

模型-数据双驱动的低空图像语义通信方法

刘 任^{1,2}, 辛 利^{2,3}, 王帅军²

(1. 华南理工大学电子与信息学院, 广州 510000; 2. 广州海格通信集团股份有限公司研究院, 广州 510000; 3. 南京航空航天大学电子信息工程学院, 南京 211106)

摘 要: 低空空地通信信道具有时变多径与视距/非视距分量动态切换特征, 传统纯数据驱动联合信源信道编码方法缺乏无线信道先验信息, 存在图像重建鲁棒性弱、训练易振荡和收敛效果差等问题。为此, 本文提出一种模型-数据双驱动的端到端低空图像语义通信架构UP-WITT。该架构以低空时变信道时域基带传输模型为基础, 采用双分支协同架构设计: 数据驱动分支依托Swin Transformer构建编解码骨干网络, 完成图像多尺度全局语义特征提取, 并映射为满足传输规范的时域离散基带符号序列; 模型驱动分支将投影梯度下降(Projected gradient descent, PGD)迭代求解过程展开为可学习网络层, 搭建迭代步数可自定义的PGD重建模块。依托信道时域传输物理先验实现信道损伤迭代补偿, 同时结合频域快速计算降低迭代复杂度, 在当前实验条件下缓解了时变信道逆问题求解的病态程度。基于CIFAR数据集开展仿真测试, 结果表明, 相较于纯数据驱动基线模型, 所提方法在不同信道带宽比(Channel bandwidth ratio, CBR)及信道环境下, 峰值信噪比(Peak signal-to-noise ratio, PSNR)、多尺度结构相似性(Multi-scale structural similarity, MS-SSIM)均有明显提升, 且表现出更平滑的训练收敛过程与更好的信道自适应能力, 能够为低空场景可靠图像语义传输提供技术参考。

关键词: 通信与信息系统; 低空空地通信; 图像语义通信; 联合信源信道编码; 模型-数据双驱动

中图分类号: TN919.8

文献标志码: A

引用格式: 刘任, 辛利, 王帅军. 模型-数据双驱动的低空图像语义通信方法[J]. 数据采集与处理, 2026, 41(5): 1318-1332. LIU Ren, XIN Li, WANG Shuaijun. Model-data dual-driven image semantic communication for low-altitude scenarios[J]. Journal of Data Acquisition and Processing, 2026, 41(5): 1318-1332.

引 言

低空经济^[1-4]已成为我国战略性新兴产业之一, 是推动数字经济与实体经济深度融合的重要增长点。在无人机巡检、应急救援、城市安防及低空物流等典型应用场景中, 空地无线图像传输作为态势感知与信息交互的核心手段, 对通信的可靠性、实时性和带宽利用率提出了极高要求。传统通信系统基于香农分离定理, 采用信源编码、信道编码、调制、信道均衡、解调译码的分立设计架构, 在理想平稳信道条件下可逼近信道容量。然而, 低空空地通信信道具有快时变、多径效应显著以及视距(Line of sight, LOS)/非视距(Non line of sight, NLOS)分量动态切换的复杂特性: 无人机高速机动导致信道冲激响应快速变化, 多径传播会引发严重的符号间干扰(Inter symbol interference, ISI)。传统固定结构信道均衡算法高度依赖精确的信道状态信息(Channel state information, CSI), 在高动态低空场景中均衡性能显著退化; 同时, 分立架构难以同时兼顾图像传输效率与重建保真度, 无法满足低空应用的实际需求。

近年来,基于深度学习的联合信源信道编码(Joint source-channel coding, JSCC)与语义通信技术^[5-8]快速发展。该类方法采用端到端神经网络替代传统分立编解码架构,直接学习从信源输入到接收端重建的非线性映射关系,通过提取并传输核心语义信息而非原始比特流,在低信噪比和带宽受限场景下展现出显著优势,在图像、视频等多媒体传输任务中具备良好的应用前景。现有深度 JSCC 方法主要分为以下 3 类技术路线:(1)以卷积神经网络(Convolutional neural network, CNN)为骨干的纯数据驱动架构^[7-9],受限于局部感受野约束,对图像全局依赖关系的建模能力偏弱;(2)以无线图像传输 Transformer(Wireless image transmission Transformer, WITT)为代表的视觉语义通信框架^[10-11],借助自注意力机制建模长距离语义关联,图像传输性能优于 CNN 方案,为低空图像传输提供了可行技术思路;(3)基于生成模型的生成式 JSCC 方法^[12],通过生成对抗网络补全传输过程中丢失的细节信息,在极端恶劣信道条件下仍能保持较高的感知重建质量。与此同时,语义通信正从单任务独立传输向多任务协同共享方向演进,文献[13]提出了统一的层次化语义知识库架构,实现了不同任务语义信息的统一表征与高效共享,为低空场景下多任务并行传输提供了重要理论支撑。上述工作为低空图像语义通信奠定了技术基础,但现有方法仍存在物理先验缺失、信道均衡鲁棒性不足等核心问题,难以适配低空信道 LOS/NLOS 动态切换的高动态传输特性。

现有语义通信方案多采用纯数据驱动端到端设计,将信道均衡与语义重建耦合为黑箱映射方式,在低空时变信道环境中仍存在以下明显不足。(1)网络未引入无线基带传输与信道均衡的先验信息,完全依赖数据拟合补偿信道损伤;低空多径信道会加剧符号重建逆问题的病态程度,导致低信噪比和快衰落条件下图像重建鲁棒性显著下降。(2)黑箱式端到端映射缺少明确的均衡约束,难以针对低空 LOS/NLOS 动态切换特性进行定向优化,当实际信道与训练信道存在分布失配时,易出现性能断崖式下滑。(3)网络训练缺少收敛约束机制,迭代过程易发生梯度振荡、过拟合甚至数值不稳定问题;同时网络参数与传输过程的物理可解释性较弱,不利于理论分析和工程落地。

针对纯数据驱动语义通信方法的固有缺陷,模型驱动的深度展开(Deep unfolding/deep unrolling)技术^[14-15]近年来成为通信领域的研究热点。该技术的核心思想是将传统信号处理中具有严格收敛性的迭代优化过程,展开为可学习的神经网络层,通过嵌入信道先验信息引导神经网络的学习方向,既保留了模型驱动方法的物理可解释性与收敛稳定性,又具备数据驱动方法对复杂信道特性的自适应拟合能力。目前,深度展开技术已被成功应用于信道均衡、信号检测等多个通信物理层任务^[16-18],在低信噪比与快衰落场景下展现出显著优于纯数据驱动方法的鲁棒性。然而,现有深度展开语义通信方法大多针对通用高斯白噪声信道或静态多径信道设计,未充分嵌入低空时变信道的物理先验,无法直接适配低空空地通信的高动态传输需求。

针对上述问题,本文结合模型驱动与数据驱动的设计思路,提出一种模型-数据双驱动的低空无线图像语义通信方法 UP-WITT(Unrolling PGD WITT)。该方法以低空时变信道时域基带传输模型为物理基础,通过双分支结构融合互补,既利用了数据驱动网络对复杂语义特征的自适应拟合能力,又发挥了模型驱动算法在信道均衡中的可解释性与收敛稳定性。首先构建模型-数据双驱动端到端语义通信架构,以 Swin Transformer 搭建编解码分支,完成图像多尺度语义特征学习并映射为标准基带传输符号;同时引入展开式投影梯度下降(Projected gradient descent, PGD)模块,嵌入通信物理先验实现信道均衡与符号稳健重建,适配低空高动态传输场景。其次,建立适配低空空地链路的时域基带信道模型,结合莱斯分布刻画 LOS 平坦衰落分量,利用循环卷积表征 NLOS 多径干扰,精准描述低空信道传输特征,为可学习均衡模块设计提供理论支撑。最后,面向时变信道线性均衡的凸优化求解问题,将 PGD 迭代过程展开为可学习网络层,构建嵌入物理先验的迭代均衡模块,兼顾传统优化算法的收敛特性与深度学习的自适应拟合能力,有效改善逆问题病态性并提升网络训练稳定性。

