

基于深度学习的 CSI 压缩反馈方案研究

王兆圆, 李立华

(北京邮电大学信息与通信工程学院, 北京 100876)

摘要: 针对基于深度学习的信道状态信息 (channel status information, CSI) 压缩反馈技术, 本文提出了一种基于多层感知机 (multi-layer perceptron, MLP)-Transformer 深度自编码器的 CSI 压缩反馈方案。该方案从现有算法的不足出发, 本着对编码器模块的轻量级设计准则, 以及 Transformer 自注意力机制的应用, 设计了一种高精度 CSI 压缩重构的网络模型。仿真结果表明, 基于 MLP-Transformer 的深度自编码器 CSI 压缩反馈方案在频分双工 (frequency division duplex, FDD) 下行大规模多输入多输出 (multiple input multiple output, MIMO) 系统场景的不同压缩比下, 均有很高的 CSI 重构精度, 且其编码器模块的轻量级设计对比其他方案具有更低的编码器复杂度。

关键词: 通信技术; 大规模多输入多输出 (MIMO); 深度学习; CSI 压缩反馈

中图分类号: TN929.52

文献标识码: A

文章编号: 1674-2850(2023)02-0241-09

开放科学 (资源服务) 标识码 (OSID):

DOI: 10.3969/j.issn.1674-2850.2023.02.012



Research of CSI compression feedback based on deep learning

WANG Zhaoyuan, LI Lihua

(School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: For compression feedback technology of channel status information (CSI) based on deep learning, this paper proposed a CSI compression feedback based on multi-layer perceptron (MLP)-Transformer deep autoencoder. Starting from the shortcomings of existing algorithms, this scheme realized a network model of high-precision CSI compression and reconstruction recovery, based on the lightweight design criteria of encoder module and the application of Transformer self-attention mechanism. The simulation results show that the CSI compression feedback scheme of deep self encoder based on MLP-Transformer has high CSI reconstruction accuracy under different compression ratios of frequency division duplex (FDD) downlink massive multiple input multiple output (MIMO) system scenarios, and that the lightweight design of its encoder module has lower encoder complexity than other schemes.

Key words: communication technology; massive multiple input multiple output (MIMO); deep learning; CSI compression feedback

0 引言

近年来, 随着第五代移动通信 (5th generation mobile communication systems, 5G) 的发展与普及, 随之而来的是指数式增长的无线设备的接入和海量无线数据的传输, 无线通信面临更大容量、更密网络、

作者简介: 王兆圆 (1997—), 男, 硕士研究生, 主要研究方向: 无线和移动通信

通信联系人: 李立华, 教授, 主要研究方向: MIMO 及多天线传输技术、协作通信技术、无线能量收集技术、链路自适应技术和多载波传输技术. E-mail: lilihua@bupt.edu.cn

更低时延的巨大挑战^[1]。与此同时,人工智能(artificial intelligence, AI)技术在自然语言处理、计算机视觉、生物医学等领域获得了空前的成功, AI 中深度学习算法的优越性自然也引起了无线通信领域研究者们关注。研究者们期望将 AI 特别是深度学习技术与无线通信的各个层面相结合, 打造面向未来通信系统的智能通信, 并将其视为后 5G 或 6G 时代的主流发展方向之一。

CSI 的压缩反馈作为无线通信物理层链路的重要一环, 是指将在接收端信道估计得到的 CSI, 经过反馈链路反馈到基站的过程。基站则可以根据反馈的 CSI 进行预编码等操作来消除干扰, 提高频谱效率。对于大规模 MIMO 系统的 CSI 压缩反馈, 大量天线的使用使得传统方案的码本设计复杂度和相应的反馈量显著增长, 因此, 改变传统基于码本的反馈方法, 用深度学习技术对完整的 CSI 进行压缩并反馈, 以实现在额定反馈开销的情况下反馈更准确的信道信息, 是未来 CSI 压缩反馈领域研究的重点方向。

