征提取和预测输出 3 个部分构成。特征分布建模首先计算输入特征在所有通道上的均值 $u_x = [u_{x,1}, u_{x,2}, \dots, u_{x,i}, \dots, u_{x,C'}]^T$ 和标准差 $\delta_x = [\delta_{x,1}, \delta_{x,2}, \dots, \delta_{x,i}, \dots, \delta_{x,C'}]^T$, 其中, $u_{x,i}$ 表

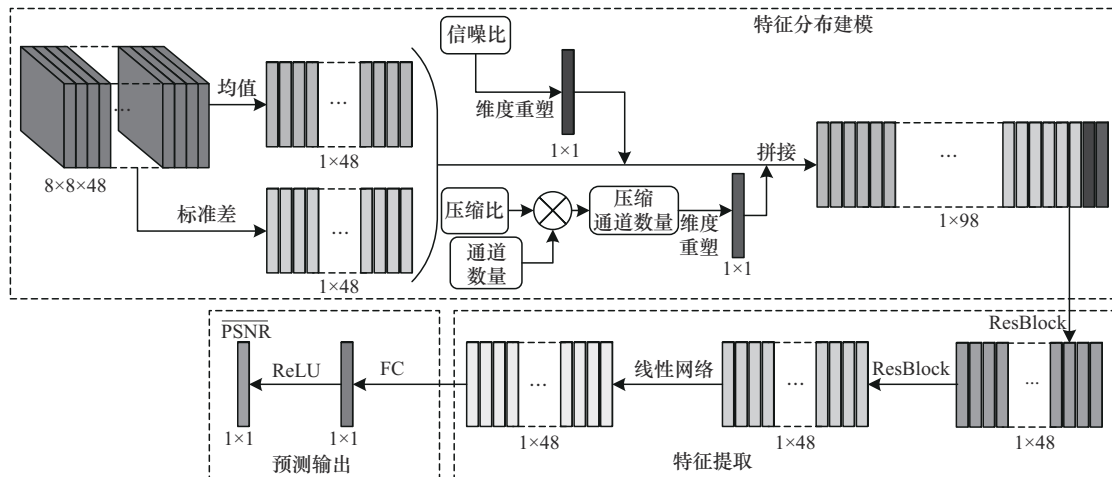


图5 PreNet原理

示第 i 个通道上的均值, $\delta_{x,i}$ 表示第 i 个通道上的标准差。通过提取预测图像的重要统计信息来实现特征表示。均值 μ_x 的计算过程表示为

$$\mu_{x,i} = \frac{1}{H \times W} \sum_{j=1}^H \sum_{k=1}^W x_{i,j,k} \quad (18)$$

其中, $x_{i,j,k}$ 表示第 i 个通道上的特征在 (j,k) 处的像素值。标准差 δ_x 的计算式为

$$\delta_{x,i} = \sqrt{\frac{1}{H \times W} \sum_{j=1}^H \sum_{k=1}^W (x_{i,j,k} - \mu_{x,i})^2} \quad (19)$$

为了使模型更好地理解不同条件下的图像质量变化, 从而增强其判别能力, 特征分布建模将 SNR 与 CR 预处理为与特征表示具有相同维度的特征向量, 并将其与特征表示在通道维度上进行拼接, 以实现多维特征融合。均值和标准差能够描述图像的内容特征, SNR 和 CR 则分别反映了系统传输条件和压缩条件。这种将不同来源信息相结合的特征分布建模方式能够结合图像统计特征、传输和压缩过程等信息, 实现综合特征表示, 使模型能够更灵活地适应不同的信道和压缩条件, 提高在多样化应用场景下的适应性和预测质量。因此, 式(11)可重新表示为

$$\overline{\text{PSNR}} = S(\mu_x, \delta_x, \phi, \tau, R) \quad (20)$$

进一步地, 根据式(12)可将损失函数重新表示为

$$\mathcal{L}_\phi(\mu_x, \delta_x; \tau, R) = (\text{PSNR} - \overline{\text{PSNR}})^2 \quad (21)$$

特征提取通过 2 个级联的 ResBlock 和一个线性网络实现。ResBlock 主要用于提取和变换输入特征, 通过增强模型的非线性表达能力, 来捕获复杂的特征关系。线性网络用于对 ResBlock 的输入特征实现进一步的变换, 通过特征映射来生成和学习更复杂的特征表示。

预测输出通过一个全连接 (FC, fully connected) 层和 ReLU 激活函数实现。FC 通过对特征表示从高维特征向量映射到一维标量值 ψ , 实现了重构质量预测值的生成。ReLU 激活函数通过引入非线性变换, 将预测结果中小于 0 的值置 0, 大于 0 的值则保持不变, 从而增加模型的非线性表达能力。

PreNet 对系统 CR 的进一步优化过程, 可建模为

$\min R$

$$\text{s.t. } S(\mu_x, \delta_x, \phi, \tau, R) \geq \text{PSNR}_{\text{th}}, 0 < R \leq 1 \quad (22)$$

其中, PSNR_{th} 为图像重构的 PSNR 阈值。

2.3 模型训练过程

DF-JSCC 模型能够实现 SIT 的深度灵活编码, 而 PreNet 需要以 DF-JSCC 的编码结果作为参考, 实现对 SIT 效果的高质量预测。DF-JSCC 与 PreNet 需要进行独立训练。

1) DF-JSCC

DF-JSCC 的训练目标是在信道 SNR 和系统 CR 的约束下, 通过对原始图像与重构图像的 MSE 均值最小化, 联合优化 SIT 系统的编解码器参数 θ 和 ϕ 。将目标函数定义为

$$(\theta^*, \phi^*) = \arg \min_{\theta, \phi} \mathbb{E}_{m, \tau, R} \mathcal{L}_{\theta, \phi}(m, \hat{m}; \tau, R) \quad (23)$$

在训练过程中, 分别设置约束条件 SNR $\tau \in [0, 27]$ dB 和 CR $R \in [0.05, 0.5]$, 参数 τ 和 R 由系统在区间范围内随机生成。通过对一定批量的样本进行训练, DF-JSCC 模型能够自动学习和压缩图像特征, 实现特定条件下对 SIT 的最优联合编解码过程。算法 1 详细描述了 DF-JSCC 模型单个 epoch 的训练过程。

算法 1 DF-JSCC 模型单个 epoch 的训练过程

输入 训练数据集 κ , 批次大小 B , 学习率 η

输出 编解码参数集 (θ, ϕ)

- 1) 从 κ 中采样一个批次样本 $\mathbf{M} = \{m_1, m_2, \dots, m_k, \dots, m_B\}$;
- 2) 循环 $m_k \in \mathbf{M}$
- 3) 随机生成 $\tau_k \in [0, 27]$ dB 和 $R_k \in [0.05, 0.5]$;
- 4) 通过式(1)得到语义编码向量 \mathbf{f}_k ;
- 5) 通过式(3)计算掩码向量 $\boldsymbol{\varepsilon}_k$;
- 6) 通过式(4)计算掩码后的特征向量 \mathbf{x}_k ;
- 7) 根据式(5)计算接收到的特征向量 \mathbf{y}_k ;
- 8) 根据式(6)重构图像 \hat{m}_k ;
- 9) 计算平均损失 $\mathcal{L} = \frac{1}{NB} \sum_{q=1}^N \sum_{k=1}^B \|m_{kq} - \hat{m}_{kq}\|^2$;
- 10) 更新模型参数 (θ, ϕ) 。

