

基于 SFNet 的大规模 MIMO 系统的 CSI 反馈算法

张昀¹, 黄经纬¹, 徐孙武¹, 高贵¹, 于舒娟¹, 赵生妹²

(1. 南京邮电大学电子与光学工程学院、柔性电子(未来技术)学院, 江苏南京 210023;

2. 南京邮电大学通信与信息工程学院, 江苏南京 210003)

摘要: 在频分双工大规模多输入多输出 (MIMO) 系统中, 为解决现有的基于深度学习的信道状态信息 (CSI) 反馈方法复杂度高、反馈精度低以及未考虑量化损失的问题, 基于传统 CNN 和 Transformer 架构, 结合一种利用全局信息而设计的空间频率模块 (SFB) 以及一种融合局部和全局特征的特征多尺度自适应空间注意力门 (MASAG), 提出了用于 CSI 反馈的深度学习算法 SFNet。通过使用快速傅里叶卷积以及特征融合网络动态来激活更多的输入信息, 同时调整接受野, 以确保有选择地突出空间相关的特征, 最大限度地减少干扰, 使网络以非常低的计算复杂度实现了先进的性能。实验结果表明, 所提算法在低复杂度情况下具有较好的估计性能, 并且在不同环境下表现出较好的鲁棒性。

关键词: 深度学习; CSI 反馈; 大规模 MIMO; 信道状态信息

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025097

CSI feedback algorithm for massive MIMO systems based on SFNet

ZHANG Yun¹, HUANG Jingwei¹, XU Sunwu¹, GAO Gui¹, YU Shujuan¹, ZHAO Shengmei²

1. College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract: To address the issues of high computational complexity, low feedback accuracy, and neglect of quantization loss in existing deep learning-based channel state information (CSI) feedback methods for frequency-division duplex massive multiple-input multiple-output (MIMO) systems, the deep learning algorithm SFNet for CSI feedback was proposed. SFNet integrated a traditional convolutional neural network (CNN) and Transformer architecture, incorporating a spatial-frequency block designed to leverage global information and a multi-scale adaptive spatial attention gate for fusing local and global features. Fast Fourier convolution and a dynamic feature fusion mechanism were utilized to activate more input information, adjust the receptive field, selectively highlight spatially correlated features, suppress interference, and allow the network to achieve advanced performance with extremely low computational complexity. The experimental results show that the proposed algorithm achieves advanced estimation performance with significantly low computational complexity. Furthermore, the trained model exhibits strong robustness across various environments.

Keywords: deep learning, CSI feedback, massive MIMO, channel state information

收稿日期: 2025-03-20; 修回日期: 2025-05-19

通信作者: 黄经纬, 1022020615@njupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62375140); 江苏省研究生科研与实践创新计划基金资助项目 (No.KYCX23_0994)

Foundation Items: The National Natural Science Foundation of China (No.62375140), The Postgraduate Research & Practice Innovation Program of Jiangsu Province (No.KYCX23_0994)

0 引言

大规模多输入多输出 (MIMO, multiple input multiple output) 技术被认为是当前 5G 无线通信标准和即将到来的 6G 无线通信技术的核心技术之一。系统通过集中式^[1]或分布式^[2]的方式为基站 (BS, base station) 配备数百甚至数千个天线, 它能够提供更高的数据吞吐量, 显著减少多用户干扰, 因而能够大幅提高频谱效率。基站处的实时信道状态信息 (CSI, channel state information) 在实现大规模 MIMO 最佳性能方面发挥着重要作用。在当前的频分双工 (FDD, frequency division duplex) 系统中, 下行链路的 CSI 从用户设备 (UE, user equipment) 处获取, 并通过反馈链路返回基站。然而在大规模 MIMO 系统中, 随着天线数量不断增加, 码本变得非常大, 导致反馈开销显著增加。为了解决上述问题, 研究人员引入了 CSI 压缩方法, 希望通过在 UE 端压缩信道矩阵来减少反馈开销和功耗。传统的基于压缩感知 (CS, compressed sensing) 的方法 (如 LASSO 解算器^[3]、AMP^[4]和 TV AL3^[5]) 存在局限性。例如, 在随机投影期间忽略信道统计, 或过度迭代处理导致的时延反馈。而且, 这些传统方法依赖于信道矩阵的稀疏性, 其 CSI 恢复质量并不令人满意。

近年来, 深度学习在 CSI 压缩中的应用已经占据了主导地位。文献[6]给出的 CSINet 首次将基于卷积神经网络 (CNN, convolutional neural network) 的变分自动编码器 (VAE, variational autoencoder) 引入压缩问题中, 性能显著优于传统的基于压缩感知的方法, 但是其网络结构简单, 难以充分提取 CSI 特征。文献[7]提出的神经网络模型使用了更大的 CNN 核, 可以通过增大感受野来更好地捕捉角延迟域中的空间相关性。CRNet^[8]采用了一种多分辨率卷积核对数据进行压缩反馈, 这种方式获得的反馈性能显著优越于 CSINet。文献[9]设计了一个特定的输入层, 利用 1×1 逐点卷积来耦合同一通道系数的实部和虚部以防相位信息丢失。该方法不仅反馈性能优于现有文献的模型, 而且显著降低了复杂度。

截至目前, 大多数模型中使用了 CNN 进行 CSI 反馈。随着 Transformer^[10]架构在无线通信领域受到广泛关注, 文献[11]首次将其引用到 CSI 反馈中, 采用全注意力机制, 但其信道矩阵重构性能仅仅略优于 CSINet, 与其他网络相比并无竞争力。文

献[12]引入了一种双层 Transformer 架构, 该架构的性能大大优于上述几个模型, 但该网络的计算复杂度非常高, 在实践中无法承受。CSIFormer^[13]通过引入局部分组自注意力在一定程度上降低了计算复杂度, 但是在性能上有着太大牺牲。尽管 Transformer 的使用极大地提升了 CSI 重建性能, 但是 CSI 重建还需要底层细节, 以尽量减少重建误差。在此基础上, STNet^[14]通过结合 CNN 和 Transformer, 专注于局部信息的同时通过 CNN 的空间不变性来捕获低层次的细节。虽然 STNet 同时兼顾了 CNN 和 Transformer 的特性, 并且极大地降低了复杂度, 但是 CNN 模块仅仅使用了传统卷积, 不能最大程度上捕获信息细节, 也做不到局部和全局信息的融合, 难以做到在捕获细节特征的同时, 又能提取更广泛的全局信道信息, 故很难实现最优性能。

针对现有的深度学习模型在 CSI 反馈信道矩阵中局部和全局信息缺失以及随着网络深度的增加, 复杂度过高且容易出现性能退化的问题, 本文提出了一种基于改进的空间频率模块 (SFB, spatial frequency block)^[15]以及多尺度自适应空间注意力门 (MASAG, multi-scale adaptive spatial attention gate)^[16]的 CSI 反馈方法。本文主要贡献总结如下。

1) 创新性地构建了一种全新的 SFB, 并将其嵌入 CSI 反馈网络解码端的 CNN 路径中。在该模块的频率分支里, 借助快速傅里叶卷积 (FFC, fast Fourier convolution)^[17]来提取全局信息。而在空间分支中, 运用基于深度可分离卷积改进的残差模块增强局部特征的表达能。SFB 不仅显著降低了计算成本, 还能够提取到更为全面、详细且稳定的全局特征。

2) 区别于传统架构, 创新性地构建了全新的 MASAG, 引入 CBAM (convolutional block attention module) 注意力机制, 显著提升特征图关键信息筛选能力。随后将 MASAG 嵌入网络末端, 提出全新的 SFNet 模型。该模型可以实现 CNN 路径全局信息与 Transformer 路径局部信息的深度融合, 能精准捕捉信道增益细粒度细节, 高效提取高层次特征, 有效克服传统模型在应对信道动态变化和不确定性方面的不足, 大幅提升 CSI 反馈的精度和鲁棒性, 为领域研究开辟了新方向。