对于基于深度学习的 CSI 压缩反馈技术, 国内外现有基于深度学习的 CSI 反馈方案大部分^[2~8]都是借鉴压缩感知^[9]理论中截取稀疏矩阵的思想设计的, 然而并不是所有的信道响应均在其角度-时延域拥有稀疏特性, 对于更复杂场景的 MIMO 信道, 其信道数据的预处理操作变得不再适用。其次, 虽然 2021 年至 2022 年最新提出的 3 篇^[6~8]基于 Transformer^[10]的 CSI 压缩反馈算法已呈现出不错的效果, 但是这些模型均有一定的弊端, 即它们的编码器(Encoder)和解码器(Decoder)部分均使用了 Transformer Encoder 或 Decoder 层, 这使得 Encoder 部分的网络参数量过大, 而 Encoder 往往应用于终端等小体量设备上, 所以, 轻量级的 Encoder 网络设计才是 CSI 压缩反馈模型的关键。因此, 本文针对 FDD 模式下大规模 MIMO 系统中基于深度学习的 CSI 压缩反馈技术, 提出了一种基于 MLP-Transformer 的 CSI 压缩反馈方案, 并对其在 FDD 下行大规模 MIMO 信道模型下进行仿真对比与分析。

1 系统模型

对于大规模 MIMO 的下行链路, 在时分双工(time division duplex, TDD)制式下, 基站可通过上行链路发送的导频, 在基站端进行信道估计, 再利用信道互易性获取下行链路的 CSI; 而在 FDD 制式下, 上行链路和下行链路工作在不同的频点上, 上下行信道的互易性很弱, 因此, FDD 下行链路的 CSI 需要先由终端通过下行导频估计获得, 再通过反馈链路传回基站端。

针对 FDD 单小区下行大规模 MIMO 场景, 基于深度学习的 CSI 压缩反馈是指在接收端信道估计之后, 将估计信道 $\hat{H}$ 通过压缩神经网络进行压缩形成压缩码字, 再将压缩码字量化为反馈比特反馈给基站, 基站将接收到的反馈比特进行解量化处理后通过解压神经网络, 最终在基站端恢复出信道响应。本文考虑一个 FDD 模式下的单小区下行大规模 MIMO 系统, 其基于深度学习的 CSI 压缩反馈的系统模型如图 1 所示。

图 1 中的编码器(Encoder)和解码器(Decoder)即为用于 CSI 压缩和恢复的神经网络。基于深度学习的大规模 MIMO CSI 压缩反馈算法是由文献[2]首先正式提出的, 该算法基于自编码器(Auto-encoder)架构, 即分为编码器(Encoder)和解码器(Decoder)两部分的网络: 用编码器进行 CSI 的压缩, 用解码器重构出信道信息。简单的自编码器即为 MLP 构成的先下采样再上采样的结构。在 FDD 单小区的下行大规模 MIMO 系统中, 设终端(接收端)信道估计所得的信道响应为 $\hat{H} \in \mathbb{C}^{M \times N \times K}$, 其中: M 为基站天线数; N 为每用户终端接收天线数; K 为频带子载波数。那么基于自编码器的 CSI 压缩反馈可由图 2 所示。

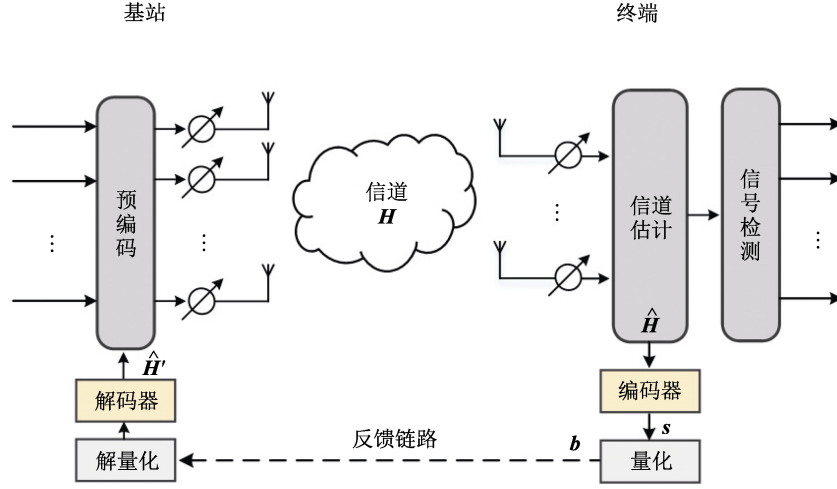


图 1 单小区下行大规模 MIMO 基于深度学习的 CSI 压缩反馈系统模型

Fig. 1 Deep learning-based CSI compression feedback system model for downlink massive MIMO in single-cell scenario