本文后续安排如下:第1节建立低空时域基带信道模型与端到端语义通信系统模型,明确优化目标并分析纯数据驱动方法的固有局限;第2节详细阐述所提双驱动架构整体框架与核心模块设计,重点说明模型驱动可学习信道均衡的实现思路;第3节设置仿真实验方案并开展结果对比分析;第4节总结全文工作并对后续研究方向进行展望。

1 系统模型与问题描述

1.1 适配低空场景的时域基带信道模型

低空空地通信链路的最显著特征是LOS路径与NLOS路径呈现动态混合特性。当无人机处于开阔空域时,通信以LOS直射路径为主,莱斯因子取值较高,信道表现为平坦衰落特征;当无人机飞行至楼宇遮挡、城市峡谷等复杂环境时,NLOS多径分量占比显著增大,信道呈现频率选择性衰落,易引发严重符号间干扰。与此同时,无人机高速移动会使得信道时域冲激具备强快时变特性。

针对低空信道上述传播特点,本文引入块衰落假设:在单帧图像传输的相干时间内,信道时域冲激响应保持恒定;不同传输块之间,信道状态随时间动态变化。上述块衰落假设在低空高速飞行场景下依然成立,定量分析如下:以载波频率3.5 GHz、无人机最大飞行速度30 m/s(约108 km/h)为例,最大多普勒频移约为350 Hz,根据工程常用近似,对应的信道相干时间约为1.4 ms;本文单帧图像对应的基带符号序列在典型符号速率(如 10×10^6 Bd)下的传输时长仅为数十微秒量级,远小于信道相干时间,因此“单个传输块内信道冲激响应保持恒定”的假设具有充分的物理依据;不同传输块之间信道状态的动态演化,则通过信道预测模块在块间进行跟踪适配。此外,LOS与NLOS分量以加权系数 β 线性叠加的建模方式,源于空地(Air-to-ground)信道测量研究中广泛采用的莱斯两径/多径混合模型,刻画视距路径能量占比随飞行环境的动态变化。

在块衰落假设下,单传输块内的时域离散基带传输模型可表示为

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} + \mathbf{n} \quad (1)$$

式中: $\mathbf{y} = [y_1, y_2, \dots, y_N]^T \in \mathbb{C}^N$ 为发射端输出的时域离散复数基带符号序列, y_1 到 y_N 按时间先后顺序依次发射,每个符号对应一个独立的传输时隙,序列长度 N 由信道带宽比(Channel bandwidth ratio, CBR)约束; $\hat{\mathbf{y}} \in \mathbb{C}^N$ 为接收端收到的时域带噪符号序列,与发射符号一一对应,是接收端信道均衡模块的输入; $\mathbf{n} \in \mathbb{C}^N$ 为时域加性复高斯白噪声序列,服从分布 $\mathbf{n} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$, σ_n^2 为噪声功率,信道信噪比定义为 $\text{SNR} = 10 \lg(E[|y|^2]/\sigma_n^2)$,单位为dB; $\mathbf{H} \in \mathbb{C}^{N \times N}$ 为时域信道矩阵,精准适配低空信道LOS/NLOS动态切换的特性,建模为LOS平坦分量与NLOS多径分量的线性叠加,即

$$\mathbf{H} = \beta \mathbf{D} + (1 - \beta) \mathbf{C} \quad (2)$$

式中: $\beta \in [0, 1]$ 为LOS路径占比系数; $\mathbf{D} = \text{diag}(\mathbf{h}_d) = \text{diag}(h_{d,1}, h_{d,2}, \dots, h_{d,N}) \in \mathbb{C}^{N \times N}$ 为对角矩阵,对应LOS路径的时域平坦衰落,其第 i 个对角元 $h_{d,i}$ 为第 i 个传输符号对应的LOS路径复增益; $\mathbf{C} \in \mathbb{C}^{N \times N}$ 为循环矩阵,对应NLOS多径分量带来的时域符号间干扰,其第1行为多径信道的时域冲激响应 $\mathbf{h}_c = [h_{c,0}, h_{c,1}, \dots, h_{c,N-1}]^T \in \mathbb{C}^N$,其中 $h_{c,l}$ 为第 l 条多径的复增益,服从瑞利分布,符合无直射路径下散射多径的传输特性。

LOS路径复增益在标准复莱斯模型基础上引入扰动项,其精确表达式为

$$h_{d,i} = \sqrt{\frac{K}{K+1}} \cdot (e^{j\varphi} + \epsilon_i) + \sqrt{\frac{1}{K+1}} \cdot h_s \quad (3)$$

式中: K 为莱斯因子; $e^{j\varphi}$ 为确定性视距分量的相位项, φ 在 $[0, 2\pi)$ 内取值; h_s 为零均值、单位方差的循环对称复高斯随机变量,对应视距路径伴随的散射分量;为刻画低空场景下无人机姿态抖动、大气折射对

视距路径的影响,在确定性分量幅度上叠加零均值高斯扰动项 ϵ_i 。

本文所建信道模型能够真实反映低空空地信道传播规律:通过调整占比系数 β ,可精准适配LOS/NLOS分量的动态切换,对角矩阵 D 与循环矩阵 C 分别对应平坦衰落与频率选择性衰落两种典型信道场景,可为后续可学习信道均衡模块的设计提供完备物理基础。

为降低高维矩阵运算的计算复杂度,本文基于离散傅里叶变换的循环卷积定理,将时域循环卷积等价转换为频域逐点运算,通过快速傅里叶变换(Fast Fourier transform, FFT)实现信道矩阵前向传输与共轭转置的低复杂度计算,其数学等价形式为

$$\begin{cases} H\mathbf{y} = \beta \mathbf{h}_d \odot \mathbf{y} + (1 - \beta) \cdot \text{IFFT}(\text{FFT}(\mathbf{h}_c) \odot \text{FFT}(\mathbf{y})) \\ H^H \mathbf{r} = \beta \mathbf{h}_d \odot \mathbf{r} + (1 - \beta) \cdot \text{IFFT}(\text{FFT}(\mathbf{r}) \odot \text{Conj}(\text{FFT}(\mathbf{h}_c))) \end{cases} \quad (4)$$

式中:“ $\odot$ ”为哈达玛积(逐元素相乘), $\text{FFT}(\cdot)$ 与 $\text{IFFT}(\cdot)$ 分别表示快速傅里叶变换与逆变换, $\text{Conj}(\cdot)$ 为共轭运算, H^H 为信道矩阵的共轭转置。需要说明的是,本文系统在发送端对每个传输块插入长度不小于最大多径时延扩展的循环前缀(Cyclic prefix, CP),接收端去除CP后,多径信道对应的时域线性卷积等效转换为循环卷积,其信道矩阵因此具有循环矩阵结构,可基于离散傅里叶变换的循环卷积定理,通过FFT/IFFT在频域以逐点乘运算低复杂度实现信道前向传输与共轭转置。该频域等价变换仅作为工程实现的降复杂度手段,并未改变信道时域传输的物理本质。