3) 在实际部署中, 量化与逆量化的损失是长期难题。基于此本文提出在 SFNet 中引入基于 μ 律

量化器的轻量化非均匀量化方法^[18]。该方法凭借自适应量化策略，在不增加模型负载的前提下，能够有效降低量化误差，显著提升信号重建精度。

1 基于SFB和MASAG的CSI反馈网络

本文考虑的是单个小区中的简单下行链路FDD大规模MIMO系统，该系统由一个基站和一个用户设备组成。该系统在基站端配置具有 N_t ($N_t \geq 1$)个发射天线的均匀线性天线阵列，在用户设备端配置单根天线，并采用正交频分复用 (OFDM, orthogonal frequency division multiplexing)，具有 N_c 个子载波。

在该系统中，第 n 个子载波的接收信号由复值表示，可以表示为

$$y_n = \tilde{\mathbf{h}}_n^H \mathbf{v}_n x_n + z_n \quad (1)$$

其中， $\tilde{\mathbf{h}}_n \in \mathbb{C}^{N_c \times 1}$ 、 $\mathbf{v}_n \in \mathbb{C}^{N_t \times 1}$ 、 $x_n \in \mathbb{C}$ 和 $z_n \in \mathbb{C}$ 分别表示频域中的信道矢量、预编码矢量、传输数据符号和第 n 个子载波的加性噪声。

在FDD系统中，用户设备应当通过反馈链路向BS返回 $\tilde{\mathbf{H}}$ ，然而，整个信道矩阵的维数是 $N_c \times N_t$ ，并且对于信道的实部和虚部，反馈元素的总数都是 $2N_c N_t$ ，这个数据量非常大，需要占据大量珍贵的带宽资源。为了减少反馈开销，需要通过执行2D离散傅里叶变换 (DFT, discrete Fourier transform) 将反馈变换到角时延域中，过程如式(2)所示。

$$\mathbf{H} = \mathbf{F}_d \tilde{\mathbf{H}} \mathbf{F}_a^H \quad (2)$$

其中， \mathbf{F}_d 和 \mathbf{F}_a 分别是 $N_c \times N_c$ 和 $N_t \times N_t$ 的离散傅里叶变换矩阵。为了进一步压缩信道矩阵，UE可以压缩为维度为 $M \times 1$ (码字) 的一维向量，其中压缩比 (CR, compression ratio) 可以定义为 $\gamma =$

$\frac{M}{2N_c N_t}$ 。编码过程描述为

$$s = f_c(\mathbf{H}) \quad (3)$$

其中， s 表示的码字被定义为维度为 $M \times 1$ 的矩阵。本文SFNet模型在编码器和解码器之间加入量化模块，其中每个码字被量化为 B 位，并传输回基站，过程可以被描述为

$$\mathbf{v}_q = \mathcal{Q}(s) = \mathcal{Q}(f_c(\mathbf{H})) \quad (4)$$

其中， $f_c(\cdot)$ 和 $\mathcal{Q}(\cdot)$ 分别表示编码过程和量化过程。

在基站接收到量化码字 \mathbf{v}_q 后，去量化模块和解码器将重构信道矩阵 $\hat{\mathbf{H}}$ 。恢复的通道矩阵可以描述为

$$\hat{\mathbf{H}} = f_d(\mathbf{v}_q) \quad (5)$$

其中， $f_d(\cdot)$ 表示解码过程。基站最终恢复的CSI矩阵可由填充零值与逆二维傅里叶变换得到，这与编码端的前端数据处理操作是相对应的。

本文提出的SFNet系统架构主要由2个部分组成，用于CSI压缩的编码器和用于CSI解压缩和重建的解码器，整体的系统架构如图1所示。SFNet主要在解码器部分进行改进，首先，在CNN路径上构建并引用了全新的SFB，通过快速傅里叶卷积以及残差模块提取更为全面且稳定的特征。其次，在网络末端构建并嵌入了MASAG，通过MASAG实现CNN路径全局信息与Transformer路径局部信息的深度融合，使模型不仅能够捕捉信道增益的细粒度细节信息，还能提取广泛的高层次特征，提升CSI反馈精度。最后，考虑到实际通信系统中量化损失的影响，本文引入了一种轻量级非均匀量化方案，通过在编码器和解码器之间加入量化模块来对SFNet进行量化操作，有效减少了量化误差，增加了信号的重建精度。

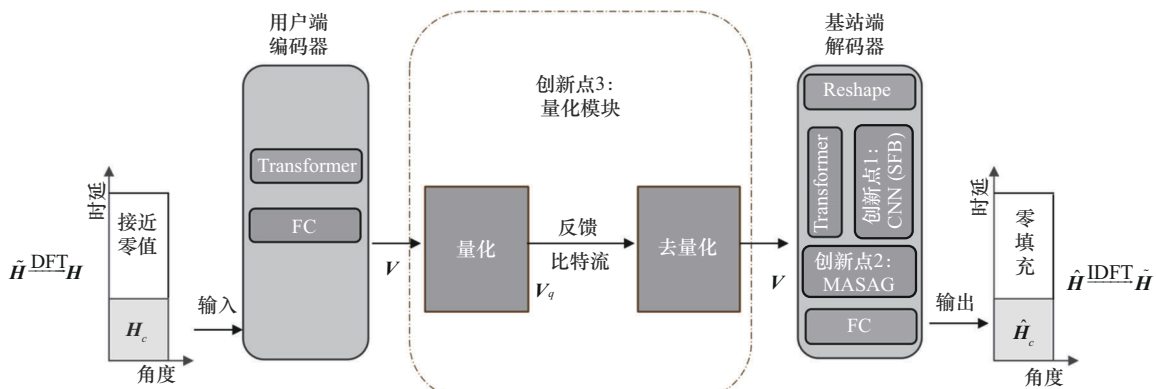


图1 SFNet系统架构

1.1 SFNet 架构

SFNet 使用的自编码结构由一个编码器和一个解码器组成, 结构如图 2 所示。编码器由 3 个 Conv 层、一个 Conv_t 层以及一个具有空间可分离注意力机制的 Transformer^[14] 模块组成。输入首先经过一个 1×1 的卷积层用于保留信息的同时降低计算成本, 其次进行一个 5×5 的卷积操作对输入特征进行更高级的特征提取, 然后进入 Transformer 模块, 最后通过 4×4 的卷积转置(反卷积)操作以及一个 4×4 的卷积操作, 提高模型对图像细节的感知, 提取更高级别的抽象特征。

在解码器设计上, SFNet 采用混合双径方法, 有机结合 CNN 与 Transformer。通过并行构建 2 条不同的卷积路径, 使网络能够灵活捕捉不同尺度特征, 显著增强对特征粒度的适应能力。

在 CNN 路径中, 引入 SFB, 以快速傅里叶卷积与残差模块取代传统 3×3 卷积, 实现全局与局部特征融合。其中, 频率分支利用 FFC 提取全局信息, 空间分支则通过空间可分离卷积改进的残差模块强化局部特征表达, 在降低计算成本的同时, 能够提取更全面、详细且稳定的特征。

在 2 条路径的衔接处, 嵌入特征融合网络 MASAG, 该网络能有效捕捉细节与全局信息。其独特的跳跃连接设计, 通过动态加权和组合粗粒度 CNN 路径特征与细粒度 Transformer 路径特征映射,