2) PreNet

PreNet 的训练目标是给定信道 SNR 条件和系统 CR, 通过优化网络参数 ϕ 提升 SIT 质量的预测精度, 以进一步指导系统对 CR 的优化。目标函数定义为

$$\phi^* = \arg \min_{\phi} \mathbb{E}_{\mathbf{x}, \tau, R} \mathcal{L}_{\phi}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\delta}_{\mathbf{x}}; \tau, R) \quad (24)$$

PreNet 依托已完成训练的 DF-JSCC 模型开展训练过程。对于输入图像 m ，通过 DF-JSCC 模型可以获得压缩后的语义编码向量 \mathbf{x} 以及相应的重构图像 \hat{m} ，并根据式(8)计算 SIT 质量的 PSNR 真实值。为了实现预测功能，以语义编码向量 \mathbf{x} 的均值 $\boldsymbol{\mu}_{\mathbf{x}}$ 和标准差 $\boldsymbol{\delta}_{\mathbf{x}}$ 作为特征表示来捕获图像特征的统计特性。在训练过程中，通过在网络结构中融入 SNR 和 CR，提高 PreNet 在不同条件下的适应能力和预测能力。SNR 和 CR 参数的生成方式与 DF-JSCC 模型的训练过程一致。通过 PreNet 的训练，能够实现对相应的 DF-JSCC 模型在不同信道条件和系统 CR 约束下的 SIT 质量的精准预测，即根据式(20)计算预测值 $\overline{\text{PSNR}}$ ，并通过预测质量反馈，实现对系统传输和参数的进一步优化，促进传输效率和资源利用率的提升。算法 2 详细描述了 PreNet 单个 epoch 的训练过程。

算法 2 PreNet 单个 epoch 的训练过程

输入 训练数据集 κ ，批次大小 B ，学习率 γ

输出 PreNet 参数 ϕ

- 1) 从 κ 中采样一个批次样本 $\mathbf{M} = \{m_1, m_2, \dots, m_k, \dots, m_B\}$
- 2) 循环 $m_k = \in \mathbf{M}$
- 3) 随机生成 $\tau_k \in [0, 27]$ dB 和 $R_k \in [0.05, 0.5]$;
- 4) 通过预训练模型 DF-JSCC 获得语义编码向量 \mathbf{x} 和重构图像 \hat{m}_k ;
- 5) 根据式(8)计算 m_k 与 \hat{m}_k 的 PSNR 真实值;
- 6) 根据式(18)和式(19)分别计算 \mathbf{x} 的均值 $\boldsymbol{\mu}_{\mathbf{x}}$ 和标准差 $\boldsymbol{\delta}_{\mathbf{x}}$;
- 7) 根据式(20)计算 DF-JSCC 模型的性能预测值 $\overline{\text{PSNR}}$;
- 8) 计算平均损失 $\mathcal{L} = \frac{1}{NB} \sum_{q=1}^N \sum_{k=1}^B (\text{PSNR}_{kq} - \overline{\text{PSNR}}_{kq})$;
- 9) 更新模型参数 ϕ 。

3 实验结果与分析

本节首先介绍仿真实验采用的数据集和参数设置，然后展示并分析 DF-JSCC 与 PreNet 在不同数据集上的仿真实验结果。

3.1 参数设置

采用数据集 CIFAR10 和 ImageNet 对所提方法进行模型训练，采用 CIFAR100 和 Kodak 进行性能测试。首先，采用 CIFAR10 的训练集和测试集分别作

为所提方法的训练集和验证集，对所提出的 DF-JSCC 和 PreNet 进行模型训练，采用 CIFAR100 的测试集对所提模型进行性能测试。CIFAR10 包含 10 个类别，每个类别有 6 000 张彩色图像；CIFAR100 包含 100 个类别，每个类别有 600 张彩色图像。CIFAR10 与 CIFAR100 数据集的图像均为 32 像素 \times 32 像素，都包含 50 000 张训练样本和 10 000 张测试样本。在 CIFAR10 之外，再从 ImageNet 的验证集中随机采样 50 000 张图像，将其中的 45 000 张图像作为训练集，其余的 5 000 张图像作为验证集。在训练前，将训练样本预处理为 128 像素 \times 128 像素。采用 Kodak 数据集对所提方法进行测试。Kodak 数据集包含 24 张高质量、无压缩的自然场景彩色图像，每张图像的分辨率为 768 像素 \times 512 像素。

表 2 详细描述了 DF-JSCC 的参数设置。PreNet 的训练过程与 DF-JSCC 类似，但由于 PreNet 结构的复杂性远低于 DF-JSCC，因此为了提高训练效率，减少训练样本，将 CIFAR10 数据集中的训练 epoch 设置为 150。通过对所提方法在多个数据集上的训练和测试，来验证对于不同分辨率的图像，DF-JSCC 的自适应传输能力以及 PreNet 灵活的性能预测能力。

表 2 DF-JSCC 参数设置

参数	取值	
训练集	CIFAR10(32像素 \times 32像素)	ImageNet(128像素 \times 128像素)
测试集	CIFAR100(32像素 \times 32像素)	Kodak(768像素 \times 512像素)
优化器	Adam	
学习率 γ	10^{-4}	
损失函数	MSE	
批次大小 B	128	24
epoch	400	50
深度学习框架	PyTorch	

本文将所提出的 DF-JSCC 与现有最先进的联合编码方法 DeepJSCC-V^[22] 以及传统的分离编码方法 BPG-LDPC 进行对比。在传统编码方案中，为了适应实时信道条件，引入了自适应调制与编码 (AMC, adaptive modulation and coding) 机制，采用了多个调制与编码组合方案 (MCS, modulation and coding scheme)。此外，将所提出的 PreNet 与

Oracle 网络^[22]进行对比, 以验证基于 DF-JSCC 的 PreNet 网络的性能预测优势。

3.2 CIFAR100 数据集下的实验结果

本节采用 CIFAR100 数据集将 DF-JSCC 与 DeepJSCC-V 在相同 CR 条件下的 PSNR 性能进行比较, 实验结果如图 6 所示。在训练过程中, SNR 和

CR 的取值分别在 [0, 27] dB 和 [0.05, 0.5] 范围内随机生成, 以更好地适应信道条件和系统 CR。为了公平比较, 除了编解码器结构中引入 FEM 和 MS-FFM 外, DF-JSCC 与 DeepJSCC-V 的语义掩码等结构相同。图 6 中的 FEM 与 MS-FFM 分别表示在实验中只引入相应的单个模块进行消融实验。