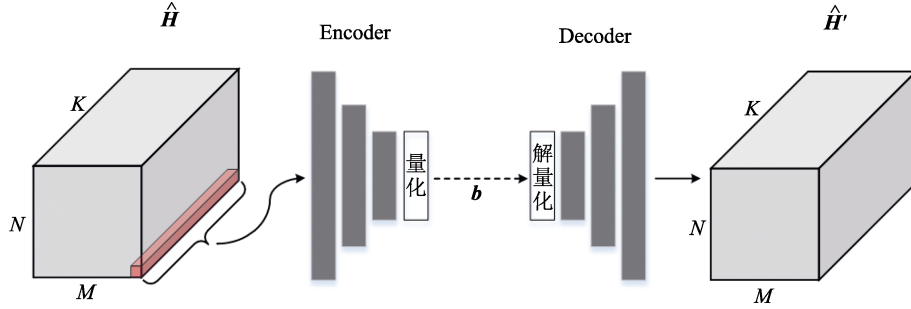


图 2 基于自编码器的 CSI 压缩反馈模型

Fig. 2 Autoencoder-based CSI compression feedback model

在图 2 中，待压缩的信道响应 $\hat{H}$ 作为 Encoder 的输入，Encoder 由下采样结构的多层 MLP 层组成，每层 MLP 层的输出节点数逐层递减，目的是逐步提取并缩小 $\hat{H}$ 的特征，最终将实数化信道矩阵 $\hat{H} \in \mathbb{R}^{M \times N \times K}$ 压缩成一定长度的码字向量 $s \in \mathbb{R}^{S \times 1}$ ，再将码字向量量化成反馈比特流 b 进行反馈。类似地，Decoder 是由上采样结构的多层 MLP 层组成，每层 MLP 层的输出节点数逐层递增以恢复出压缩前的信道响应。在基站端，对接收到的反馈比特流 $\hat{b}$ 进行解量化得到 $\hat{s} \in \mathbb{R}^{S \times 1}$ ，并将 $\hat{s}$ 作为 Decoder 的输入，逐级上采样最终重构出信道响应 $\hat{H}'$ 。对于以上过程，Encoder 部分可以表示为

$$\begin{aligned} s &= f_{\text{en}}(\hat{H}), \\ b &= Q(s), \end{aligned} \quad (1)$$

Decoder 部分可以表示为

$$\begin{aligned} \hat{s} &= \text{DQ}(\hat{b}), \\ \hat{H}' &= f_{\text{de}}(\hat{s}), \end{aligned} \quad (2)$$

其中： $f_{\text{en}}(\cdot)$ 和 $f_{\text{de}}(\cdot)$ 分别表示 Encoder 网络和 Decoder 网络； $Q(\cdot)$ 和 $\text{DQ}(\cdot)$ 分别表示量化规则和解量化规则。对于下行的 MIMO 系统，Encoder 一般存在于小型的终端设备，而 Decoder 一般布置于基站端。所以 Encoder 常使用网络层数较少、复杂度低的轻量级网络，而 Decoder 则因为有较为充足的计算资源，可以使用更深层次的网络模型，以提升 CSI 重构的精度。

2 方案及网络架构

自编码器中的 Encoder 和 Decoder 架构灵活多变, 因此衍生出了各式各样的网络模型用于 CSI 压缩反馈。考虑到注意力机制在网络模型应用中的优势, 以及 Transformer 架构与自编码器架构的相似性, 同时针对现有的基于 Transformer 的 CSI 压缩反馈算法 Encoder 复杂度过高的缺点, 本文提出了基于 MLP-Transformer 的 CSI 压缩反馈算法。MLP-Transformer 的 Encoder 只采用下采样的 MLP 层作为压缩神经网络, 将网络的重心放在 Decoder 部分, 以改进自 Transformer Decoder 的 Transformer Block 层作为特征提取的关键层。网络模型架构如图 3 所示。