推动特征融合。具体而言, 借助多尺度自适应空间注意门动态调整感受野(局部与全局上下文信息), 确保网络有针对性地突出空间相关特征。

融合后的输出会再次通过 Transformer 模块, 进一步优化和精细化融合后的特征, 大幅提升信息表达能力。最终, 利用 Sigmoid 激活函数对输出范围进行归一化处理, 这不仅提升了训练的稳定性, 还加快了收敛速度, 使模型表现更为出色。

1.2 空间频率模块 SFB

在当前深度学习领域, 处理复杂数据集时的大规模学习需求, 致使模型结构日益庞大, 计算开销显著增加。针对这一问题, 本文提出了创新解决方案, 即引入低消耗的特征提取模块。鉴于快速傅里叶卷积具备提取全局特征的能力, 本文将传统的 3×3 卷积替换为快速傅里叶卷积与残差模块, 以此融合全局和局部特征, 并构建了 SFB。该模块有效提升了模型的表示能力, 成功解决了信息超载问题, 在降低计算成本的同时增强了模型性能。

SFB 网络架构由 2 个主要组件组成, 左侧的深度可分离卷积运算和右侧的 FFC。通过将左右输出连接起来, 并执行卷积运算以获得最终结果, 如式(6)所示。

$$X_{SFB} = H_{SFB}(X) \tag{6}$$

其中, X 表示经过 5×5 卷积后的特征图, $H_{SFB}(\cdot)$ 表示 SFB, X_{SFB} 表示 SFB 经过各种运算后的输出特征

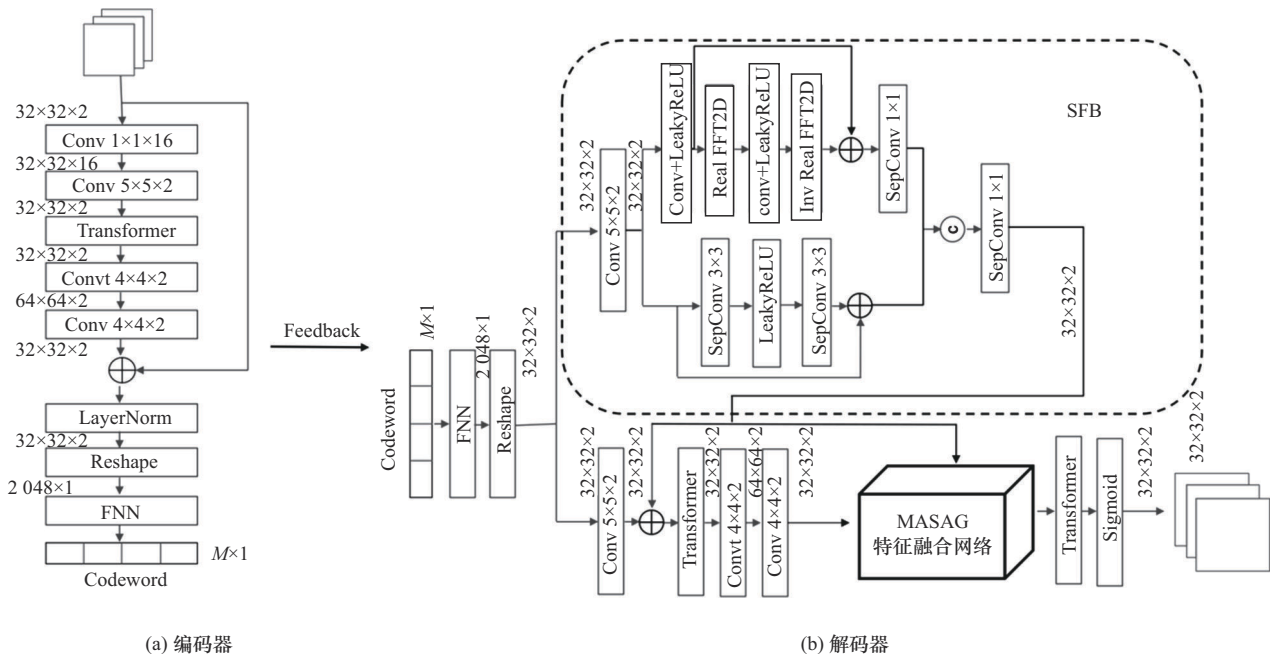


图 2 SFNet 结构

图。本文将 X 发送到 2 个不同的域, 即 X 空间和 X 频率。 X_{spatial} 用于空间域, $X_{\text{frequency}}$ 旨在捕获频域中的长程上下文, 如式(7)和式(8)所示。

$$X_{\text{spatial}} = H_{\text{spatial}}(X) \quad (7)$$

$$X_{\text{frequency}} = H_{\text{frequency}}(X) \quad (8)$$

其中, $H_{\text{spatial}}(\cdot)$ 表示深度可分离空间卷积核, $H_{\text{frequency}}(\cdot)$ 表示频率 FFC 核。 X_{spatial} 也可表示为

$$X_{\text{spatial}} = H_{\text{CLC}}(X) + X \quad (9)$$

其中, $H_{\text{CLC}}(\cdot)$ 表示头部和尾部的 3×3 卷积层, 并且在卷积层之间进行 LeakyReLU 运算。

在右频率模块中, 本文使用二维快速傅里叶变换 (FFT, fast Fourier transform) 将传统的空间特征转换到频域中以提取全局信息, 然后进行逆二维 FFT 运算以获得空间域特征。 $X_{\text{frequency}}$ 也可表示为

$$X_{\text{frequency}} = H_{\text{CL}}(X) \quad (10)$$

$$X_{\text{frequency}} = H_{\text{C}}(H_{\text{FLF}}(X) + X) \quad (11)$$

其中, $H_{\text{CL}}(\cdot)$ 表示卷积层和 LeakyReLU。然后通过卷积运算将通道的数量减少一半, 则有

$$X_{\text{SFB}} = H_{\text{C}}\left(\left[X_{\text{spatial}} \parallel X_{\text{frequency}}\right]\right) \quad (12)$$

其中, $H_{\text{C}}(\cdot)$ 表示卷积层, \parallel 表示级联算子。

为了降低计算成本, 本文在 SFB 中采用了可分离卷积技术, 将传统的卷积运算分解为逐点卷积和逐深度卷积两步, 显著减少了卷积层的计算复杂度和参数量。具体原理如图 3 所示。

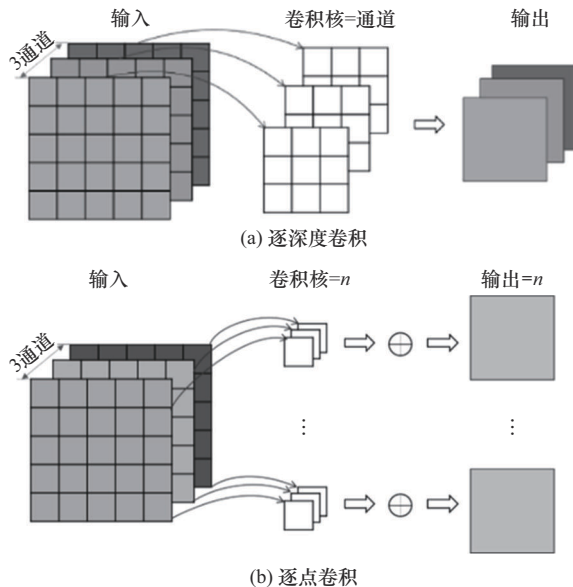


图 3 深度可分离卷积原理

深度可分离卷积将传统的卷积操作分解为逐深度卷积和逐点卷积两部分。逐深度卷积对每个输入通道分别进行卷积操作, 从而减少通道间的冗余计算。随后逐点卷积通过 1×1 卷积将不同通道的信息进行线性组合, 实现特征的融合与传递。