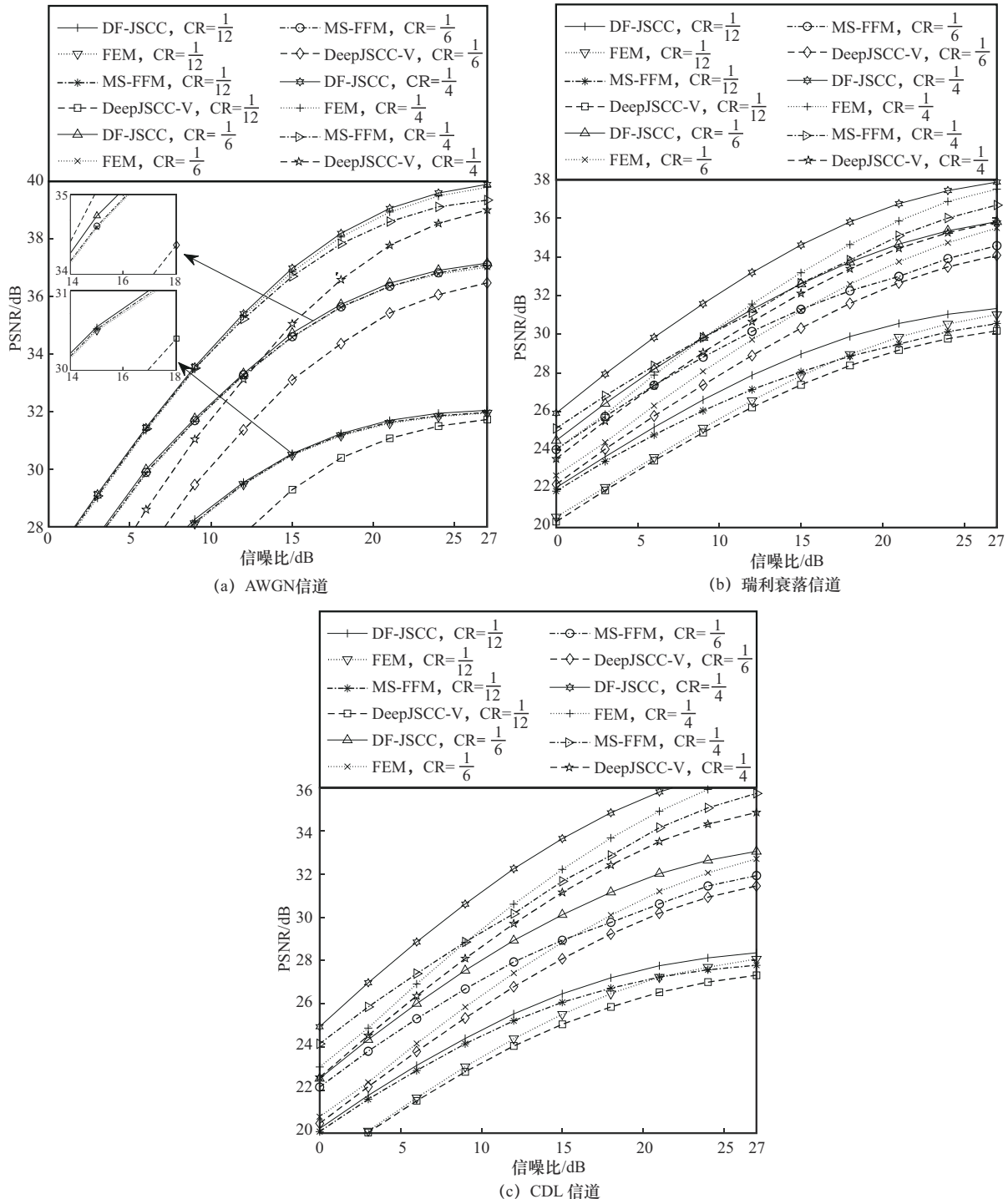


图6 DF-JSCC与DeepJSCC-V在相同CR条件下的PSNR性能对比

从图6中可以看出,在整个SNR范围内,DF-JSCC的PSNR在不同程度上均显著高于DeepJSCC-V。在AWGN信道中,当SNR=27 dB且CR= $\frac{1}{4}$ 时,DF-JSCC相较于DeepJSCC-V的PSNR提升了0.89 dB。随着SNR的降低,DF-JSCC与DeepJSCC-V之间的性能差距逐渐扩大,并且随着CR值的增大,DF-JSCC的性能优势更加显著,PSNR增益最高可达3.8 dB。虽然FEM与MS-FFM的PSNR值始终高于DeepJSCC-V,但是在低SNR区域,FEM和MS-FFM的PSNR值与DF-JSCC较为接近。在瑞利衰落(Fading, Rayleigh fading)信道中,FEM在高SNR区域的PSNR增幅较大,而在低SNR区域增幅相对较小。与此相反,MS-FFM在低SNR区域表现出更大的PSNR增益。相较于DeepJSCC-V,FEM和MS-FFM的PSNR均得到了显著提升,性能优势更加明显。例如,当SNR=15 dB且CR= $\frac{1}{16}$ 时,DF-JSCC、FEM和MS-FFM分别实现了约1.42 dB、0.41 dB、1.03 dB的PSNR增益。在聚类延迟线(CDL, clustered delay line)信道中,FEM和MS-FFM表现出与Fading信道中较为类似的现象。综合以上结果可以得出结论:FEM与MS-FFM在AWGN信道中的性能表现相似,而在Fading和CDL信道中则展现出互补性。具体而言,FEM在高SNR区域的PSNR提升较为显著,而MS-FFM则在低SNR区域表现出更大的性能提升。DF-JSCC通过同时融入FEM与MS-FFM模块,结合了两者的性能优势,在3种信道条件下均实现了平均超过1.4 dB的PSNR提升,表明其对信道状态具有较强的自适应能力以及应对复杂信道环境的鲁棒性。

图7展示了DF-JSCC与DeepJSCC-V在不同信道条件下的CR优化性能对比。从图7中可观察到,在相同信道条件下,经DF-JSCC优化得到的CR显著低于DeepJSCC-V,该优势在低SNR区域尤为明显。从图7(a)可见,当SNR低于6 dB时,DeepJSCC-V无论采用何种CR,均无法实现32 dB的PSNR。而DF-JSCC在SNR=3 dB、CR= $\frac{1}{3}$ 的条件下即可达到32 dB的PSNR。此外,当SNR=0 dB时,为实现26 dB的PSNR,DF-JSCC仅需使用DeepJSCC-V所需带宽资源的37.5%,节省了62.5%的带

宽消耗,充分体现出其在低信道质量下的资源节约能力。在Fading信道中,DeepJSCC-V在SNR低于6 dB时同样无法达到30 dB的PSNR,而DF-JSCC则可在SNR=3 dB、CR= $\frac{1}{3}$ 的条件下实现该性能目标。实验结果表明,在SIT任务中,DF-JSCC能够在相同信道条件下,以更低的带宽资源实现与DeepJSCC-V相当的传输性能;或者在相同带宽约束下,提供更高的图像质量和传输效率。由此可见,DF-JSCC在节约带宽和提高频谱效率方面具有明显优势。特别是在低SNR区域,DF-JSCC对CR的自适应能力和优化能力更加显著。