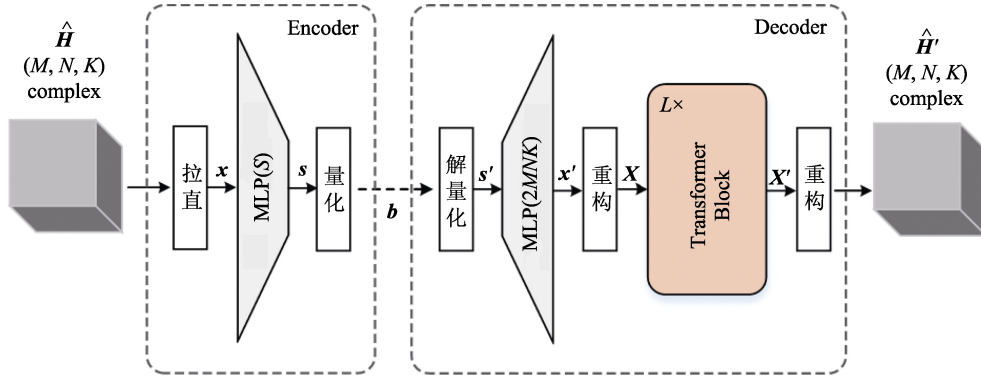


图 3 基于 MLP-Transformer 的 CSI 压缩反馈网络架构

Fig. 3 CSI compression feedback network architecture based on MLP-Transformer

如图 3 所示, 网络的输入为在 FDD 单小区下行大规模 MIMO 系统中, 终端 (接收端) 信道估计所得的信道响应 $\hat{H} \in \mathbb{C}^{M \times N \times K}$ 。对于 MLP-Transformer 的 Encoder 部分, 首先将复数 (complex) 矩阵 $\hat{H}$ 拉直成实数向量 $\mathbf{x} \in \mathbb{R}^{1 \times 2MNK}$, 再将向量 $\mathbf{x}$ 作为下采样 MLP 层的输入, 下采样 MLP 层的输出节点数为 S , 由此得到压缩码字 $\mathbf{s} \in \mathbb{R}^{1 \times S}$, 最后将压缩码字 $\mathbf{s}$ 量化成反馈比特 $\mathbf{b}$ 进行反馈传输。对于 MLP-Transformer 的 Decoder 部分, 将解量化的输出 $\mathbf{s}' \in \mathbb{R}^{1 \times S}$ 作为上采样 MLP 层的输入, 上采样 MLP 层的输出节点数为 $2MNK$, 得到向量 $\mathbf{x}' \in \mathbb{R}^{1 \times 2MNK}$, 之后对 $\mathbf{x}'$ 进行重构, 重构成适合 Transformer Block 输入的二维矩阵 $\mathbf{X} \in \mathbb{R}^{MN \times 2K}$, 将二维矩阵 $\mathbf{X}$ 通过 L 个串联的 Transformer Block 进行深度特征学习, 最终将输出的 $\mathbf{X}' \in \mathbb{R}^{MN \times 2K}$ 重构成复数信道估计 $\hat{H}' \in \mathbb{C}^{M \times N \times K}$ 完成 CSI 的重建。其中 Transformer Block 的内部结构如图 4 所示。

图 4 的 Transformer Block 可被视作 Transformer Decoder 的改进。Transformer Block 包含两个多头注意力模块 (Multi-Head Attention), 以及一个输出线性层 (Linear), 多头注意力模块之后用层归一化 (Layer Norm) 进行归一化, 整个 Block 用 3 处残差连接拉通输入和输出, 以防止因网络层数加深而可能出现的梯度消失的状况。对于多头注意力模块, 其将输入数据复制成相同的 3 份, 分别作为问 (Query)、键 (Key) 和值 (Value) 送入 n 个并行头的自注意力模块中, 每个头的输入都是相同的问、键和值, 分别经过各自的线性层后进行自注意力机制 (Scaled Dot-Product Attention) 的计算, 自注意力机制可由下式表示:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax} \left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d_k}} \right) \times \mathbf{V}, \quad (3)$$

其中: d_k 是键的维度; $\mathbf{Q}$ 、 $\mathbf{K}$ 和 $\mathbf{V}$ 分别是问、键和值向量经过各自线性层之后的输出; $(\cdot)^T$ 为转置运算由式 (3) 可以看出, 自注意力机制即是 $\mathbf{Q}$ 和 $\mathbf{K}$ 之间的互相关, 归一化后由 Softmax 函数转化为互相关概

率系数，并作为注意力权重与 V 相乘。由于输入的问、键和值是相等的，相当于对自身引入注意力计算自相关系数，所以这里称作自注意力机制。多头注意力即是把这样的自注意力并行做 n 次，再将 n 次的结果级联 (Concat) 在一起，最后通过一个下采样的 MLP 层 (Down MLP) 将数据恢复成级联前的维度。图 4 中的所有线性层 (Linear) 都为输入、输出节点数相同的全连接层。