1.2 端到端语义通信系统模型

本文所提的模型-数据双驱动端到端图像语义通信架构UP-WITT主要由3部分组成:数据驱动语义编解码分支、模型驱动可学习信道均衡分支以及低空无线信道。架构完整端到端传输流程如下:

(1)发射端输入原始图像 $\mathbf{x} \in \mathbf{R}^{H \times W \times 3}$,其中 H 和 W 分别为图像的高度与宽度,3为RGB通道数;数据驱动的语义编码器 $f_e(\mathbf{x}; \phi)$ 将原始图像映射为语义特征,再经全连接层转换为长度为 N 的时域离散基带符号序列 $\mathbf{y}$,其中 ϕ 为编码器的网络参数, $\mathbf{y}$ 经过全局功率归一化后满足发射机的平均功率约束。

(2)时域基带符号序列 $\mathbf{y}$ 按时间先后顺序经低空空地无线信道传输,引入信道衰落与加性噪声,接收端得到时域带噪符号序列 $\hat{\mathbf{y}}$ 。

(3)接收端通过模型驱动的展开式PGD迭代均衡模块,从带噪符号 $\hat{\mathbf{y}}$ 中完成信道损伤补偿与原始发射符号的鲁棒重建,得到均衡后的符号序列 $\tilde{\mathbf{y}} = f_{\text{PGD}}(\hat{\mathbf{y}})$;最后由数据驱动的语义解码器 $f_d(\tilde{\mathbf{y}}; \theta)$ 从均衡后的符号序列中恢复原始图像,输出重建图像 $\hat{\mathbf{x}} \in \mathbf{R}^{H \times W \times 3}$,其中 θ 为解码器的网络参数。

1.3 核心优化目标与问题分析

端到端语义通信系统以既定信道条件和带宽约束为前提,将最小化重建图像与原始图像间的失真作为核心优化目标,其数学表达形式为

$$\min_{\phi, \theta} E_{\mathbf{x} \sim p(\mathbf{x}), \hat{\mathbf{y}} \sim p(\hat{\mathbf{y}}|\mathbf{y})} [d(\mathbf{x}, \hat{\mathbf{x}})] \quad (5)$$

式中: $\mathbf{y} = f_e(\mathbf{x}; \phi)$, $\hat{\mathbf{y}} = H\mathbf{y} + \mathbf{n}$, $\hat{\mathbf{x}} = f_d(f_{\text{PGD}}(\hat{\mathbf{y}}); \theta)$, $p(\mathbf{x})$ 为图像数据的分布, $d(\cdot, \cdot)$ 为失真度量函数。本文采用均方误差(Mean square error, MSE)与多尺度结构相似性(Multi-scale structural similarity, MS-SSIM)作为优化损失函数,分别对应PSNR与视觉质量的优化目标。

接收端符号均衡与图像重建本质属于线性逆问题,即在已知时域带噪符号序列 $\hat{\mathbf{y}}$ 的条件下,恢复原始发射符号序列 $\mathbf{y}$ 。纯数据驱动语义通信方法将逆问题求解与语义重建融合为端到端黑箱映射,省略了信道均衡的物理建模过程,应用于低空通信场景时存在明显固有缺陷。

在低空时变多径环境下,时域信道矩阵 H 呈现高维稠密特性,矩阵条件数随LOS/NLOS分量动态

切换实时变化,导致逆问题具备较强病态性。纯数据驱动方案未引入信道先验信息约束,在低信噪比环境中图像重建效果会明显退化。同时,黑箱均衡效果完全依赖训练集的信道分布特征,当实际低空环境的莱斯因子、多径数量与训练样本存在偏差时,均衡性能会出现明显衰减,难以适配信道分量动态切换的高动态场景。此外,纯数据驱动网络缺少明确的迭代收敛约束,端到端训练过程容易出现梯度振荡、梯度消失甚至数值不稳定现象,在低空动态信道条件下模型收敛难度增大,泛化能力也难以得到有效保障。

为克服纯数据驱动方法的上述缺陷,本文引入模型驱动的优化展开思想,将信道均衡的凸优化求解过程拆解为可学习网络层,依托信道先验信息引导神经网络实现自适应信道均衡,有效弥补了纯数据驱动方案的固有不足。

2 模型-数据双驱动的语义通信方法

2.1 整体端到端架构

本文以现有语义通信神经网络 WITT^[10]为基础框架进行改进与拓展,构建模型-数据双驱动的端到端语义通信架构,整体结构如图1所示。所提方法分为数据驱动与模型驱动两条设计主线:语义特征编解码部分沿用 WITT 网络设计思路,信道均衡环节则融入信道先验信息,新增展开式 PGD 均衡模块以提升信道适应与损伤补偿能力,两部分协同构成本文算法核心框架。改进后的网络主要包含4大功能模块: Swin Transformer 语义编码器、信道自适应调制模块、模型驱动展开式 PGD 均衡模块和 Swin Transformer 语义解码器。

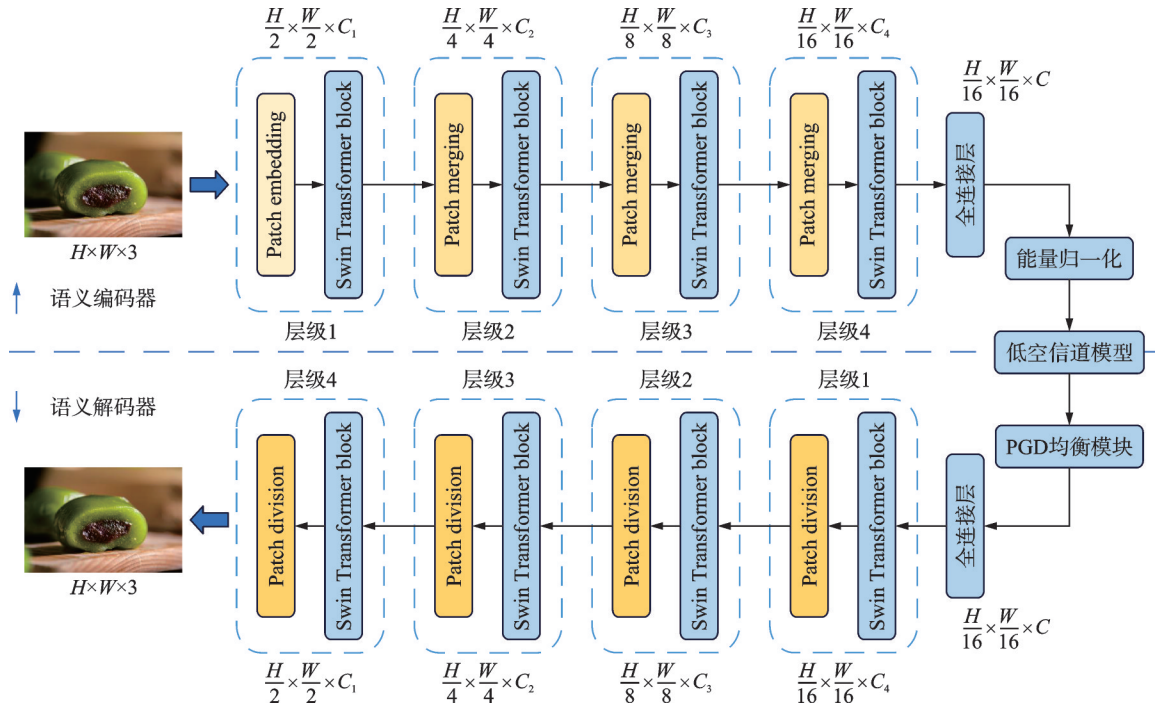


图1 模型-数据双驱动的端到端语义通信架构

Fig.1 Model-data dual-driven end-to-end semantic communication architecture

所提架构的端到端传输流程遵循双驱动协同设计逻辑。原始图像首先经 Patch embedding 映射为 Token 序列,输入分层 Swin Transformer 编码器,依托移位窗口自注意力机制提取多尺度全局语义特征,完成图像语义压缩,发挥数据驱动网络对复杂特征的自适应拟合优势。编码器输出的多尺度语义特征经由全连接层变换,生成长度为 N 的时域离散基带符号序列;再通过信道自适应调制与全局功率归一化处理,在满足发射机平均功率约束后,按时序送入低空无线信道进行传输。