1.3 特征融合网络

无线信道的动态变化、空间相关性以及用户设备特性的多样性对 CSI 反馈的准确性和效率提出了重大挑战。传统方法通常依赖于固定的感受野和简单的特征融合策略, 这在捕获信道信息的多样性方面存在局限性。为了解决这一问题, 本文在混合 CNN-Transformer 架构的跳跃连接中引入了 MASAG 模块, 以提升 CSI 反馈的质量和鲁棒性, 具体结构如图 4 所示。MASAG 包含 4 个关键阶段, 分别为多尺度特征融合、空间选择、空间交互和交叉调制, 以及重新校准。多尺度特征融合阶段以语义上相似的方式聚合 X_1 和 X_2 , 为后续处理准备融合后的特征图; 空间选择阶段通过动态调整感受野优先考虑关键特征; 空间交互和交叉调制阶段进一步丰富特征图, 融合局部细节和全局上下文信息。在重新校准阶段, 初始输入通过 MASAG 生成的选择性注意力图进行优化, 从而确保反馈的空间精确性。通过这种方法, 网络能够有效应对无线信道环境中的多样化挑战, 确保 CSI 反馈的准确性和稳定性, 为大规模 MIMO 系统的性能提升提供可靠保障。

MASAG 模块的多尺度特征融合层集成了局部上下文提取和全局上下文提取, 将 CNN 路径的高分辨率空间细节 X_1 与 Transformer 路径的信息 X_2 进行融合。局部上下文提取通过深度卷积和膨胀卷积扩展了 X_1 的空间范围, 而全局上下文提取通过通道级池化操作从 X_2 中捕获广泛的上下文信息。这种融合生成了综合特征图, 为 MASAG 中的空间选择阶段奠定了基础, 从而确保在多样化的信道条件下实现准确的 CSI 反馈重建。多尺度特征融合可表示为

$$U = \underbrace{\text{Conv}_{1 \times 1}(\text{DW-D}(\text{DW}(X_1)))}_{\text{LocalContextExtraction}} + \underbrace{\text{Conv}_{1 \times 1}([P_{\text{Avg}}(X_2); P_{\text{Max}}(X_2)])}_{\text{GlobalContextExtraction}} \quad (13)$$

其中, DW 和 DW-D 分别表示深度卷积和扩展卷积, $\text{Conv}_{1 \times 1}$ 表示局部上下文提取的逐点卷积, P_{Max} 和 P_{Avg} 分别表示全局上下文提取的最大池化和平均池化, 符号 $[\cdot; \cdot]$ 表示拼接。

在 MASAG 的空间选择阶段, 由式(13)得到的融合特征映射 U 被投影到 2 个通道, 与输入特征 X_1 和 X_2 对齐, 通过通道级 softmax 计算的空间选择权重对特征 X_1 和 X_2 的注意力进行细化, 生成了空间选择后的特征图 X_1' 和 X_2' 。

在此基础上, 本文引入了 CBAM 注意力机制, 进一步优化了特征图 X_1' 和 X_2' 的重要区域, 提升了关注度对关键信息的筛选能力, 同时减少了对不相关信息的注意力, 增强了感受野的选择能力。此外, X_1' 和 X_2' 的 2 个残差连接提高了梯度流和特征利用率, 进一步通过精确的上下文感知注意力细化了分割结果。此过程表示为

$$SW_i = S(\text{Conv}_{1 \times 1}(U)), \forall i \in [1, 2] \quad (14)$$

$$X_1' = SW_1 \otimes X_1 + X_1, X_2' = SW_2 \otimes X_2 + X_2 \quad (15)$$

其中, $SW_i \in R^{2 \times H \times W}$, 由通道相关的 softmax (S) 导出, 确保 2 个通道 ($i \in [1, 2]$) 的权重在每个空间位置之和为 1, 表示根据内容对输入特征的相对贡献。

在 MASAG 的空间交互和跨调制阶段, 该网络通过将空间上细化后的输入特征 X_1' 和 X_2' (源自式(15)) 进行动态增强, 实现了更加精确的分割。空间相互作用的表述为

$$X_1'' = X_1' \otimes \sigma(X_2'), X_2'' = X_2' \otimes \sigma(X_1') \quad (16)$$

$$U' = X_1'' \otimes X_2'' \quad (17)$$

其中, $\sigma(X_2')$ 和 $\sigma(X_1')$ 分别表示局部和全局空间权值, U' 表示乘入融合特征映射。

在重新校准阶段, 来自空间交互与交叉调制的融合特征图首先通过逐点卷积进行精细化处理, 然后经过 Sigmoid 激活函数生成一个聚焦的注意力图。该注意力图用于重新校准编码器的初始输入 X_1 。通过这一过程, X_1 得到了由前述阶段自适应调整的多尺度感受野, 从而为与网络输出的整合做好充分准备。最终, 经过这一精细调整的 X_1 将具备精确且上下文感知的特征, 能够有效促进 CSI 反馈的高精度分割。该重新校准过程表示为

$$X_1 = \text{Conv}_{1 \times 1}(\sigma(\text{Conv}_{1 \times 1}(U')) \otimes X_1) \quad (18)$$

此阶段的重新校准有效地提升了特征图的精度和相关性, 为后续的处理提供了更加可靠的输入, 确保了分割任务的高效执行。

1.4 SFNet 的量化与去量化处理

基于深度学习的自动编码器核心构成部分为编码器网络与解码器网络。在 UE 端, 首先对估计得到的 CSI 矩阵展开预处理, 以此获取反馈数据。随后, 编码器网络对预处理后的信息进行特征提取, 并将 CSI 矩阵压缩为低维码字, 该码字通过数字通信信道传输回基站。