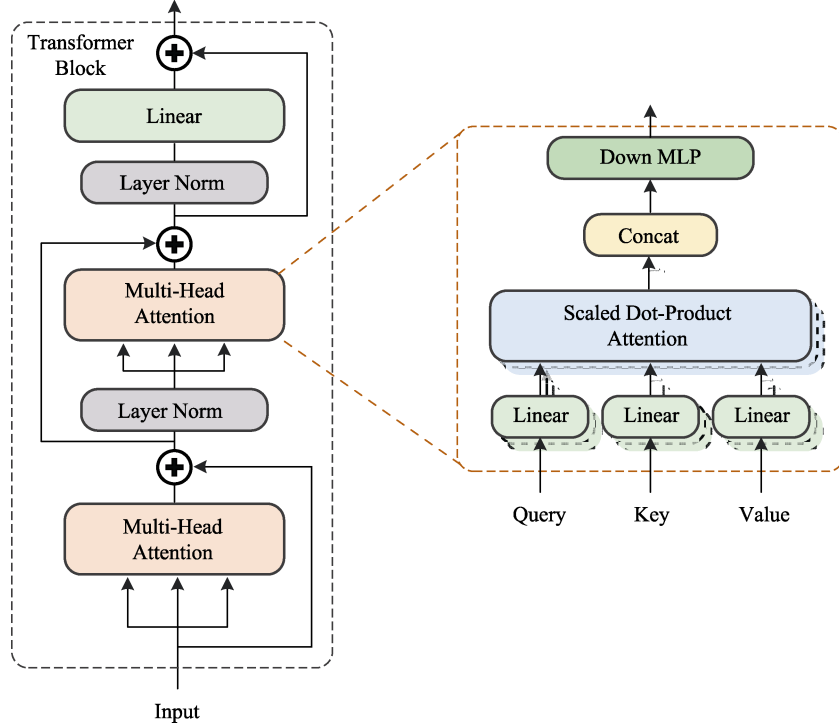


图 4 Transformer Block 架构

Fig. 4 Transformer Block architecture

Transformer Block 与原始 Transformer Decoder 的区别在于，Transformer Block 的第一个多头注意力模块并没有加入掩码，原因是原始的 Transformer 是用于自然语言处理领域的，掩码的加入是为了增强对文本序列的预测，考虑到信道数据进行二维实数化变换后与文本序列数据呈现不一样的自相关特性，因此掩码的引入在 CSI 压缩反馈中没有理论支持，也不一定会提升网络的精度，这一点也在后面的仿真中得以证实。

通过以上的描述，图 3 基于 MLP-Transformer 的 CSI 压缩反馈网络模型，其 Encoder 部分可以表示为

$$\begin{aligned} \mathbf{x} &= \text{reshape}_1(\hat{\mathbf{H}}); \\ \mathbf{s} &= f_{\text{MLP-Down}}(\mathbf{x}); \\ \mathbf{b} &= Q(\mathbf{s}). \end{aligned} \quad (4)$$

Decoder 部分可以表示为

$$\begin{aligned} \mathbf{s}' &= DQ(\mathbf{b}'); \\ \mathbf{x}' &= f_{\text{MLP-Up}}(\mathbf{s}'); \\ \mathbf{X} &= \text{reshape}_2(\mathbf{x}'); \\ \mathbf{X}' &= f_{\text{TransBlock}}(\mathbf{X}); \\ \hat{\mathbf{H}}' &= \text{reshape}_3(\mathbf{X}'). \end{aligned} \quad (5)$$

其中: $f(\cdot)$ 表示神经网络部分; $\text{reshape}_{1\sim3}(\cdot)$ 表示3次矩阵重构(包括拉直); $Q(\cdot)$ 和 $DQ(\cdot)$ 分别表示量化规则和解量化规则。联立式(4)和式(5),将量化、反馈和解量化三步共同视作一个随机误差 ε 的引入,即 $s' = s + \varepsilon$,则整个 MLP-Transformer 网络可表示为

$$\hat{H}' = r_3 \left\{ f_{\text{TransBlock}} \left(r_2 \left[f_{\text{MLP-Up}} \left(f_{\text{MLP-Down}} \left(r_1 \left(\hat{H} \right) \right) + \varepsilon \right) \right] \right) \right\}, \quad (6)$$