传输符号序列经历时变衰落及加性噪声干扰,接收端最终获得含噪时域符号序列。随后将带噪符号输入展开式 PGD 均衡模块,以低空信道物理传输特性为先验支撑,通过可自定义迭代步数的可学习优化策略,实现信道损伤精准补偿与发射符号稳健重建,充分发挥模型驱动方法在收敛稳定性与物理可解释性上的优势。均衡后的符号序列经全连接层重构为语义特征,送入结构对称的 Swin Transformer 解码器,由 Patch division 层完成特征上采样,逐步恢复图像空间分辨率,最终输出重建图像。

该架构的核心设计思路是突破纯数据驱动端到端黑箱映射的局限,将信道均衡环节从网络黑箱中解耦出来,利用模型驱动的信道先验约束引导神经网络实现自适应均衡,实现模型驱动与数据驱动机制的有机结合。其中数据驱动分支侧重完成图像语义特征的提取与恢复,模型驱动分支专注于信道损伤的均衡与补偿,二者分工协作,可有效适配低空场景高动态、高可靠的图像传输需求。

2.2 数据驱动的语义编解码骨干

面向低空场景图像传输的应用需求,本文选取 Swin Transformer 作为数据驱动分支的编解码骨干网络。相比于传统 CNN 模型, Svin Transformer 依托移位窗口自注意力机制,可在线性计算复杂度条件下建模图像全局语义依赖,克服了 CNN 局部感受野带来的表征局限;相比于原生 Vision Transformer,其分层多级结构能够构建多尺度特征图,更契合图像编解码任务对多尺度特征提取与还原的设计要求。

2.2.1 语义编码器

语义编码器采用分层级联结构,每一级由 Patch embedding/Patch merging 层与 Swin Transformer block 级联构成。图 1 以 $256 \text{ 像素} \times 256 \text{ 像素}$ 的 RGB 图像与 4 层 Swin Transformer block 级联为例,具体设计如下:

输入 RGB 图像首先经过 Patch embedding 层,被划分成互不重叠的 4×4 图像块;每个图像块展平为 48 维一维向量,再通过线性映射变换为 128 维 Token 序列,完成由原始图像到语义 Token 的表征转换。

Swin Transformer block 为网络基础特征提取单元,由窗口多头自注意力(Window multi-head self-attention, W-MSA)、移位窗口多头自注意力(Shifted W-MSA, SW-MSA)以及多层感知机(Multi-layer perceptron, MLP)交替堆叠而成,层间配置层归一化(Layer normalization, LN)与残差连接以保障训练稳定性。连续两个模块依次采用 W-MSA 与 SW-MSA 结构,通过窗口移位策略实现不同局部窗口间的信息交互,在控制计算开销的同时完成图像局部与全局语义特征联合提取,模块详细结构如图 2 所示。

Patch merging 层负责特征下采样,对相邻 2×2 窗口内的 Token 进行拼接与线性映射,将 Token 数量缩减至原有 $1/4$,同时把特征维度提升一倍,逐步构建多尺度分层特征。本文 4 级编码器对应的特征维度依次为 128、192、256、320,记为 $C_1 \sim C_4$;各级配置的 Swin Transformer block 数量分别为 1、1、2、6。

编码器输出的多尺度语义特征经由全连接层映射,转换为时域离散基带符号序列,实现图像语义特征到通信传输符号的映射。本文定义信道带宽压缩比为 $R = \frac{C}{2 \times 3 \times 2^n \times 2^n}$,其中 n 为语义编码器层级数。

2.2.2 语义解码器

语义解码器采用与编码器对称的四级分层结构,以 Patch division 层替代编码端的 Patch merging 层,实现特征图 2 倍上采样,逐级恢复图像空间分辨率。解码端每一级由 Patch expanding 层与 Swin Transformer block 级联构成,最终通过线性映射将 Token 序列还原为 RGB 图像,完成图像重建输出。

2.3 模型驱动的展开式 PGD 迭代均衡模块

依托本文建立的低空时域基带信道模型,以 PGD 优化理论为基础,将信道均衡凸优化问题的迭代求解过程展开为可学习网络层,构建模型驱动的展开式 PGD 迭代均衡模块。该模块引入信道先验约束引导神经网络完成自适应信道均衡,同时保留深度学习方法对复杂信道环境的自适应拟合特性。

2.3.1 基于 PGD 的线性均衡基本原理

低空信道下的符号传输可抽象为线性逆问题,考虑发射功率约束,对应的最小二乘线性均衡优化模型可写为

$$\min_{\mathbf{y}} \frac{1}{2} \|\mathbf{H}\mathbf{y} - \hat{\mathbf{y}}\|_2^2 + \iota_{\mathcal{X}}(\mathbf{y}) \quad (6)$$

式中 $\iota_{\mathcal{X}}(\mathbf{y})$ 为可行域 $\mathcal{X}$ 的指示函数,当 $\mathbf{y}$ 满足发射功率约束时取值为 0,否则为无穷大。该优化问题属于带凸约束的凸优化问题,可通过投影梯度下降算法迭代求解,其标准迭代格式为

$$\begin{cases} \mathbf{g}_k = \nabla f(\mathbf{y}_k) = \mathbf{H}^H(\mathbf{H}\mathbf{y}_k - \hat{\mathbf{y}}) \\ \mathbf{z}_k = \mathbf{y}_k - \eta_k \mathbf{g}_k \\ \mathbf{y}_{k+1} = \mathcal{P}_{\mathcal{X}}(\mathbf{z}_k) \end{cases} \quad (7)$$

式中 k 为迭代步数, $f(\mathbf{y}) = \frac{1}{2} \|\mathbf{H}\mathbf{y} - \hat{\mathbf{y}}\|_2^2$ 为最小二乘损失函数, $\mathbf{g}_k$ 为第 k 步的梯度,对应均衡过程的残差补偿方向; $\mathbf{H}^H$ 为信道矩阵的共轭转置; η_k 为迭代步长,控制残差补偿的幅度; $\mathcal{P}_{\mathcal{X}}(\cdot)$ 为功率归一化算子,对迭代结果进行逐样本能量归一化,保障迭代过程数值稳定。需要指出的是,本文采用的逐样本单位能量归一化与优化理论中标准的功率约束投影并不严格等价:标准投影仅在信号超出功率约束时将其缩放至可行域边界,而本文操作在信号已满足功率约束时仍会改变其幅度,本质上更接近一种保障数值稳定的网络归一化手段。本文采用该设计是为了与发射端的全局功率归一化保持一致。

传统 PGD 迭代具备完善的收敛理论支撑,可稳定求解线性均衡逆问题,但实际应用中依赖精确信道状态信息与固定迭代参数,难以适配低空莱斯时变信道的动态变化特征。为此本文将 PGD 迭代机制展开为可学习网络层,在继承凸优化迭代结构与物理可解释性的同时,赋予网络自主拟合信道时变特性的能力。