在基站处, 解码器网络将接收到的码字恢复为 CSI 矩阵, 接着历经二维离散傅里叶变换 (2DDFT)、零填充等后处理步骤, 完成原始 CSI 矩阵的重建。

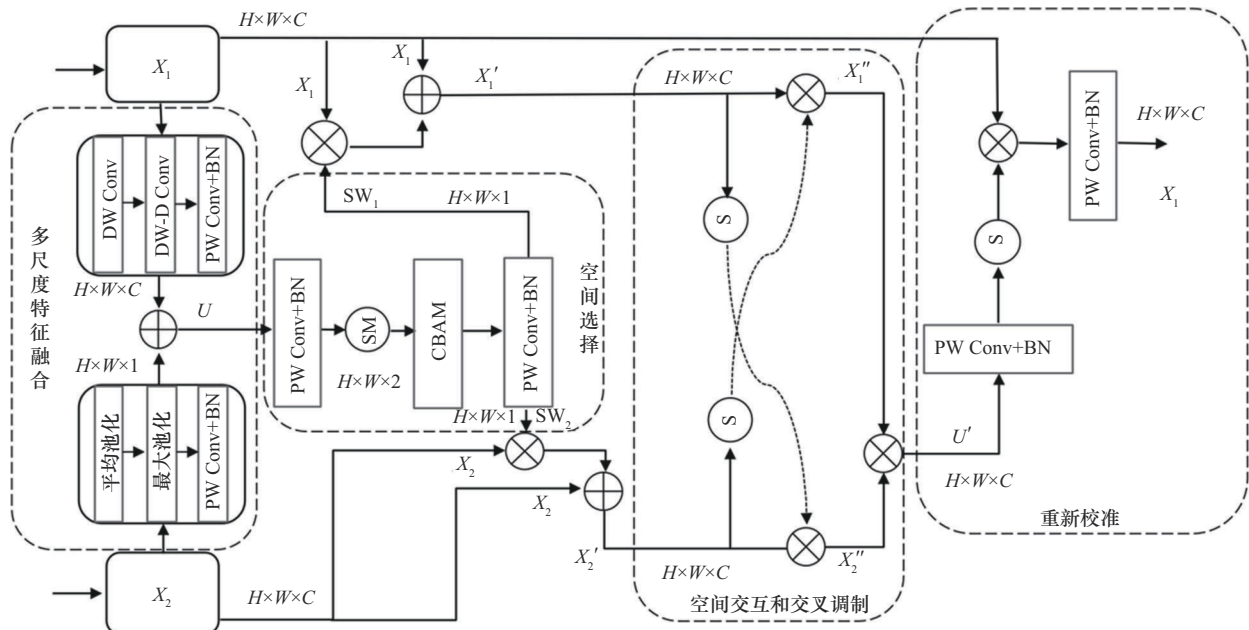


图 4 MASAG 结构

在数字通信系统中，信号需转换为比特流以实现反馈，具体结构可参考图 1。

经编码器压缩后，生成的码字由完整的 32 位单精度浮点数 (FP32) 构成。若直接传输这些浮点数，会给反馈信道带来极大的负担，导致难以承受的开销。鉴于此，在反馈前，需将码字量化为较低位宽。本文采用文献[18]提出的轻量级量化方案，对自主构建的网络 SFNet 进行量化处理，针对这一过程，本文分别设定量化比特数为 2、4、6 和 8，即每个经过量化后的实数码字将以 2、4、6 或 8 比特流进行传输。此方案在保障性能的同时，有效降低了反馈信道的传输压力。

本文采用了一种 μ 律量化器，它在 CSI 反馈任务中表现良好。 μ 律量化器是一种非均匀量化，通过一个非递减的压缩函数 $\Phi: (0,1) \rightarrow (0,1)$ ，在零附近提供更窄的量化区间，在 1 附近提供更宽的量化区间，可以很好地处理零集中数据。实验表明，CSI 反馈任务的压缩码字集中在零附近。 μ 律量化器中的压缩函数描述为

$$\Phi(x) = \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad (19)$$

其中， x 是归一化输入， μ 是改变扩展程度的参数，需要根据给定数据的分布进行调整。在操作 μ 律量化器时，首先将归一化后的数据放入压缩函数 Φ 中，用均匀量化器进行量化，然后通过压缩函数 Φ^{-1} 展开得到最终量化值。在实际部署中，可将压缩函数 Φ 近似为分段折线函数，大大降低了计算复杂度。

2 仿真结果分析

2.1 数据与参数设置

本文使用了文献[6]中描述的同数据集。采用了 COST 2100 的通道模型，这是一个基于几何结构的随机模型。它涉及 2 种环境：室内微微蜂窝设置在 5.3 GHz 下运行，室外农村场景设置在 300 MHz 下运行。该系统使用了具有 32 个天线和 1 024 个子载波的均匀线性阵列，其中基站端位于正方形区域的中心，用户设备端随机位于同一区域内。COST 2100 模型包括 3 种不同的集群类型。通过叠加这些集群的特性得到的信道脉冲响应形成 MIMO 信道矩阵。仿真参数如表 1 所示。

表 1 MIMO 信道仿真参数

参数名称	室内	室外
载波频率/MHz	5 300	300
双工模型	FDD	FDD
子载波数	1 024	1 024
发射天线	32	32
接收天线	1	1

本文提出的 SFNet 模型选择天线数 $N_c=32$ ，窗口大小 $W=8$ ，多头注意力的头数 $P=4$ 。数据集包括 150 000 个样本，其中 100 000 个用于训练，30 000 个用于验证，20 000 个用于测试。通过选择最初的 32 行，将信道矩阵转换为角延迟域中的 32×32 矩阵。批量大小设置为 200，epoch 设置为 1 500，学习率为 0.001，损失函数为 Adam 优化器的均方误差 (MSE, mean squared error)。

$$\text{MSE} = \frac{1}{B} \sum_{i=1}^B \| \mathbf{H} - \hat{\mathbf{H}} \|^2 \quad (20)$$

其中， \mathbf{H} 是输入通道矩阵， $\hat{\mathbf{H}}$ 是重构的通道矩阵， B 是批量大小。

2.2 模型性能分析

本文使用归一化均方误差 (NMSE, normalized mean square error) 作为性能指标，其定义为

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\| \mathbf{H} - \hat{\mathbf{H}} \|^2}{\| \mathbf{H} \|^2} \right\} \quad (21)$$

为了测量波束成形矢量的质量，本文还考虑了余弦相似性。

$$\rho = \mathbb{E} \left\{ \frac{1}{\tilde{N}_c} \sum_{n=1}^{\tilde{N}_c} \frac{|\hat{\mathbf{h}}_n^H \tilde{\mathbf{h}}_n|}{\|\hat{\mathbf{h}}_n\|_2 \|\tilde{\mathbf{h}}_n\|_2} \right\} \quad (22)$$

其中， $\hat{\mathbf{h}}_n$ 是时刻 t 的第 n 个子载波重建信道向量， $\tilde{\mathbf{h}}_n$ 是第 n 个子载波的信道向量。

本节将本文提出的 SFNet 与其他 4 种基于深度学习的 CSI 反馈算法 (CSINet、CRNet、CLNet 和 STNet) 进行性能比较，从 CSI 重构的效果与模型复杂度 2 个核心方面，对仿真结果进行了全面分析。仿真参数一致，epoch 均设置为 1 500，且所有模型均在相同的硬件平台与环境下进行训练和测试。重构性能采用归一化均方误差和余弦相似度作为评价指标，而复杂度则通过模型的参数量和每秒浮点运算次数 (FLOPS) 来表征，分别对应存储需

求和计算开销。本文分别在室内和室外场景下用不同的压缩比 ($\frac{1}{4}$ 、 $\frac{1}{8}$ 、 $\frac{1}{16}$ 、 $\frac{1}{32}$ 、 $\frac{1}{64}$) 进行实验。

表2对比了不同网络在不同压缩比和场景下的NMSE性能。从表2中可以清晰地看出, SFNet在整体性能上优于其他基准网络, 展现了优秀的CSI重构能力。在室内场景下, 本文提出的SFNet优于其他网络, 特别在压缩比为 $\frac{1}{4}$ 时的NMSE达到了-31.3 dB, 显著优于CSINet、CRNet、CLNet和STNet, 分别提升了13.94%、21.6%、16.8%和1.75%。即使在低压缩比 $\frac{1}{64}$ 时, SFNet仍以-7.8 dB保持领先, 相比次优的STNet减少了0.8 dB的误差。

表2 不同网络NMSE性能比较

压缩比 γ	场景	NMSE/dB				
		CSINet	CRNet	CLNet	STNet	SFNet
$\frac{1}{4}$	室内	17.36	-25.72	-26.88	-30.84	-31.3
	室外	-8.75	-11.6	-12.45	-12.65	-12.4
$\frac{1}{8}$	室内	-13	-15.6	-15.1	-20.62	-21.7
	室外	-7.6	-8.1	-8.4	-8.17	-8
$\frac{1}{16}$	室内	-8.65	-10.46	-10.59	-14.6	-14.6
	室外	-4.51	-5.03	-5.4	-5.43	-5.53
$\frac{1}{32}$	室内	-6.24	-8.36	-8.8	-8.85	-9.57
	室外	-2.81	-3.2	-3.25	-3.5	-2.88
$\frac{1}{64}$	室内	-5.84	-6.1	-6.24	-6.96	-7.8
	室外	-1.93	-2.13	-2.22	-2.49	-2.4