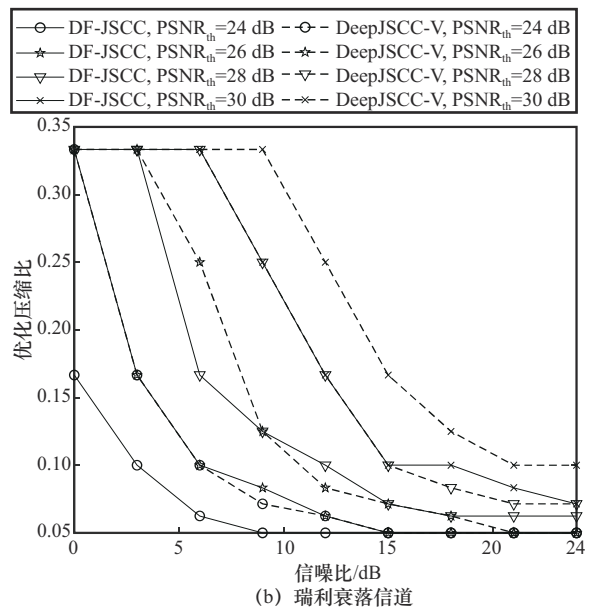
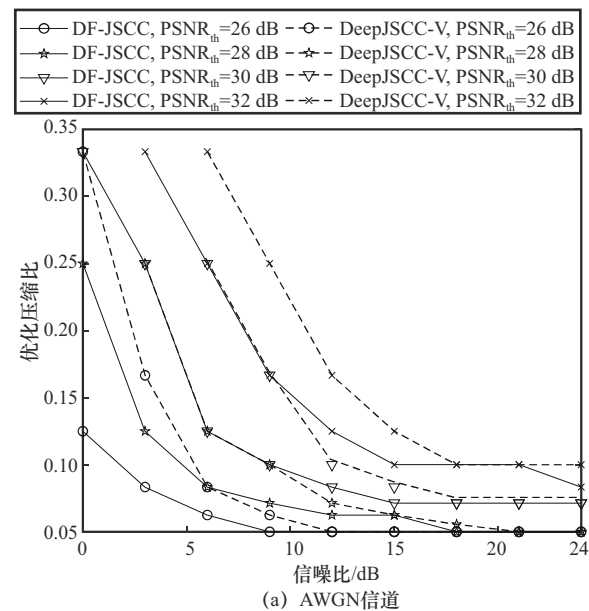


图7 DF-JSCC与DeepJSCC-V在不同信道条件下的CR优化性能对比

本文采用MSE函数对PSNR性能预判模型的测试值与预测值之间的误差进行度量。PreNet与OraNet的性能预测误差MSE曲面图对比结果如图8所示。从图8中可以看出,基于DF-JSCC的PreNet模型的MSE明显低于OraNet,表明PreNet的预测误差较小,具有更高的预测质量。此外,与PreNet相比,在AWGN信道中的低SNR区域且CR数值较小区间内,以及在Fading信道中的相同传输条件和高SNR区域且CR数值较大区间内,OraNet的预测误差波动较大,稳定性较差。PreNet在不同的信道环境中能够感知SNR和系统CR的真实情况,实现高质量的预测效果,尤其在复杂环境下可以提供更精确的预测结果,表现出更优越的自适应性和稳定性。

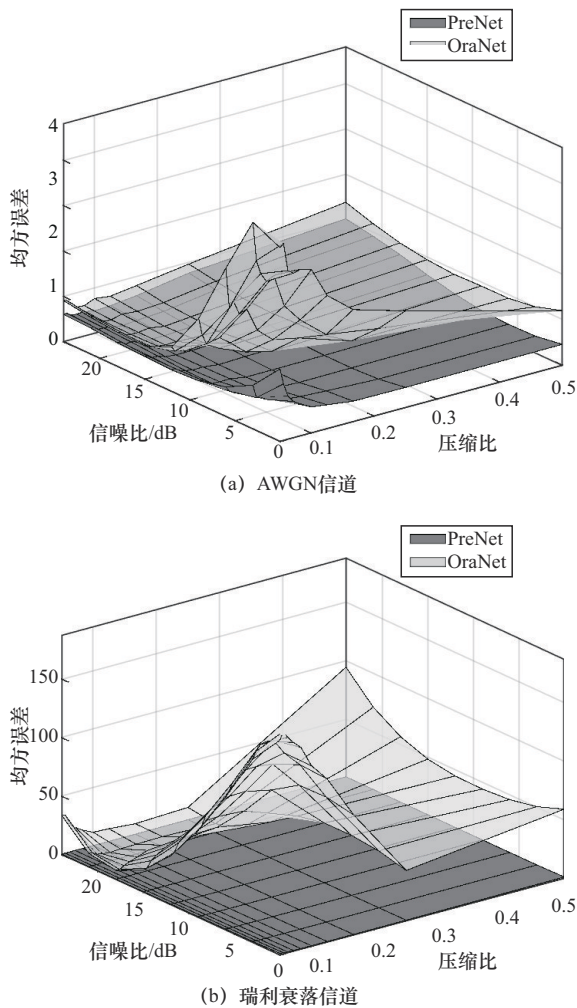


图8 PreNet与OraNet的性能预测误差MSE曲面图对比结果

本节通过仿真实验证明了在SIT中,所提出的DF-JSCC比DeepJSCC-V表现出了更好的适应性和

灵活性,在不同传输条件下实现了更高的PSNR。此外,进一步证明了与现有性能预测模型相比,所提出的PreNet能够降低预测误差,获得更高质量的PSNR。

3.3 Kodak数据集中的实验结果

3.2节已通过CIFAR100数据集上的实验证明了所提模型对低分辨率图像的处理优势。Kodak数据集中的图像细节更加丰富,本节在此数据集上开展仿真实验。

图9为DF-JSCC与传统BPG+LDPC方案在不同信道条件下的PSNR性能对比。BPG+LDPC是经典的分离编码技术,BPG能够实现高质量的图像压缩,LDPC提供了强大的纠错能力,两者的结合在带宽有限且信号存在干扰的通信环境中表现优异。在实验中,引入了AMC机制,预设了多种MCS组合,包括LDPC码率为 $\frac{1}{2}$ 、 $\frac{2}{3}$ 和 $\frac{3}{4}$ 等数值,以及BPSK、4QAM和16QAM等调制方式,以模拟传统分离编码方案在不同信道条件下对多种MCS的动态选择性能。为了在比较中使BPG+AMC占据码长优势,进一步增加了端到端压缩比(E2E-CR, end-to-end compression ratio)大于 $\frac{1}{4}$ 的对比方法。

从图9中可以看出,首先,BPG+AMC在高SNR区域图像重构质量表现较为出色,但是在低SNR条件下,重构性能急剧下降,出现明显的“悬崖效应”;DF-JSCC则随着SNR的减小,重构质量实现优雅的退化,有效缓解了传统分离编码的缺陷。其次,在2个信道环境中,随着SNR的增大,BPG+AMC的PSNR性能先于DF-JSCC趋于饱和,这可能导致高SNR条件下的资源浪费,并且限制了图像传输质量和系统效率的进一步提升。此外,当SNR=0 dB时,DF-JSCC在AWGN信道中的重构质量可以实现28 dB以上的PSNR,而在Fading信道中的重构性能最低也可以超过27 dB。从图9(a)中还可以看出,当CR= $\frac{1}{4}$ 时,DF-JSCC在SNR为[0,9]dB以及[12,20]dB范围的重构质量均高于BPG+AMC,PSNR的最大差值超过2 dB。在图9(b)中,DF-JSCC的优势更加显著,当CR= $\frac{1}{6}$,DF-JSCC在SNR低于12 dB范围内实现了比所有BPG+AMC基线方法更高的PSNR。当CR增大到 $\frac{1}{4}$,DF-JSCC在