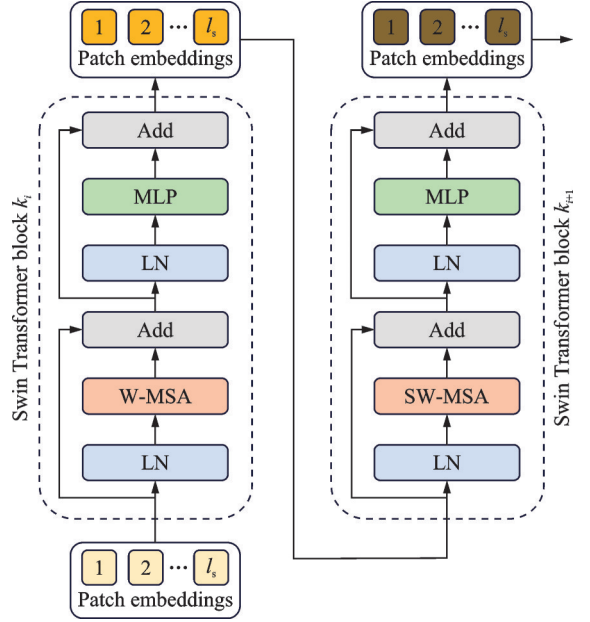


图2 多层次串联 Swin Transformer block 结构

Fig.2 Structure of the cascaded multi-stage Swin Transformer block

2.3.2 展开式PGD迭代均衡模块的详细设计

本文将PGD线性均衡完整迭代流程映射为多级可学习网络层,单次迭代对应一个网络子层,迭代步数可根据信道环境与性能需求灵活配置。模块核心设计思路如下:针对本文所提出的低空信道混合结构 $H = \beta D + (1 - \beta)C$,设计了轻量卷积网络从接收的时域带噪符号中预测信道的LOS分量冲激响应 h_d 与NLOS多径冲激响应 h_c 。网络输入为接收符号序列实部与虚部构成的双通道一维序列;主体由3层一维卷积级联构成,卷积核尺寸依次为7、5、3,输出通道数依次为32、64、128,步长均为1,采用零填充以保持序列长度不变;除输出层外,各卷积层后接批归一化(BatchNorm 1d)与ReLU激活函数;网络末端经两个相互独立的全连接层分别输出LOS平坦衰落分量的冲激响应(N 维复向量,以实部、虚部拼接形式表示)与NLOS多径冲激响应(L 维复向量),并对预测结果进行逐样本 L_2 归一化,维持信道矩阵良态性,规避数值异常问题。训练方式上,该信道预测网络不引入真实信道标签进行监督,而是与语义编解码器、展开式PGD均衡模块构成统一网络进行端到端联合训练,全网共享图像重建MSE损失函数,优化器、学习率等训练超参与主网络完全一致。

每一层迭代设置独立的可学习步长参数 $\tilde{\eta}_k$,初始值设置为0.001,同时通过绝对值约束保证步长恒为非负,使各层迭代沿残差补偿方向更新。需要说明的是,步长非负并不构成严格的收敛性理论保证,迭代稳定性还与信道矩阵的谱特性相关;在本文实验设置下,训练与推理过程未观察到迭代发散现象,3.2.3小节的消融实验表明模块性能随迭代步数平稳提升并趋于饱和。

本文信道预测模块工作在盲估计方式下,网络不依赖真实信道标签监督,完全由最终图像重建误差驱动进行端到端学习,其本质是一种数据驱动的隐式信道适配机制,而非严格意义上的信道参数估计。盲信道估计固有的尺度、相位与分量分解模糊性,在本文框架下由端到端重建目标隐式消解——网络仅需学习对重建任务有效的等效信道表征,无需唯一恢复真实信道参数。

功率约束采用与发射端一致的逐样本单位能量归一化投影方式,将迭代结果约束在功率可行域内,投影算子表达式为

$$\mathcal{P}_x(z) = \frac{z}{\sqrt{E[|z|^2] + \epsilon}} \quad (8)$$

式中: $E[|z|^2]$ 为符号序列的平均能量, $\epsilon = 10^{-8}$ 为数值稳定项,避免分母为零。该约束方式可限制迭代过程符号能量无限制增长,进一步提升网络训练与迭代推理的数值稳定性。

整体的展开式PGD迭代均衡模块由 K 级迭代子层级联构成,每一级完成一次完整PGD均衡迭代,前向计算流程如算法1所示。

算法1 展开式PGD迭代均衡流程

输入:时域带噪符号序列 $\hat{y}$,迭代步数 K

初始化:均衡初始值 $y_0 = \hat{y}$

for $k = 0, 1, \dots, K - 1$

信道前向传播: $Hy_k = \beta h_d \odot y_k + (1 - \beta) \cdot \text{IFFT}(\text{FFT}(y_k) \odot \text{FFT}(h_c))$

计算残差: $r_k = Hy_k - \hat{y}$

匹配滤波与梯度计算: $g_k = \beta h_d \odot r_k + (1 - \beta) \cdot \text{IFFT}(\text{FFT}(r_k) \odot \text{Conj}(\text{FFT}(h_c)))$

残差补偿更新: $z_k = y_k - |\tilde{\eta}_k| \cdot g_k$

投影约束: $y_{k+1} = z_k / \sqrt{E[|z_k|^2] + \epsilon}$

输出:均衡后时域符号序列 y_K

完成 K 步迭代均衡后,输出符号序列 y_K 送入后续的Swin Transformer解码网络,完成最终图像重建,流程如图3所示。

2.3.3 模块的核心优势

该模块是UP-WITT双驱动架构的核心衔接单元,实现了模型先验与数据拟合的有机结合:

(1)从物理可解释性来看,网络每一级迭代均对应线性均衡的完整物理过程,残差计算、匹配滤波、残差补偿及投影约束均具备明确通信物理含义,弥补了纯数据驱动黑箱均衡可解释性弱的缺陷,也便于针对低空信道莱斯因子、多径数量等参数进行定向优化。

(2)从收敛与鲁棒性来看,依托优化算法展开的迭代机制在当前实验条件下表现出更平滑的训练过程和更好的重建性能,一定程度上缓解了纯数据驱动模型训练梯度振荡的问题;同时嵌入的低空信道物理先验有助于缓解时变信道逆问题的病态程度,在低信噪比、LOS/NLOS动态切换等复杂工况下仍可保持良好均衡性能。

(3)从工程适配性来看,可学习步长与自适应信道预测模块赋予模型自主适配低空动态信道的能力,无需针对不同传输场景重新设计均衡算法,具备良好的场景兼容性与工程落地潜力。

3 实验结果与分析

3.1 实验设置

3.1.1 数据集

本文选用CIFAR数据集作为低分辨率实验训练集与测试集,DIV2K数据集作为高分辨率训练集,CLIC2021作为高分辨率测试集。此类数据能够匹配低空快速场景概览、态势感知等图像传输应用需求。训练阶段采用随机水平翻转、亮度微调等数据增强策略,进一步提升模型的泛化能力与场景适配性。

3.1.2 对比方案

选取语义通信领域中基于视觉Transformer的代表性WITT模型作为基线对比方案。WITT采用纯数据驱动端到端架构,在编码输出端与解码输入端分别增设信道调制网络,以实时信噪比作为网络输入生成自适应缩放系数,通过特征逐元素加权实现信道环境适配。本文所提UP-WITT方法舍弃了WITT额外的信道调制网络,依托模型驱动的展开式PGD迭代均衡模块实现信道损伤自适应补偿。