在室外场景下, SFNet同样展现出性能优势。当压缩比为 $\frac{1}{4}$ 时, SFNet的NMSE为-12.4 dB, 相比CSINet、CRNet和CLNet分别提升了37.14%、3.45%和3.61%。当压缩比增大至 $\frac{1}{16}$ 时, SFNet在所有网络中的性能最佳, 达到-5.53 dB。综合来看, SFNet在不同压缩比和场景下均展现出更优的性能, 可以看出空间频率模块和特征融合网络的结合使网络在保证高重构精度的同时适应多种反馈条件, 展现了SFNet在CSI反馈任务中卓越的重构效果与实际应用潜力。

在余弦相似度方面, 基于深度学习的方法相较于传统压缩感知算法有着显著的性能提升, 这是由于基于传统压缩感知的算法中在低压缩比的情况下, 高斯随机矩阵采样不能够提取到有效的信息,

没有办法重建出有效的信号。基于深度学习的方法虽然也受压缩比的影响, 但是与传统压缩感知算法相比, 在低压缩比的情况下神经网络能够提取到更多的原始信号, 重建性能更佳。因此, 本文对近期主流的深度学习模型进行余弦相似度的分析比较, 如图5所示。从余弦相似度的结果来看, SFNet在室内和室外2种场景下均展现了显著的性能优势。在室内场景中, SFNet在压缩比为 $\frac{1}{16}$ 时表现尤为突出, 相较于CLNet的0.89提升了约10.1%。这种显著的性能提升表明, SFNet能够更有效地捕获室内场景的特征信息。在室外场景中, SFNet同样表现出稳定的性能优势, 进一步说明了其对室外复杂信道环境的适应能力。

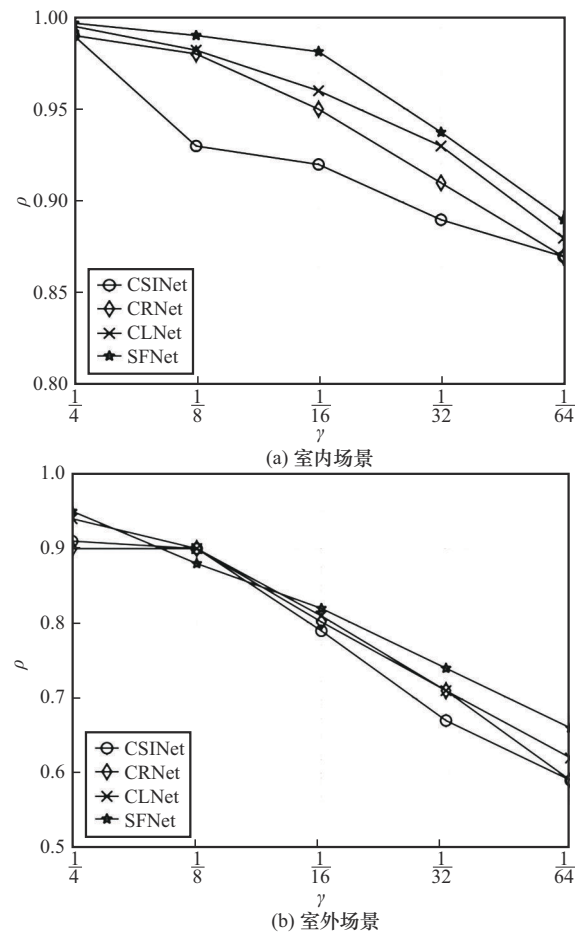


图5 不同网络的余弦相似度比较

为了进一步验证本文提出的SFB与MASAG特征融合网络的优势, 本文进行了消融实验, 无添加以及加入SFB与MASAG情况下的NMSE性能如表3所示。

表3 消融实验

压缩比 γ	场景	NMSE/dB		
		无添加	加入 SFB	加入 MASAG
$\frac{1}{4}$	室内	-30.84	-31.06	-31.3
	室外	-12.65	-12	-12.4
$\frac{1}{8}$	室内	-20.62	-21.02	-21.7
	室外	-8.17	-7.7	-8
$\frac{1}{16}$	室内	-14.6	-14.42	-14.6
	室外	-5.43	-5.5	-5.53
$\frac{1}{32}$	室内	-8.85	-9.5	-9.57
	室外	-3.5	-2.73	-2.88
$\frac{1}{64}$	室内	-6.96	-7.5	-7.8
	室外	-2.49	-2.4	-2.4

由表3可以得到,随着SFB与MASAG的陆续加入,在大部分的场景下,NMSE性能都会随之提升。例如,在室内的 $\frac{1}{4}$ 场景下,NMSE性能由-30.84 dB提升到-31.06 dB,随着MASAG的加入更是提升到了-31.3 dB,展示了SFB与MASAG的卓越表现。此外,本文还对CBAM模块进行了验证,在室内和室外场景下取压缩比为 $\frac{1}{4}$ 、 $\frac{1}{8}$ 和 $\frac{1}{16}$ 进行了消融实验,结果如表4所示。

表4 CBAM模块消融实验

压缩比 γ	场景	NMSE/dB	
		无添加	加入CBAM
$\frac{1}{4}$	室内	-30.2	-31.3
	室外	-11.2	-12.4
$\frac{1}{8}$	室内	-21.37	-21.7
	室外	-7.84	-8
$\frac{1}{16}$	室内	-12.44	-14.6
	室外	-4.83	-5.53

为了验证本文模型中引入的CBAM模块对性能的实际影响,本文在不同压缩比($\gamma=\frac{1}{4}$ 、 $\frac{1}{8}$ 和 $\frac{1}{16}$)及通信场景(室内与室外)下进行了消融实验。结果表明,引入CBAM模块后,模型在所有设置下均获得了性能提升。例如,在压缩比为 $\frac{1}{4}$ 时,室内场景的NMSE由-30.2 dB提升至-31.3 dB,

室外场景的NMSE由-11.2 dB提升至-12.4 dB。在压缩比为 $\frac{1}{16}$ 时,尽管压缩率较高,CBAM模块依然带来了显著的性能增益。这一结果充分说明CBAM模块能够有效提升模型对关键信道特征的关注能力,增强特征表达效果,从而提升CSI重建精度,并在不同通信场景下均展现出良好的适应性与稳定性。

此外,本文研究了SFNet中解码器端的头部卷积设计,分别设置卷积核为0、1、3和5进行研究,结果如表5所示。

表5 不同头部卷积层下的NMSE性能比较(压缩比为 $\frac{1}{4}$)

头部卷积层	室内		室外	
	NMSE/dB	FLOPS	NMSE/dB	FLOPS
无卷积	-30.3	4.912×10^6	-12.2	4.912×10^6
1×1	-29.8	4.916×10^6	-12	4.916×10^6
3×3	-30.2	4.942×10^6	-12.2	4.942×10^6
5×5	-31.3	4.995×10^6	-12.4	4.995×10^6

一般来说,基于纯残差架构的神经网络需要头部卷积层来大致提取特征。然而,在SFNet中,本文发现头部卷积层对最终性能影响不大。对于室内场景 $n=4$,如果去掉5×5的头部卷积层,NMSE只会降低约1 dB。此外,不同头部层的表现相对相似。这可能得益于SFNet的多路径架构,能够更好地适应CSI特征的变化。尽管如此,为了进一步优化性能,本文在SFNet中仍然保留了一个5×5的头部卷积层,因为其引入的额外计算量(浮点运算)可以忽略不计。

本文还研究了网络末端不同激活函数的选择对网络性能的影响。通过使用当下流行的不同激活函数,在室内压缩比为 $\frac{1}{4}$ 的情况下对网络的NMSE性能进行分析,具体结果如表6所示。