所有 SNR 范围获得了比所有基线方法更出色的 PSNR，实现了更高的图像重构质量。

进一步地，将 DF-JSCC 与 DeepJSCC-V 在 Kodak 数据集中的 PSNR 性能进行对比。为了实现公平比较，DeepJSCC-V 的训练参数与 DF-JSCC 保持一致，实验结果如图 10 所示。随着 CR 的增大，码长逐渐增大，DF-JSCC 与 DeepJSCC-V 的图像重构质量均得到明显提升。从图 10(a)中可以看出，DF-JSCC 在低 SNR 条件下的 PSNR 显著高于 DeepJSCC-V，最大差值达到 3.72 dB。随着信道条件的改善，DF-JSCC 与 DeepJSCC-V 之间的 PSNR 差距逐渐减小，数值趋于一致。这种现象表明，与 DeepJSCC-V 相比，DF-JSCC 具备更强的抗噪能力。在 Fading 信道中，当 CR 分别为 $\frac{1}{16}$ 和 $\frac{1}{12}$ 时，随着 SNR 的弱化，DF-JSCC 相较于 DeepJSCC-V 的优势逐渐显著，PSNR 差距逐渐增大；当 CR 分别增大到 $\frac{1}{8}$ 、 $\frac{1}{6}$ 和 $\frac{1}{4}$ 时，DF-JSCC 在所有 SNR 范围内的优势保持稳定。在 CDL 信道中，当 CR 分别取值为 $\frac{1}{16}$ 、 $\frac{1}{12}$ 以及 $\frac{1}{8}$ 时，DF-JSCC 在低 SNR 条件下相

较于 DeepJSCC-V 的优势较为显著；当 CR 取值为 $\frac{1}{6}$ 和 $\frac{1}{4}$ 时，DF-JSCC 在整个 SNR 测试区间实现了稳定的 PSNR 增益。

此外，为了从视觉上比较 DF-JSCC 与 DeepJSCC-V 的图像重构质量差异，本文将样本 Kodim07 和 Kodim21 在 SNR=12 dB、CR= $\frac{1}{3}$ 条件下的重构图像展示在图 11 中。其中，图 11(a)和图 11(f)分别为 Kodim07 和 Kodim21 的原始图像，图 11(b)~图 11(e)分别为 DF-JSCC 和 DeepJSCC-V 在 AWGN 和 Fading 信道中对 Kodim07 的重构图像，图 11(g)~图 11(j)分别为 DF-JSCC 和 DeepJSCC-V 在 AWGN 和 Fading 信道中对 Kodim21 的重构图像。从图 11 中可以看出，通过 DF-JSCC 恢复的图像在色彩、纹理等方面更接近于原始图像，而利用 DeepJSCC-V 重构的图像可能丢失了部分细节信息，在背景、纹理、色彩等方面均存在比较明显的失真，这种现象在结构更为复杂的 Kodim21 样本中表现更为突出。具体来说，经 DF-JSCC 重构的 Kodim07 样本中的花朵颜色比 DeepJSCC-V 更鲜艳，且纹理更清晰。对于样本 Kodim21，经 DeepJSCC-V 重构的图像中