为保证对比实验的公平性,两种模型采用完全一致的Swin Transformer编解码骨干、训练数据集及超参配置,仅在信道适配与均衡机制上存在差异,可有效验证本文模型驱动可学习均衡方法相对纯数据驱动信道调制方案的性能提升效果。

3.1.3 评价指标

本文采用图像通信领域两项通用客观评价指标定量衡量模型的图像重建性能。峰值信噪比(Peak signal-to-noise ratio, PSNR)以像素均方误差为评估依据,单位为dB,指标数值越高,代表重建图像与原始图像的像素偏差越小,还原精度越高;多尺度结构相似性(MS-SSIM)贴合人眼视觉特性设计,指标取值区间为0~1,数值越趋近于1,表明图像纹理、轮廓等结构信息保留越完整,视觉观感更优。

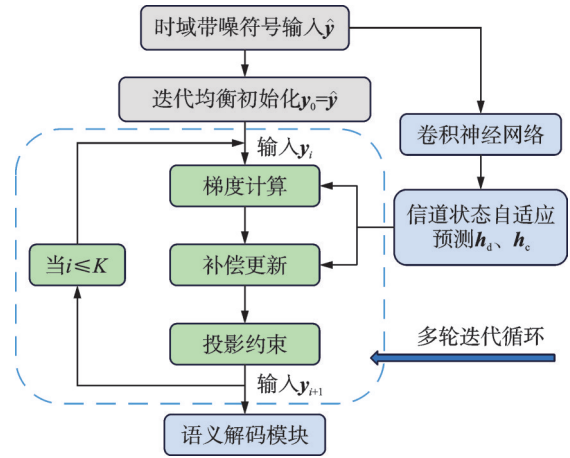


图3 模型驱动的展开式PGD迭代均衡模块

Fig.3 Model-driven unrolled PGD iterative equalization module

3.1.4 训练与场景设置

训练过程采用本文构建的低空时域基带信道模型,信道信噪比在1~13 dB区间内均匀随机采样,使训练所得模型可适配该信噪比范围内的低空动态信道变化规律。具体参数如表1所示。

表1 低空时域基带信道模型参数

Table 1 Parameters of the low-altitude time-domain baseband channel model

参数类别	参数名称	符号	取值与分布说明
信道结构	混合权重	β	[0,1]区间均匀分布
	多径信道参数	h_e	$\mathcal{CN}(0,0.25)$
	平坦衰落信道	h_d	Rician分布
Rician分布参数	Rician K因子	K	[4,8]区间均匀分布
	幅度扰动	ϵ	$\mathcal{N}(0,0.02^2)$
噪声模型	信噪比	SNR	随机选取{1,4,7,10,13}

网络训练选用SGD优化器,初始学习率设为 1×10^{-4} ,批次大小128,完整训练迭代轮次250轮。展开式PGD均衡模块固定迭代步数为5,可学习迭代步长初始值设置为0.001,训练损失函数采用均方误差(Mean squared error, MSE)。所有模型均在单张NVIDIA GeForce RTX 4070 Ti SUPER显卡上完成训练与仿真测试。

3.2 实验结果与分析

3.2.1 CIFAR数据集实验

(1) 固定CBR下不同SNR性能对比

为对比不同信道质量下的图像重建效果,分别设置信道带宽比 $CBR=0.5$ 与0.25两种压缩工况,绘制不同信噪比条件下UP-WITT与WITT模型的PSNR、MS-SSIM对比曲线,直观分析信道波动对两类模型重建性能的影响。

$CBR=0.5$ 工况下各信噪比对应的模型性能曲线如图4、5所示。从图4、5可以看出,随着信道信噪比提升,两种模型的PSNR与MS-SSIM指标均呈稳步上升趋势。在全信噪比范围内,UP-WITT各项指标均优于WITT基线模型;尤其在低信噪比恶劣传输条件下,本文方法仍能保持更优重建效果,说明展开式PGD均衡模块可有效抑制信道噪声干扰,弥补纯数据驱动模型抗噪能力不足的短板。

图6、7给出 $CBR=0.25$ 带宽压缩条件下的性能对比。带宽压缩比例增大后,图像传输可用资源减

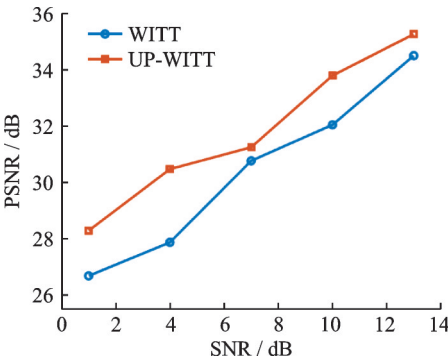


图4 CBR=0.5时PSNR对比曲线

Fig.4 PSNR performance versus SNR at CBR=0.5

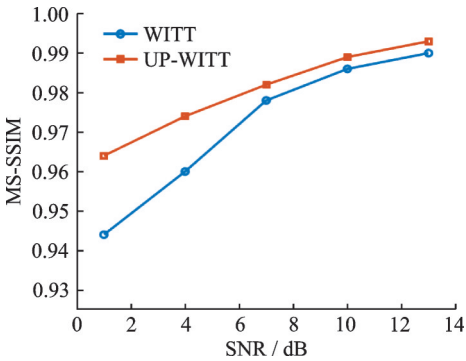


图5 CBR=0.5时MS-SSIM对比曲线

Fig.5 MS-SSIM performance versus SNR at CBR=0.5

少,两类模型整体重建性能均有所回落。即便在传输资源受限的条件下,UP-WITT 仍保持稳定性能优势。相比于依靠特征自适应调制的 WITT 模型,本文引入的物理迭代均衡机制对带宽约束适配性更强,能够在低带宽条件下有效保留图像基础语义与细节特征。

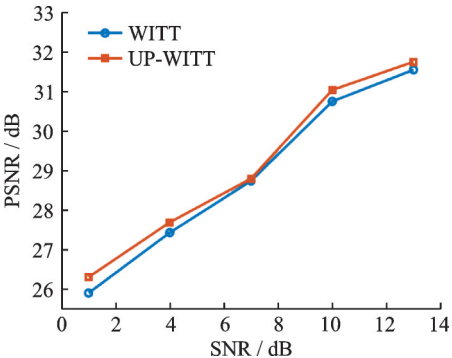


图6 CBR=0.25时 PSNR 对比曲线

Fig.6 PSNR performance versus SNR at CBR=0.25

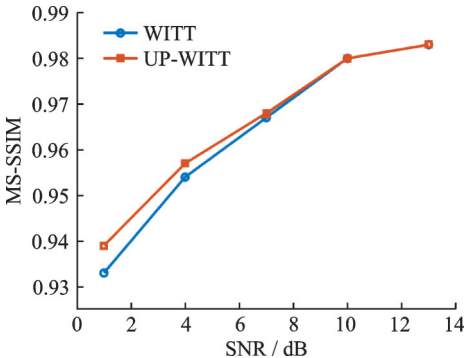


图7 CBR=0.25时 MS-SSIM 对比曲线

Fig.7 MS-SSIM performance versus SNR at CBR=0.25

(2) 不同 CBR 条件下整体性能横向对比

为量化分析带宽压缩程度对模型重建效果的影响,统计两种 CBR 工况下全信噪比区间的平均 PSNR 与平均 MS-SSIM,结果如表 2 所示。由表 2 数据可见,在两种带宽约束条件下,UP-WITT 的平均 PSNR 与 MS-SSIM 均高于基线模型。随着带宽压缩比例提升,两类模型性能均出现一定衰减,但本文方法始终保持性能领先,体现出更优的带宽自适应适配能力,在不同传输资源条件下均可稳定提升图像重建质量。