根据表6可知,Sigmoid激活函数在室内和室外场景下均表现最优,能够有效提升性能,说明其在CSI反馈的优化中具有最好的性能。LeakyReLU激活函数紧随其后,在2个场景中的表现也十分出色,但略逊于Sigmoid。其他激活函数在这2个场景中的表现较为接近,均逊色于Sigmoid和ReLU。出现这一现象的原因可能包括以下几点。首先,Sigmoid的输出范围限定在0~1,相比ReLU激活函

数更能有效抑制解码结果中的极端值波动, 有助于保障模型输出的稳定性与合理性。其次, Sigmoid 在整个定义域上保持平滑的非线性变化, 这种压缩特性可能更适用于对高维 CSI 数据中细节结构的拟合, 尤其是在重建误差较小时, 表现出更强的微调能力。此外, Sigmoid 激活函数对梯度的保持更为连续, 有利于模型训练过程的稳定收敛。最后, 从任务目标出发, CSI 重建的输出本身通常为归一化后有界的复数实部和虚部, Sigmoid 的输出范围与之匹配程度更高。因此, 在本文中采用 Sigmoid 激活函数有助于提升 CSI 重建的整体表现。因此, 在实际应用中, Sigmoid 激活函数可能是处理 CSI 反馈时的最佳选择。

表 6 不同激活函数性能比较

激活函数	NMSE/dB	
	室内	室外
Sigmoid	-31.3	-12.4
ReLU	-27.7	-11.3
LeakyReLU	-29.8	-11.9
Tanh	-29.7	-12.2
R_ReLU	-28.8	-12.2

2.3 SFNet 量化性能分析

本文在 SFNet 的编码器和解码器中分别加入量化模块和去量化模块, 通过引入量化层来减少信道矩阵传输过程中的反馈开销。具体参数设置如下, 量化比特数 B 分别设置为 2、4、6 和 8, μ 律量化器的压缩函数参数 μ 设为 50, 其他参数保持不变。本文考虑了 SFNet 在不同量化精度下的 NMSE 性能, 对比结果如图 6 所示。

实验结果表明, 在相同压缩比条件下, 随着量化精度的不断提高, NMSE 在大部分情况下呈现逐渐降低的趋势。值得注意的是, 在对网络进行 8 bit 非均匀量化的实验中, 本文方法在部分通信场景下的 CSI 反馈性能甚至优于未进行量化处理的情况。该现象表明, 在保持较低量化误差的前提下, 适度的非均匀量化不仅不会削弱模型的重建能力, 反而可能在一定程度上提升模型的泛化能力。这一效应类似于深度学习中常用的正则化策略 (如 L2 正则化或 Dropout), 其通过引入轻微扰动防止模型对训练集的过拟合, 从而提高在测试集上的表现。因而,

合理设计的量化模块不仅能够显著降低反馈开销, 还可能提升网络在不同信道场景下的鲁棒性和泛化性。

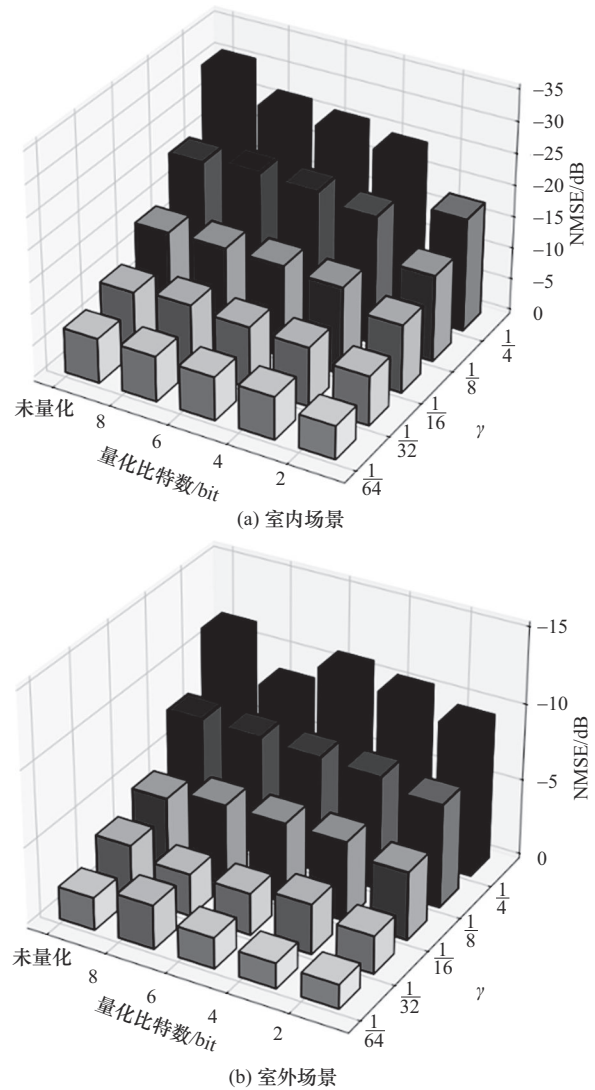


图 6 不同量化精度下的 NMSE 性能对比

为了更加清楚直观地展示, 本文在不同的网络模型中也加入了量化模块, 取量化比特数为 8, 在相同条件下对它们的 NMSE 性能进行了比较, 结果如图 7 所示。

在不同环境和压缩比条件下, SFNet 均展现出良好的性能稳定性和适应能力。在室内场景下, SFNet 在 $\frac{1}{4}$ 、 $\frac{1}{8}$ 和 $\frac{1}{64}$ 压缩比下的重建精度始终优于现有主流方法。而在复杂多变的室外环境中, SFNet 在 $\frac{1}{32}$ 和 $\frac{1}{64}$ 压缩比场景下依然取得了小幅提升, 说明该模型不仅能适应高密度、多路

径的室内信道结构，也具备在环境差异较大的条件下维持性能的能力，反映出较强的模型泛化能力。

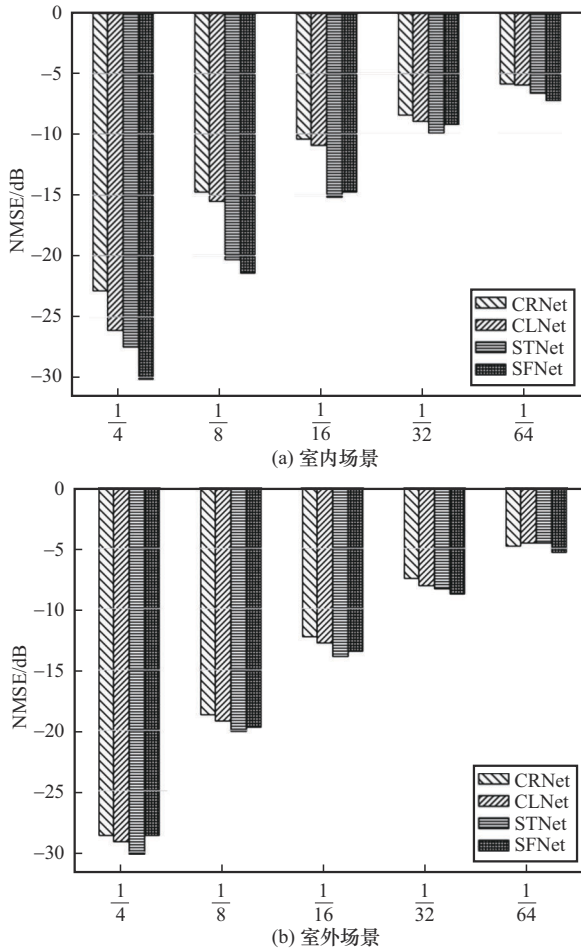


图7 不同模型的NMSE性能比较

量化模块的加入证明了本文提出的SFNet模型在实际的反馈场景中也有着不错的性能表现。在实际应用中，需要综合考虑总体反馈成本和测试结果，来选择合适的压缩比和量化级别，以达到最佳的重建性能。