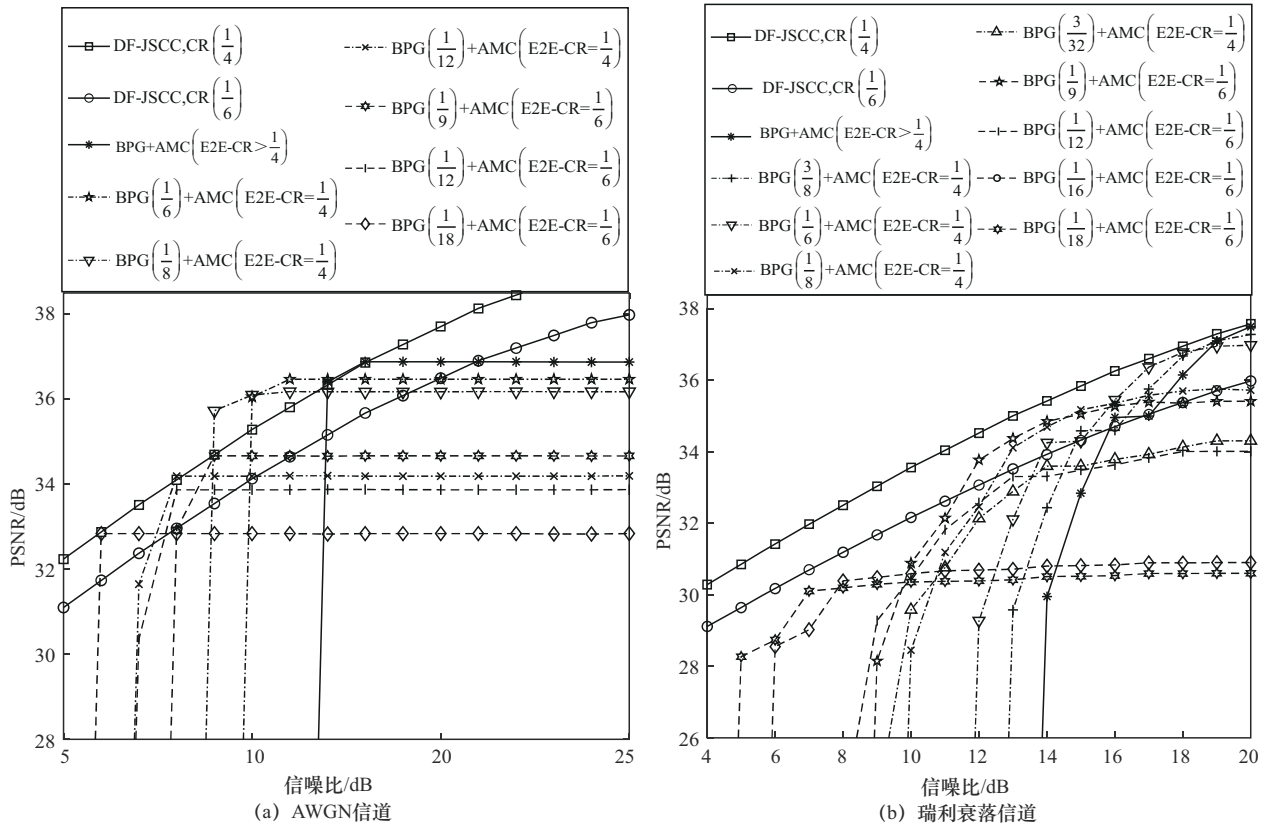


图 9 DF-JSCC 与传统 BPG+LDPC 方案在不同信道条件下的 PSNR 性能对比

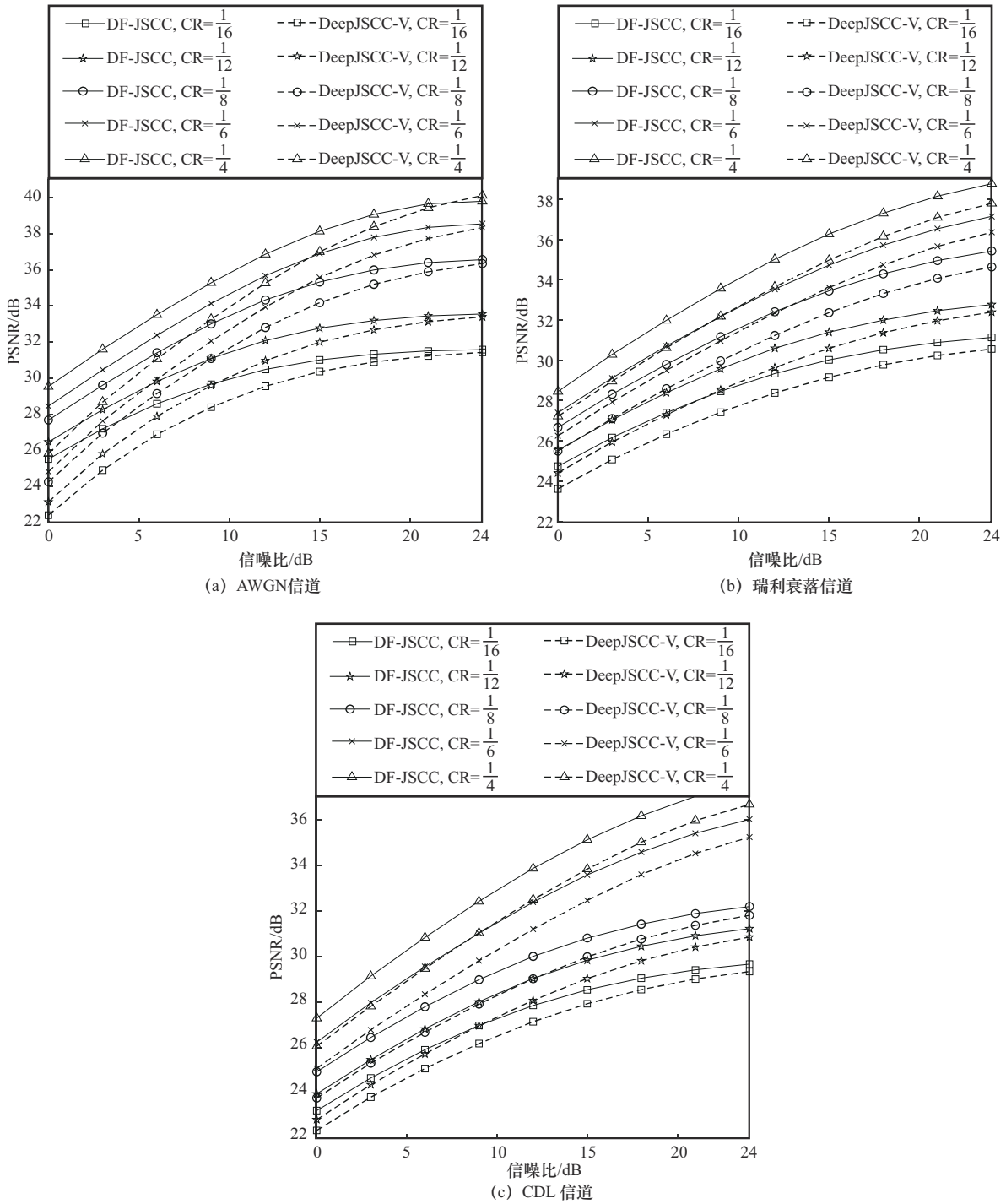


图 10 DF-JSCC 与 DeepJSCC-V 在 Kodak 数据集中的 PSNR 性能对比结果

的天空和护栏存在比较明显的模糊现象。DF-JSCC 在图像细节的捕获和还原中表现更为出色，实现了更高的保真度。实验结果表明，与 DeepJSCC-V 相比，DF-JSCC 不仅对高分辨率图像具有更高的图像重构精度，同时在复杂信道环境中对抗噪声表现出优越的自适应性、鲁棒性和高效性。

以样本 Kodim08、Kodim21 和 Kodim24 为例，

将 DF-JSCC 和 DeepJSCC-V 图像重构质量与 CR 之间的关系绘制成曲线，如图 12 所示。在 SNR 确定的情况下，随着 CR 从 0.05 逐渐增大到 0.5，3 个样本的 PSNR 均明显增大。在相同的 CR 条件下，对于所有的测试样本，DF-JSCC 比 DeepJSCC-V 获得了更高的 PSNR。此外，3 个样本的结构复杂度差异明显，Kodim21 和 Kodim24 的复杂度较低，

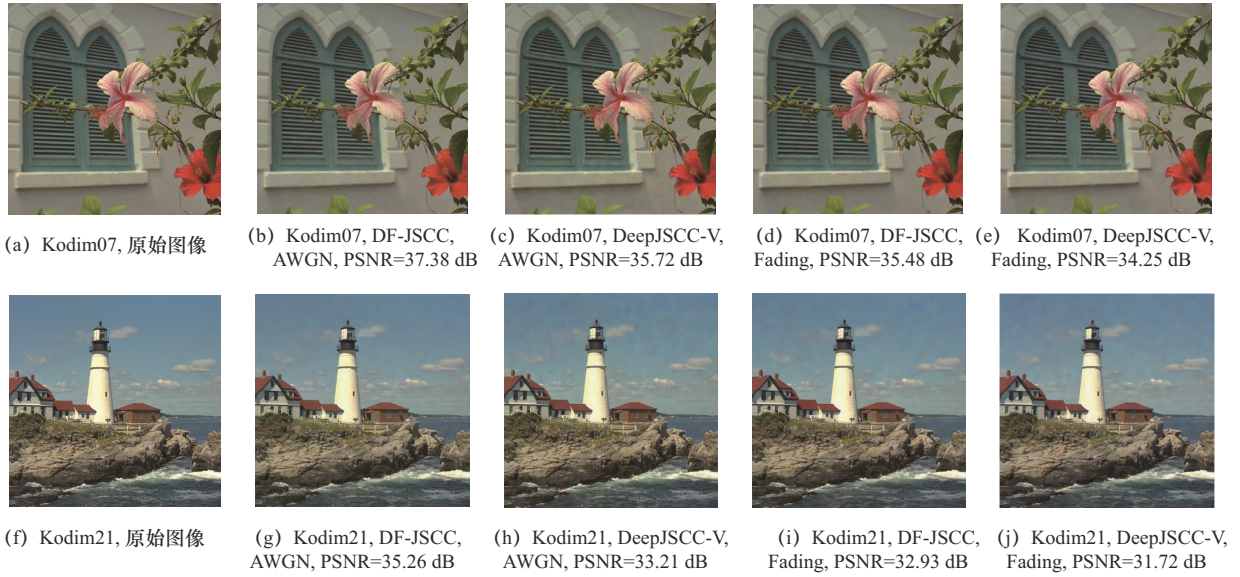


图 11 样本 Kodim07 和 Kodim21 在 SNR=12 dB、CR=1/3 条件下的重构质量对比

Kodim08 复杂度较高, 为了实现相同的图像重构质量, Kodim08 需要更大的 CR 来进行更多的信息传递, DeepJSCC-V 则比 DF-JSCC 需要更多的带宽资源来进行传输, 这种差异在图像传输质量需求更高的场景中表现更为突出。例如, 为了在 Fading 信道中获得 30 dB 的 PSNR, 如果通过 DF-JSCC 进行传输, Kodim21 仅需要 8.33% 的带宽, Kodim08 则需要 18% 左右的带宽。如果通过 DeepJSCC-V 对 Kodim08 进行传输, 需要 32% 左右的带宽。实验结果表明, DF-JSCC 在复杂环境中不仅对不同结构复杂度的图像具有更好的自适应性和鲁棒性, 而且在带宽利用效率方面表现更