3.2.2 DIV2K/CLIC2021 数据集实验

为验证所提方法在高分辨率图像传输任务中的可扩展性,本文增设高分辨率场景扩展实验。实验采用 DIV2K 训练集(800 张高分辨率图像,训练时随机裁剪为 256 像素 $\times$ 256 像素图像块)训练模型,在 CLIC2021 测试集上评估性能。网络采用 4 级 Swin Transformer 编解码器(各级 Block 数为[1, 1, 2, 6],特征维度为[128, 192, 256, 320],窗口尺寸为 8),与 WITT 高分辨率配置一致;展开式 PGD 均衡模块迭代步数为 5,可学习步长初始值为 0.001。信道模型采用本文第 1.1 节低空时域基带信道模型,参数同表 1,信道带宽比 CBR=1/16;训练信噪比在 1~13 dB 区间均匀采样,测试信噪比取{1, 4, 7, 10, 13}dB。对比方法为 CNN 骨干的 DeepJSCC 与纯数据驱动的 WITT;批次大小设为 16,其余训练超参与 3.1 小节一致。

(1) 固定 CBR 下不同 SNR 性能对比

由表 3 可见,在 CBR=1/16 的高分辨率传输工况下,3 种方法的 PSNR 与 MS-SSIM 指标均随信道信噪比提升而稳步上升。横向对比可以发现:(1)CNN 骨干的 DeepJSCC 受局部感受野限制,在高分辨率图像上与两种 Swin Transformer 方法的性能差距明显拉大,这与 WITT^[10]文中观察到的“高分辨率下 CNN 方法性能退化”现象一致,印证了本文选用 Transformer 骨干的必要性;(2)UP-WITT 在全信噪比范围内均保持领先,且在低信噪比恶劣条件下增益更为显著,表明展开式 PGD 均衡模块引入的物理先

表 2 不同带宽比下模型整体平均性能统计

Table 2 Overall average performance of the models under different bandwidth ratios

CBR	模型	平均	平均
		PSNR/dB	MS-SSIM
0.5	WITT	30.370	0.972
0.5	UP-WITT	31.807	0.980
0.25	WITT	28.876	0.963
0.25	UP-WITT	29.119	0.965

验在高分辨率、强噪声场景下依然具备良好的信道损伤抑制能力。

表 3 CLIC2021数据集上不同 SNR 下各方法性能对比 (CBR=1/16)

Table 3 Performance comparison of different methods at different SNRs on the CLIC2021 dataset (CBR=1/16)

SNR/dB	PSNR/dB			MS-SSIM		
	DeepJSCC	WITT	UP-WITT	DeepJSCC	WITT	UP-WITT
1	20.2	22.6	23.9	0.895	0.922	0.933
4	22.7	25.1	26.4	0.918	0.941	0.950
7	24.9	27.6	28.7	0.936	0.955	0.962
10	26.7	29.6	30.6	0.950	0.966	0.971
13	28.3	31.6	32.5	0.961	0.973	0.978

(2) 不同 CBR 条件下整体性能横向对比

由表 4 可见,在 CLIC2021 高分辨率测试集的全信噪比区间上,UP-WITT 平均 PSNR 达 28.420 dB,较 WITT 提升约 1.1 dB,较 DeepJSCC 提升约 3.9 dB;平均 MS-SSIM 达 0.959,同为 3 种方法最优。上述结果表明,本文方法可从 32×32 小尺寸图像稳定扩展至 256 像素 $\times$ 256 像素分辨率图像传输任务,在不同信道带宽与信噪比条件下均保持了模型驱动均衡机制带来的性能增益,具备良好的可扩展性,能够匹配低空无人机巡检、安防监控等场景的高清图像回传需求。

3.2.3 消融实验

为分析展开式 PGD 模块迭代步数对重建性能的影响,在 CBR=0.5、SNR=4 dB 条件下开展消融实验,测试结果如表 5 所示。实验结果表明,随着 PGD 迭代步数增加,模型图像重建性能逐步提升。迭代步数由 0 增至 5 时,PSNR 提升 2.596 dB,MS-SSIM 提升 0.014,性能增益最为明显;当迭代步数超过 5 步后,指标增长趋于饱和,继续增加迭代次数仅能获得小幅性能提升,同时会带来计算开销增大的问题。综合考虑图像重建精度与低空无人机平台轻量化运算需求,本文选取 5 步迭代作为模块最优配置。

3.2.4 训练稳定性分析

为保证训练过程对比的公平性,所有模型设置相同随机种子、批次采样规则与数据加载方式。以 CBR=0.5、全区间信噪比训练过程为例,绘制两类模型训练损失收敛曲线进行对比,结果如图 8 所示。对比收敛曲线可以看出,纯数据驱动的 WITT 模型训练损失存在明显振荡现象,收敛过程不够平稳;而 UP-WITT 依托展开式 PGD 模块引入的物理先验与迭代约束,训练损失曲线变化更为平滑,有效改善了网络训练的收敛稳定性。本文采用的低空参数化信道模型涵盖了 LOS/NLOS 动态切换的典型工况。后续可结合更宽参数范围的信道测试与实测空地信道数据,进一步拓展验证框架的适用边界。

表 4 CLIC2021数据集上全 SNR 区间平均性能统计

Table 4 Average performance over the full SNR range on the CLIC2021 dataset

模型	平均	平均
	PSNR/dB	MS-SSIM
DeepJSCC	24.560	0.932
WITT	27.300	0.951
UP-WITT	28.420	0.959

表 5 迭代步数对 UP-WITT 性能的影响

Table 5 Effect of the number of iterations on the performance of UP-WITT

迭代步数	PSNR/dB	MS-SSIM
0	27.872	0.960
1	28.521	0.965
3	29.876	0.971
5	30.468	0.974
7	30.592	0.975
10	30.645	0.975

4 结束语

针对低空空地通信中纯数据驱动语义通信方法普遍存在物理先验缺失、信道均衡鲁棒性偏弱、训练收敛易振荡等问题,本文提出一种模型-数据双驱动的端到端图像语义通信方法。该方法以低空莱斯时变时域基带信道模型为基础,通过双分支协同架构,兼顾深度学习对复杂语义特征的拟合能力,以及模型驱动均衡算法的可解释性与收敛稳定性。并且以 Swin Transformer 搭建数据驱动编解码骨干网络,完成图像多尺度语义特征提取,并映射为满足传输规范的时域离散基带符号;同时将线性均衡凸优化求解过程展开为可学习网络层,构建 PGD 迭代均衡模块,依托信道先验信息引导网络完成时变信道自适应补偿,在当前实验条件下缓解了时变信道逆问题求解的病态程度。基于 CIFAR 数据集的仿真结果表明,在不同信道带宽比与莱斯衰落条件下,本文方法的 PSNR 与 MS-SSIM 指标均优于纯数据驱动基线模型,在训练信道分布范围内表现出更平稳的训练收敛特性与更好的信道自适应能力。

后续研究可面向低空高速移动场景,进一步优化信道预测模块的时变信道跟踪性能;同时结合联邦学习与模型轻量化手段,探索算法在无人机集群下的分布式训练与边缘部署方案,提升方法在实际低空通信场景的工程应用价值。此外,本文信道预测模块采用盲估计方式,其与导频辅助信道估计方案的性能边界、以及信道估计误差对重建性能的影响机理是有待深入研究的理论问题,将作为后续工作的重点方向之一。

参考文献:

- [1] 吴启晖,董超,贾子晔,等.低空智能组网与控制理论方法[J].航空学报,2024,45(3):28809.
WU Qihui, DONG Chao, JIA Ziye, et al. Networking and control mechanism for low-altitude intelligent networks[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(3): 28809.
- [2] 丁旭辉,卢琦,张园园,等.面向低空经济的通信技术综述[J].南京航空航天大学学报(自然科学版),2025,57(6):1026-1045.
DING Xuhui, LU Qi, ZHANG Yuanyuan, et al. Review on communication technologies for low-altitude economy[J]. Journal of Nanjing University of Aeronautics & Astronautics (Natural Science Edition), 2025, 57(6): 1026-1045.
- [3] 董超,经宇骞,屈毓铤,等.面向低空智能网频谱认知与决策的云边端融合体系架构[J].通信学报,2023,44(11):1-12.
DONG Chao, JING Yuqian, QU Yuben, et al. Cloud-edge-device fusion architecture oriented to spectrum cognition and decision in low altitude intelligence network[J]. Journal on Communications, 2023, 44(11): 1-12.
- [4] 董超,崔灿,贾子晔,等.面向低空智能网的多维信息统一表征技术综述[J].电子与信息学报,2025,47(5):1215-1229.
DONG Chao, CUI Can, JIA Ziye, et al. Survey of unified representation technology of multi-dimensional information for low altitude intelligent network[J]. Journal of Electronics & Information Technology, 2025, 47(5): 1215-1229.
- [5] GÜNDÜZ D, WIGGER M A, TUNG T Y, et al. Joint source-channel coding: Fundamentals and recent progress in practical designs[J]. Proceedings of the IEEE, 2025, 113(9): 888-919.
- [6] FARSA D N, RAO M, GOLDSMITH A. Deep learning for joint source-channel coding of text[C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2018: 2326-2330.
- [7] BOURTSOULATZE E, KURKA D B, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 567-579.
- [8] KURKA D B, GÜNDÜZ D. Deep joint source-channel coding of images with feedback[C]//Proceedings of 2020 IEEE

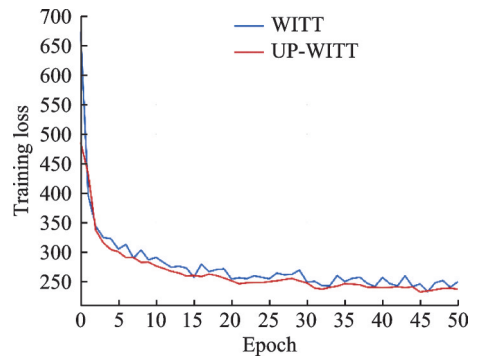


图8 CBR=0.5时模型训练损失对比

Fig.8 Comparison of training loss at CBR=0.5

- International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2020: 5235-5239.
- [9] XU J, 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, 2021, 32(4): 2315-2328.
- [10] YANG K, WANG S, DAI J, et al. WITT: A wireless image transmission transformer for semantic communications[C]//Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2023: 1-5.
- [11] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 10012-10022.
- [12] 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.
- [13] WANG L, WU W, ZHOU F, et al. A unified hierarchical semantic knowledge base for multi-task semantic communication [C]//Proceedings of 2024 IEEE International Conference on Communications. [S.l.]: IEEE, 2024: 2937-2943.
- [14] GREGOR K, LECUN Y. Learning fast approximations of sparse coding[C]//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: [s.n.], 2010: 399-406.
- [15] HERSHEY J R, ROUX J L, WENINGER F. Deep unfolding: Model-based inspiration of novel deep architectures[J]. arXiv preprint arXiv: 1409.2574, 2014.
- [16] ZHAO H, YANG C, XU Y, et al. Model-driven based deep unfolding equalizer for underwater acoustic OFDM communications[J]. IEEE Transactions on Vehicular Technology, 2022, 72(5): 6056-6067.
- [17] SAHIN S, CIPRIANO A, POULLIAT C. Optimization and analysis of deep unfolding based double loop turbo equalizers [C]//Proceedings of URSI-France Scientific Days Workshop: Future Networks: 5G and Beyond. [S.l.]: [s.n.], 2020: 1-7.
- [18] IGARASHI K. Parameter optimization of MIMO adaptive equalization using deep unfolding[C]//Proceedings of 2024 IEEE Opto-Electronics and Communications Conference (OECC). [S.l.]: IEEE, 2024: 1-3.

作者简介:



刘任(1993-),通信作者,男,博士,研究方向:深度学习、语义通信、大模型,以及低空经济、具身智能等人工智能相关应用, E-mail:18625172750@163.com。



辛利(1984-),男,硕士,研究方向:低空经济、6G通信,低空智联网与eVTOL通信导航技术。



王帅军(1992-),男,博士,研究方向:6G通信、空天地海一体化网络。

(编辑:王静)

Model-Data Dual-Driven Image Semantic Communication for Low-Altitude Scenarios

LIU Ren^{1,2*}, XIN Li^{2,3}, WANG Shuaijun²

(1. School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510000, China; 2. Research Institute, Guangzhou Haige Communications Group Incorporated Company, Guangzhou 510000, China; 3. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Low-altitude air-ground communication channels are characterized by time-varying multipath propagation and dynamic switching of line-of-sight (LOS) and non-line-of-sight (NLOS) components. Traditional pure data-driven joint source-channel coding methods lack physical communication priors, leading to poor image reconstruction robustness and oscillating training convergence. This paper proposes an end-to-end image semantic communication architecture (UP-WITT) with model-data dual-driven design for low-altitude scenarios. Taking the time-domain baseband transmission model of low-altitude time-varying channels as the physical basis, the architecture adopts a two-branch collaborative structure. The data-driven branch employs Swin Transformer as the codec backbone to extract multi-scale global semantic features, which are further mapped into time-domain discrete baseband symbol sequences complying with transmission constraints. Based on projected gradient descent (PGD) theory, the model-driven branch unrolls the iterative optimization process into learnable network layers and constructs a configurable PGD reconstruction module. With the physical prior of time-domain channel transmission, the module performs effective compensation of channel impairments and reduces iterative complexity via frequency-domain fast computation, thereby alleviating the ill-posedness of the inverse problem under the experimental conditions considered. Experimental results on the CIFAR dataset show that compared with pure data-driven baseline models, the proposed method achieves better peak signal-to-noise ratio (PSNR) and multi-scale structural similarity (MS-SSIM) under different channel bandwidth ratios (CBRs) and channel conditions. It also exhibits a smoother training convergence process and better reconstruction performance within the trained channel distribution, providing a useful reference for reliable image semantic transmission in low-altitude applications.

Highlights:

1. Propose a model-data dual-driven end-to-end image semantic communication architecture (UP-WITT) for low-altitude scenarios, which adopts a swin transformer codec backbone to extract multi-scale global semantic features and map them into time-domain baseband symbol sequences.
2. Establish a low-altitude time-domain baseband channel model that captures the dynamic switching between Rician line-of-sight flat fading and circulant multi-path non-line-of-sight components, precisely characterizing the time-varying air-ground transmission channel.
3. Unroll the projected gradient descent (PGD) iteration into learnable network layers to build a channel-adaptive equalization module; embedding physical channel priors and FFT-based fast computation alleviate the ill-posedness of the inverse problem, improving PSNR/MS-SSIM and training convergence over pure data-driven baselines.

Key words: communication and information system; low-altitude air-ground communication; image semantic communication; joint source-channel coding; model-data dual-driven

Received: 2026-06-08; Revised: 2026-09-07

*Corresponding author, E-mail: 18625172750@163.com.