2.4 复杂度分析

在复杂度方面，本文评估了模型的空间和时间复杂度，空间复杂度是指参数的数量或网络容量 (MP, model parameter)，时间复杂度表示 FLOPS。

全连接层的复杂度计算：假设输入的数据维度 (N, D) ，隐藏层的权重维度 (D, out) ，输出 (N, out) ，那么它的 FLOPS 和 MP 计算式分别为

$$FLOPS = (2D - 1) out \quad (23)$$

$$MP = D_{out} + out \quad (24)$$

卷积层的复杂度计算：假设输出特征图的宽和高分别为 W 和 H ，卷积核大小为 $K_1 \times K_2$ ， C_{in} 为输入通道数（卷积核的层数）， C_{out} 为输出通道数（卷积核的个数），那么它的 FLOPS 和 MP 计算式分别为

$$FLOPS = WHK_1 K_2 C_{in} C_{out} \quad (25)$$

$$MP = K K C_{in} C_{out} \quad (26)$$

通过上面的计算方式，可以得到 CSINet、CRNet、CLNet、STNet 以及 SFNet 的 MP 和 FLOPS，本文将得到的 MP 和 FLOPS 汇总并进行了比较，结果如表 7 所示。

表7 不同网络复杂度分析

压缩比 γ	指标	复杂度				
		CSINet	CRNet	CLNet	STNet	SFNet
$\frac{1}{4}$	MP	2.103×10^6	2.1×10^6	2.102×10^6	2.125×10^6	2.124×10^6
	FLOPS	4.201×10^6	5.12×10^6	4.42×10^6	5.97×10^6	4.995×10^6
$\frac{1}{8}$	MP	1.054×10^6	1.06×10^6	1.053×10^6	1.076×10^6	1.075×10^6
	FLOPS	2.103×10^6	4.07×10^6	3.37×10^6	5.132×10^6	4.156×10^6
$\frac{1}{16}$	MP	0.53×10^6	0.53×10^6	0.53×10^6	0.551×10^6	0.55×10^6
	FLOPS	1.055×10^6	3.55×10^6	2.85×10^6	4.712×10^6	3.736×10^6
$\frac{1}{32}$	MP	0.268×10^6	0.27×10^6	0.27×10^6	0.29×10^6	0.288×10^6
	FLOPS	0.53×10^6	3.28×10^6	2.58×10^6	4.502×10^6	3.527×10^6
$\frac{1}{64}$	MP	0.137×10^6	0.14×10^6	0.14×10^6	0.158×10^6	0.157×10^6
	FLOPS	0.269×10^6	3.16×10^6	2.45×10^6	4.397×10^6	3.422×10^6

本文模型与改良前的STNet相比,具有更低的浮点运算数FLOPS,在模型参数方面也有着轻微的优势,NMSE性能也有着明显的提升,证明了空间频率块和特征融合网络的使用能在保证NMSE性能优越的前提下,显著降低网络的复杂度。

3 结束语

针对大规模MIMO系统中CSI反馈方法复杂度高、反馈精度低以及未考虑量化损失的问题,本文提出了一种轻量化、低复杂度的CSI反馈框架SFNet,通过结合一种利用全局信息而设计的SFB以及一种融合局部和全局特征的特征多尺度融合网络MASAG,使网络可以激活更多的输入信息,以确保有选择地突出空间相关的特征,使网络能够以非常低的计算复杂度实现先进的性能。此外,考虑到实际通信系统中量化损失的影响,本文还引入了一种轻量级非均匀量化方案,通过在编码器和解码器之间加入量化模块对SFNet进行量化操作。仿真结果表明,本文提出的SFNet相较于STNet等算法,拥有更高的反馈性能和更低的计算复杂度,同时能够有效减少量化误差,增加信号的重建精度。

参考文献:

- [1] MARZETTA T L. Noncooperative cellular wireless with unlimited numbers of base station antennas[J]. IEEE Transactions on Wireless Communications, 2010, 9(11): 3590-3600.
- [2] ZHANG J, WEN C K, JIN S, et al. On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas[J]. arXiv Preprint, arXiv: 1209.5513, 2012.
- [3] DAUBECHIES I, DEFRISE M, MOL C D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on Pure and Applied Mathematics, 2004, 57(11): 1413-1457.
- [4] DONOHO D L, MALEKI A, MONTANARI A. Message-passing algorithms for compressed sensing[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(45): 18914-18919.
- [5] LI C B, YIN W T, ZHANG Y. User's guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms[J]. Department of CAAM, 2009, 20: 46-47.
- [6] 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.
- [7] GUO J J, WEN C K, JIN S, et al. Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: design, simulation, and analysis[J]. IEEE Transactions on Wireless Communications, 2020, 19(4): 2827-2840.
- [8] LU Z L, WANG J T, SONG J. Multi-resolution CSI feedback with deep learning in massive MIMO system[C]//Proceedings of the ICC 2020-2020 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2020: 1-6.
- [9] JI S J, LI M. CLNet: complex input lightweight neural network designed for massive MIMO CSI feedback[J]. IEEE Wireless Communications Letters, 2021, 10(10): 2318-2322.
- [10] XU Y, YUAN M Q, PUN M O. Transformer empowered CSI feedback for massive MIMO systems[C]//Proceedings of the 2021 30th Wireless and Optical Communications Conference (WOCC). IEEE, 2021: 157-161.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv Preprint, arXiv: 1706.03762, 2017.
- [12] CUI Y D, GUO A H, SONG C L. TransNet: full attention network for CSI feedback in FDD massive MIMO system[J]. IEEE Wireless Communications Letters, 2022, 11(5): 903-907.
- [13] 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.
- [14] MOURYA S, AMURU S, KUCHI K K. A spatially separable attention mechanism for massive MIMO CSI feedback[J]. IEEE Wireless Communications Letters, 2023, 12(1): 40-44.
- [15] ZHANG D F, HUANG F Y, LIU S Z, et al. SwinFIR: revisiting the SwinIR with fast Fourier convolution and improved training for image super-resolution[J]. arXiv Preprint, arXiv: 2208.11247, 2022.
- [16] KOLAH S G, CHAHARSOOGHI S K, KHATIBI T, et al. MSA2Net: multi-scale adaptive attention-guided network for medical image segmentation[J]. arXiv Preprint, arXiv: 2407.21640, 2024.
- [17] CHI L, JIANG B R, MU Y D. Fast Fourier convolution[J]. Advances in Neural Information Processing Systems, 2020, 33: 4479-4488.
- [18] ZHANG X D, LU Z L, ZENG R, et al. Quantization adaptor for bit-level deep learning-based massive MIMO CSI feedback[J]. IEEE Transactions on Vehicular Technology, 2024, 73(4): 5443-5453.

[作者简介]



张昀 (1975-), 女, 江苏南京人, 博士, 南京邮电大学副教授, 主要研究方向为智能化算法与通信信号处理。



黄经纬 (2001-), 男, 江苏淮安人, 南京邮电大学硕士生, 主要研究方向为深度学习与信号处理。



徐孙武 (2001-), 男, 江苏南京人, 南京邮电大学硕士生, 主要研究方向为深度学习与信号处理。



于舒娟 (1967-), 女, 江苏南京人, 南京邮电大学教授, 主要研究方向为自适应信号处理、深度学习和智能大数据处理。



高贵 (1998-), 男, 安徽阜阳人, 南京邮电大学硕士生, 主要研究方向为深度学习与信号处理。



赵生妹 (1968-), 女, 江苏丹徒人, 南京邮电大学教授, 主要研究方向为量子通信与信息处理、无线通信与信号处理。