其中, $r_{1\sim3}(\cdot)$ 是 $\text{reshape}_{1\sim3}(\cdot)$ 的简写。由式(6)可以看出,虽然整个模型分为 Encoder 和 Decoder 两部分,但其正向传播、反向传播和梯度下降参数更新时均视作一个整体进行,网络性能的好坏由整体网络构造决定,不会单独因为 Encoder 或 Decoder 进行变化。因此,这给本节 MLP-Transformer 网络中 Encoder 的轻量级设计提供了支撑,将网络参数重心转向 Decoder,Encoder 仅保持下采样网络即可,将参数量级最大的 Transformer Block 模块都放在 Decoder 部分。在下一节中的仿真对比也表明,将网络参数重心转向 Decoder 的设计并不会影响 CSI 反馈的精度。

3 仿真与性能分析

本节将通过仿真探究本文提出的基于 MLP-Transformer 的 CSI 压缩反馈算法的性能,并将其与其他3种基于深度学习的 CSI 压缩反馈算法,以及一个引入掩码的自身变种网络 MLP-Masked-Transformer 进行仿真对比。下面将对训练及仿真流程做详细的介绍。

神经网络训练与仿真所用的信道数据为 COST 2100^[11]信道模型,这是因为要进行对比的3种 CSI 压缩反馈网络: CsiNet^[2]、CsiFormer^[6]、TransNet^[7],均为基于 COST 2100 模型生成的信道数据。信道的具体参数如表1所示。

表1 信道具体参数表
Tab. 1 Channel parameter table

参数	值与描述
基站天线数 (M)	32
基站天线阵列	均匀线性阵列
终端天线数 (N)	4 天线接收
中心频率	5.3 GHz
系统带宽	300 MHz
频带子载波数 (K)	512
场景	场景 1: 室内蜂窝环境 (Indoor) 场景 2: 室外乡村环境 (Outdoor)
总样本数	100 000 (TTI)
信道数据维度(频域)	场景 1 信道: $100\,000 \text{ (TTI)} \times 32 \text{ (} M \text{)} \times 4 \text{ (} N \text{)} \times 512 \text{ (} K \text{)}$ 场景 2 信道: $100\,000 \text{ (TTI)} \times 32 \text{ (} M \text{)} \times 4 \text{ (} N \text{)} \times 512 \text{ (} K \text{)}$

注: TTI 为传输时间间隔。

由此,根据表1的参数设置,对两种信道场景各生成了100 000个传输时间间隔(TTI)采样的信道数据,即网络训练的样本集数量设置为100 000,并按照70%、20%和10%的比例依次将其划分为训练集、验证集和测试集,训练集中每个小批量大小(Batch Size)设置为100。

本小节用以上训练集和验证集的信道数据,在两种信道场景(Indoor 和 Outdoor)和4种压缩率 $\gamma = S/(2MNK) \in \{1/4, 1/16, 1/32, 1/64\}$ 下,分别对 CsiNet^[2]、CsiFormer^[6]、TransNet^[7]以及本文提出的 MLP-Transformer 和其引入10%掩码的变种 MLP-Masked-Transformer 进行 CSI 压缩反馈的神经网络训练。其中 MLP-Transformer 和其变种的 Transformer Block 的层数 L 为3。神经网络训练的具体参数如表2所示。

表 2 CSI 压缩反馈网络训练的具体参数

Tab. 2 Parameters for training CSI compression feedback network

参数	值	备注
训练集大小	70 000	
验证集大小	20 000	
Batch Size	100	
损失 (Loss) 函数	MSE loss	$Loss = E \ \hat{\mathbf{H}}' - \hat{\mathbf{H}}\ ^2$
循环训练次数	1 000	
优化算法	Adam 算法	
学习率	0.001	自适应下降, 最低为 10^{-7}
量化准则	4 比特均匀量化	量化前数据会通过一个 Sigmoid 函数映射到(0,1)区间, 量化为(0,1)区间的 4 比特均匀量化。

网络在经过以上指标充分训练后, 用剩下 10% 的测试集数据进行仿真测试。在接下来的仿真结果中, 本章将采用压缩前复信道矩阵 $\hat{\mathbf{H}}$ 和重构信道矩阵 $\hat{\mathbf{H}}'$ 之间的 NMSE 以及余弦相似度 ρ 作为信道重建质量的评判指标。NMSE 公式如下:

$$NMSE(dB) = 10 \lg_{10} E \left\{ \frac{\|\hat{\mathbf{H}}' - \hat{\mathbf{H}}\|_2^2}{\|\hat{\mathbf{H}}\|_2^2} \right\}. \quad (7)$$