加出色。

PreNet 与 OraNet 在 ImageNet 数据集上的训练损失比较结果如图 13 所示。虽然 PreNet 的结构复杂度略高于 OraNet, 但是, 从图 13 中可以明显看出, 在 AWGN 信道中, PreNet 迭代 10 次就可以达到收敛, 而 OraNet 需要迭代 20 次左右。在 Fading 信道中, PreNet 表现出相同的优势, 比 OraNet 更快接近收敛。此外, 与 OraNet 相比, PreNet 实现了更低 MSE 损失。这意味着网络结构复杂度的轻微提高, 使 PreNet 能够更好地捕获图像特征, 具备更强的表达能力, 提高了训练和优化效率, 在处理复杂信道时表现出更强的稳健性。

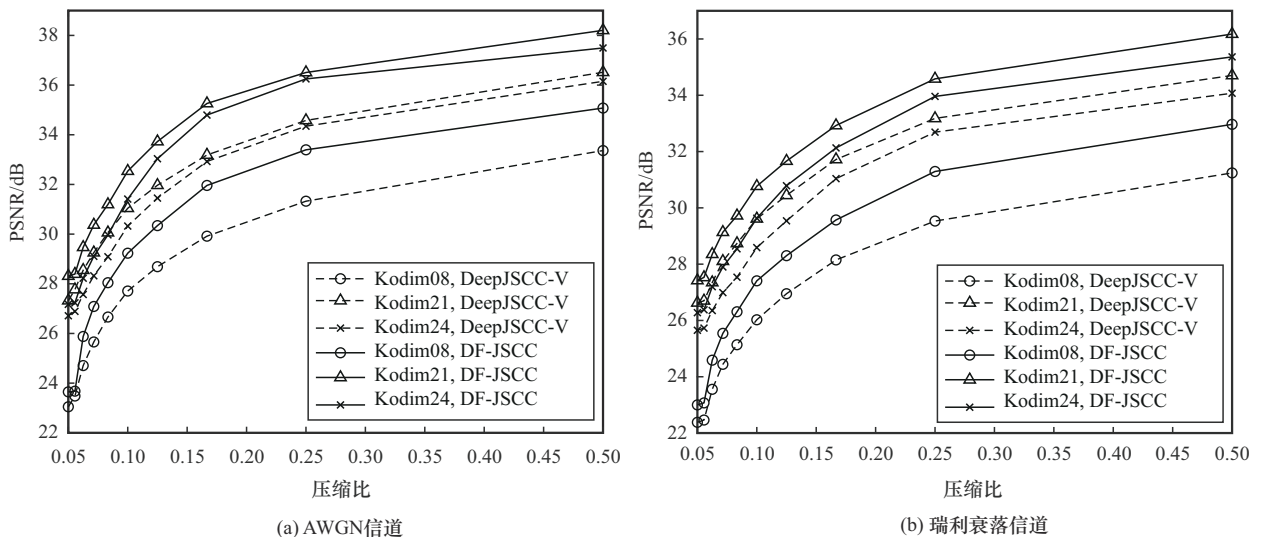


图 12 样本 Kodim08、Kodim21 和 Kodim24 在 SNR=12 dB 条件下的重构质量对比结果

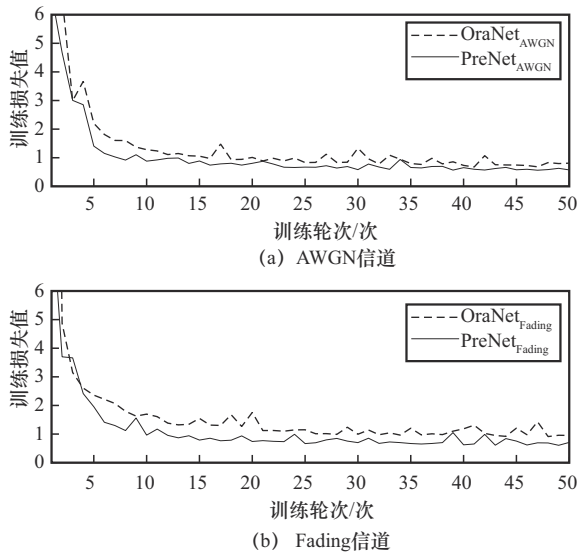


图 13 PreNet与OraNet在不同信道条件下的训练损失对比

以 Kodim08 和 Kodim21 为例，分别设置 $CR = \frac{1}{12}$ 、 $\frac{1}{8}$ 和 $\frac{1}{4}$ ，对样本在 AWGN 和 Fading 信道中的预测值 (Predict) 和测试值 (Test) 进行比较，结果如图 14 所示。从图 14 中可以看出，PreNet 的预测值与 DF-JSCC 的测试值的曲线是比较吻合的，在 SNR 为 $[0, 12]$ dB 的区域内，预测值与测试值非常接近，随着 SNR 的持续增大，预测值与测试值的差异略微增大，但最大差值仅 1.23 dB，预测最

大误差为 3.96%，依然保持在可接受的范围内。此外，从图 14 中还可以看出，在多数情况下，PreNet 在 Fading 信道中的 PSNR 预测差值比 AWGN 信道中更低，实现了更高的预测质量。实验结果表明，PreNet 能够实现高质量的 PSNR 预测，尤其在低 SNR 范围内和 Fading 信道中表现更为出色，对复杂环境具备较高的鲁棒性和适应性，这进一步证明了其在 SIT 中的潜力。

DF-JSCC 模型的计算复杂度主要来源于神经网络的系列卷积以及点乘、激活等密集运算。表 3 对比了 DF-JSCC 与 DeepJSCC-V 在 ImageNet 数据集集中的浮点运算量 (FLOP, floating point operation) 和训练时间。结果显示，DF-JSCC 的 FLOP 较 DeepJSCC-V 增加了 140.1%，训练时长增长了 78.2%。这表明，尽管 DF-JSCC 在引入了 MS-FFM 和 FEM 后模型的特征提取能力和编码灵活性得到了显著增强，在 PSNR 性能方面实现了最低 1.03 dB 以及最高 3.72 dB 的提升，但的确消耗了更高的计算开销。值得说明的是，该图像传输方案适用于对计算资源要求不敏感但对性能有较高需求的应用场景。随着边缘计算等技术的不断发展，DF-JSCC 在计算资源消耗方面的挑战有望得到进一步缓解。