用 $\hat{\mathbf{h}}_k$ 和 $\hat{\mathbf{h}}'_k$ 分别表示第 k 个子载波上的原始和重构信道向量, 则余弦相似度 ρ 的表达式如下:

$$\rho = E \left\{ \frac{1}{K} \sum_{k=1}^K \frac{\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}'_k}{\|\hat{\mathbf{h}}'_k\|_2 \|\hat{\mathbf{h}}_k\|_2} \right\}. \quad (8)$$

其中, $(\cdot)^H$ 为共轭转置运算。仿真测试结果如表 3 所示。

表 3 基于深度学习的 CSI 压缩反馈仿真测试结果

Tab. 3 Simulation and testing results of deep learning-based CSI compression feedback

γ	网络	室内场景 (Indoor)		室外场景 (Outdoor)	
		NMSE/dB	ρ	NMSE/dB	ρ
1/4	CsiNet	-17.36	0.990	-8.75	0.914
	CsiFormer	-20.45	0.992	-12.81	0.951
	TransNet	-29.22	0.998	-13.99	0.966
	MLP-Transformer	-28.37	0.997	-13.08	0.959
	MLP-Masked-Transformer	-26.69	0.996	-12.37	0.940
1/16	CsiNet	-8.65	0.912	-4.51	0.790
	CsiFormer	-13.50	0.964	-6.33	0.893
	TransNet	-14.98	0.977	-6.90	0.901
	MLP-Transformer	-15.01	0.978	-7.03	0.910
	MLP-Masked-Transformer	-14.62	0.971	-6.89	0.896
1/32	CsiNet	-6.24	0.891	-2.81	0.674
	CsiFormer	-9.32	0.916	-3.51	0.741
	TransNet	-9.83	0.921	-3.77	0.762
	MLP-Transformer	-9.91	0.930	-3.91	0.774
	MLP-Masked-Transformer	-9.82	0.921	-3.74	0.758
1/64	CsiNet	-5.84	0.872	-1.93	0.596
	CsiFormer	-6.85	0.895	-2.25	0.639
	TransNet	-5.77	0.851	-2.20	0.618
	MLP-Transformer	-6.99	0.906	-2.24	0.634
	MLP-Masked-Transformer	-6.84	0.895	-2.22	0.626

由表3可以看出,无论是室内还是室外场景,在所有压缩率下,基于Transformer的CSI压缩反馈方案的重构质量都优于CsiNet。在压缩率为1/4时,TransNet的重构质量最好;在压缩率为1/16和1/32时,本文提出的基于MLP-Transformer的CSI压缩反馈方案的重构质量最好,略优于其他方案;在压缩率为1/64时,室内场景下基于MLP-Transformer的CSI压缩反馈方案最佳,在室外场景下基于CsiFormer的CSI压缩反馈方案最佳。对于本文提出的方案,不带掩码的MLP-Transformer架构对比带掩码的MLP-Masked-Transformer架构性能在所有情况下都要好。

深度学习方案的复杂度由网络的参数量决定,以上方案的复杂度(网络参数量)对比如表4所示。

表4 方案复杂度(网络参数量)对比

Tab. 4 Comparison of solution complexity (network parameter count)

γ	网络	Encoder 参数量	Decoder 参数量	网络总参数量
1/4	CsiNet	10 336	18 478	28 814
	CsiFormer	1 593 748	7 904 660	9 498 408
	TransNet	3 166 304	5 269 600	8 435 904
	MLP-Transformer	10 336	7 899 232	7 909 568
	MLP-Masked-Transformer	10 336	7 899 232	7 909 568
1/16	CsiNet	174 560	182 702	357 262
	CsiFormer	1 757 972	8 068 884	9 826 856
	TransNet	3 330 528	5 433 824	8 764 352
	MLP-Transformer	174 560	8 063 456	8 238 016
	MLP-Masked-Transformer	174 560	8 063 456	8 238 016
1/32	CsiNet	182 816	190 958	373 774
	CsiFormer	1 766 192	8 076 960	9 843 152
	TransNet	3 338 784	5 442 080	8 780 864
	MLP-Transformer	182 816	8 071 712	8 254 528
	MLP-Masked-Transformer	182 816	8 071 712	8 254 528
1/64	CsiNet	184 896	193 038	377 934
	CsiFormer	1 768 272	8 079 040	9 847 312
	TransNet	3 340 864	5 444 160	8 782 944
	MLP-Transformer	184 896	8 073 792	8 256 608
	MLP-Masked-Transformer	184 896	8 073 792	8 256 608