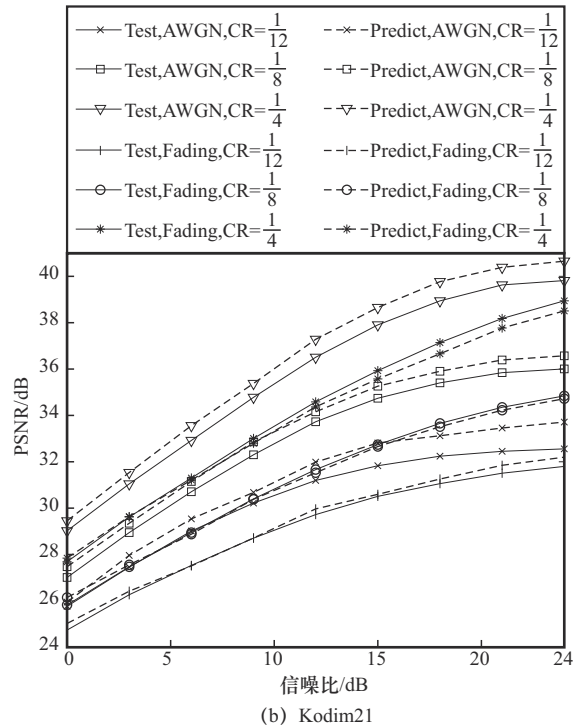
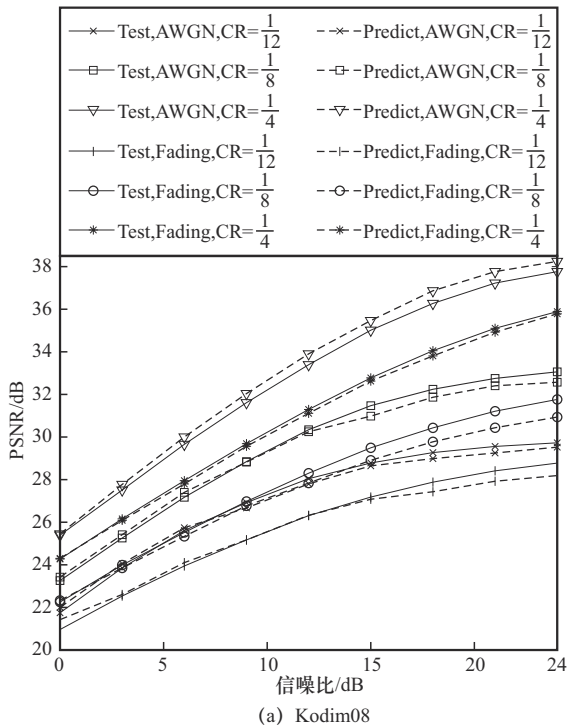


图 14 样本 Kodim08 与 Kodim21 的预测质量对比

表3 DF-JSCC与DeepJSCC-V运算复杂度对比

模型	FLOP	时间/ms
DF-JSCC	$1.881\,453\,296 \times 10^9$	$1.042\,5 \times 10^8$
DeepJSCC-V	$7.825\,569\,6 \times 10^8$	5.85×10^7

4 结束语

本文构建了一种具备深度灵活编码策略与图像传输质量预测网络的无线语义图像传输通信系统。具体而言,提出了一种DeepJSCC优化方案,即DF-JSCC,以及一种用于传输性能预测的PreNet模型。在DF-JSCC的架构中,依次引入了MS-FFM和FEM,以增强对细节的捕获能力和特征表达的丰富性,同时提高了对不同分辨率图像和复杂环境的适应能力。此外,详细描述了基于DF-JSCC的PreNet模型结构,并通过仿真实验验证了其预测质量。

实验结果表明,与DeepJSCC-V相比,DF-JSCC获得了至少1.03 dB的PSNR平均提升,在应对不同复杂度的传输内容、信道环境、系统CR等条件时均展现出出色的自适应编码能力和鲁棒性。相较于OraNet的平均MSE为0.313,PreNet的平均MSE较大幅度地降为0.25。所提出的无线语义图像传输通信系统在复杂信道环境中具备灵活地优化图像传输质量能力,在需要高保真图像传输的场景中(如远程医疗、图像监控)具有一定的价值。此外,通过对PSNR的高质量预测以及CR优化,该系统能够显著降低带宽资源消耗,有助于提升高密度用户环境中网络的整体吞吐量与带宽资源利用效率。

参考文献:

- [1] 张平. 语义通信: 未来通信系统的智简之道[J]. 中兴通讯技术, 2024, 30(1): 1.
ZHANG P. Semantic communication: the intelligent and concise way to the future communication system[J]. ZTE Technology Journal, 2024, 30(1): 1.
- [2] ZHANG P, XU W J, GAO H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks[J]. Engineering, 2022, 8: 60-73.
- [3] 秦志金, 赵葵葵, 李凡, 等. 多模态语义通信研究综述[J]. 通信学报, 2023, 44(5): 28-41.
QIN Z J, ZHAO T T, LI F, et al. Survey of research on multimodal semantic communication[J]. Journal on Communications, 2023, 44(5): 28-41.
- [4] LIANG C Y, LI D, LIN Z, et al. Selection-based image generation for semantic communication systems[J]. IEEE Communications Letters, 2024, 28(1): 34-38.
- [5] YANG W T, DU H Y, LIEW Z Q, et al. Semantic communications for future Internet: fundamentals, applications, and challenges[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 213-250.
- [6] GÜNDÜZ D, QIN Z J, AGUERRI I E, et al. Beyond transmitting bits: context, semantics, and task-oriented communications[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(1): 5-41.
- [7] 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.
- [8] 郭畅, 何占豪, 杨君刚, 等. 图像语义通信技术综述与展望[J]. 电讯技术, 2025, 65(2): 329-338.
GUO C, HE Z H, YANG J G, et al. Image semantic communication technologies: review and perspective[J]. Telecommunication Engineering, 2025, 65(2): 329-338.
- [9] GRASSUCCI E, MITSUFUJI Y, ZHANG P, et al. Enhancing semantic communication with deep generative models: an overview[C]//Proceedings of the ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2024: 13021-13025.
- [10] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 221-231.
- [11] HUANG D L, TAO X M, GAO F F, et al. Deep learning-based image semantic coding for semantic communications[C]//Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2021: 1-6.
- [12] DONG C, LIANG H T, XU X D, et al. Semantic communication system based on semantic slice models propagation[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(1): 202-213.
- [13] HU Q Y, ZHANG G Y, QIN Z J, et al. Robust semantic communications against semantic noise[C]//Proceedings of the 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). Piscataway: IEEE Press, 2022: 1-6.
- [14] 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.
- [15] FU Q F, XIE H Q, QIN Z J, et al. Vector quantized semantic communication system[J]. IEEE Wireless Communications Letters, 2023, 12(6): 982-986.
- [16] YE P G, SUN Y P, YAO S M, et al. Codebook-enabled generative end-to-end semantic communication powered by transformer[C]//Proceedings of the IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2024: 1-6.
- [17] 向际鹰, 段向阳, 冯雨龙. 语义编码与经典信道编码融合研究[J]. 中兴通讯技术, 2024, 30(S1): 24-32.
XIANG J Y, DUAN X Y, FENG Y L. Research on fusion of semantic

coding and classical channel coding[J]. ZTE Technology Journal, 2024, 30(S1): 24-32.

- [18] BOURTSOULATZE 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.
- [19] 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.
- [20] 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.
- [21] SONG M S, MA N, DONG C, et al. Deep joint source-channel coding for wireless image transmission with adaptive models[J]. Electronics, 2023, 12(22): 4637.
- [22] 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.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]// Computer Vision - ECCV 2014. Berlin: Springer, 2014: 346-361.

[作者简介]



杨舒涵 (1992-), 女, 山西运城人, 重庆邮电大学博士生, 主要研究方向为无线通信、语义通信等。



申滨 (1978-), 男, 贵州兴义人, 博士, 重庆邮电大学教授、博士生导师, 主要研究方向为认知无线电、信号处理、MIMO系统、语义通信等。



黄晓舸 (1982-), 女, 重庆人, 博士, 重庆邮电大学教授、博士生导师, 主要研究方向为无线通信资源管理、区块链技术等。