由表4可以看出,本文提出的基于MLP-Transformer的CSI压缩反馈网络的总参数量与CsiFormer和TransNet的总参数量接近,比这两种方案略少一点,但MLP-Transformer的Encoder参数量远小于其他两种基于Transformer的方案,实现了Encoder轻量级设计的同时,也实现了CSI重构精度的提升。

总的来说,3种基于Transformer的CSI压缩反馈算法在所有压缩率下的性能相差不大,在不同的压缩率下各有优劣。本文提出的基于MLP-Transformer的CSI压缩反馈方案在压缩率为1/16和1/32时优于其他网络,证明了将网络重心转向Decoder的轻量级Encoder网络架构设计是可行的,且Transformer中自注意力机制的应用对CSI压缩重构起到了很好的效果。值得注意的是,带掩码的MLP-Masked-Transformer架构对比不带掩码的MLP-Transformer架构性能会有所下降,证明原始Transformer Decoder中的掩码多头注意力模块并不是特别适用于信道数据的结构类型,掩码的引进反而会阻碍对信道特性的学习。

4 结论

本文针对 FDD 模式下大规模 MIMO 系统中基于深度学习的 CSI 压缩反馈技术，首先介绍了现有技术的发展以及现有方案的不足，其次介绍了基于深度学习自编码器的 CSI 压缩反馈技术的系统模型，最后提出了基于 MLP-Transformer 的 CSI 压缩反馈方案，并将其与各种基于深度学习的 CSI 压缩反馈方法在 FDD 下行 MIMO 信道模型下进行仿真对比与分析。仿真结果与网络复杂度分析表明，本文提出的基于 MLP-Transformer 的 CSI 压缩反馈方案在轻量级 Encoder 的设计之下，也能在多种压缩率下实现 CSI 重构精度的提升。

[参考文献] (References)

- [1] 李坤, 张静, 李潇, 等. 人工智能辅助的信道估计最新研究进展[J]. 电信科学, 2020, 36 (10): 46-55.
LI K, ZHANG J, LI X, et al. An overview of artificial intelligence assisted channel estimation[J]. Telecommunications Science, 2020, 36(10): 46-55. (in Chinese)
- [2] WEN C K, SHIH W T, JIN S. Deep learning for massive MIMO CSI feedback[J]. IEEE Wireless Communications Letters, 2018, 7(5): 748-751.
- [3] WANG T Q, WEN C K, JIN S, et al. Deep learning-based CSI feedback approach for time-varying massive MIMO channels[J]. IEEE Wireless Communications Letters, 2019, 8(2): 416-419.
- [4] LU C, XU W, SHEN H, et al. MIMO channel information feedback using deep recurrent network[J]. IEEE Communications Letters, 2019, 23(1): 188-191.
- [5] LIU Z Y, ZHANG L, DING Z. Exploiting bi-directional channel reciprocity in deep learning for low rate massive MIMO CSI feedback[J]. IEEE Wireless Communications Letters, 2019, 8(3): 889-892.
- [6] BI X J, LI S, YU C D, et al. A novel approach using convolutional transformer for massive MIMO CSI feedback[J]. IEEE Wireless Communications Letters, 2022, 11(5): 1017-1021.
- [7] CUI Y D, GUO A H, SONG C L, et al. TransNet: full attention network for CSI feedback in FDD massive MIMO System[J]. IEEE Wireless Communications Letters, 2022, 11(5): 903-907.
- [8] XU Y, YUAN M Q, PUN M O. Transformer empowered CSI feedback for massive MIMO systems[C]//2021 30th Wireless and Optical Communications Conference (WOCC). Taipei: IEEE, 2021: 157-161.
- [9] DONOHO D L. Compressed sensing[J]. IEEE Trans Inform Theory, 2006, 52(2): 1289-1306.
- [10] VASWANI A, SHAZEER N M, PARMAR N, et al. Attention is All you Need[EB/OL]. [2021-12-30]. <https://arxiv.org/abs/1706.03762>.
- [11] LIU L F, OESTGES C, POUTANEN J, et al. The COST 2100 MIMO channel model[J]. IEEE Wireless Communications, 2012, 19(6): 92-99.