

基于深度学习的IRS辅助MIMO通信系统的CSI压缩及恢复研究

黄富铿

广东工业大学, 信息工程学院, 广东 广州

收稿日期: 2022年1月22日; 录用日期: 2022年2月11日; 发布日期: 2022年2月18日

摘要

智能反射面(IRS, Intelligent Reflecting Surface)因成本低、功耗低、可提升通信质量等优点被广泛研究。在采用正交频分复用作为多载波调制方案的IRS辅助频分双工多输入多输出(MIMO, Multiple input Multiple Output)通信系统中, 为了提升系统的系统增益, 用户端(UE, User Equipment)需要将多个信道的信道状态信息(CSI, Channel State Information)通过反馈链路发送至基站端(BS, Base Station)。因此, 相比于传统的MIMO系统, 该系统中CSI的数据量和反馈开销无疑将会是更加巨大的。针对此问题, 本文提出了一种基于注意力机制的深度残差网络IARNet (Inception-Attention-Residual-Net)来对大数据量的CSI进行压缩重建。该网络在传统的Inception网络结构上结合了多卷积特征融合、混合注意力机制以及残差等子模块, 这种混合结构可以有效地将大数据量的CSI进行压缩重建。仿真结果表明, 与现有的2种深度学习网络相比, IARNet在基于热身法的模型训练方案加持下可以显著提高大数据量CSI的重建质量。

关键词

智能反射面, 深度学习, 信道状态信息反馈

Study on CSI Compression and Restoration with Deep Learning in RIS-Assisted MIMO Systems

Fukeng Huang

School of Information Engineering, Guangdong University of Technology, Guangzhou Guangdong

Received: Jan. 22nd, 2022; accepted: Feb. 11th, 2022; published: Feb. 18th, 2022

Abstract

Intelligent reflective surfaces (IRS) have been widely studied due to their advantages such as low cost, low power consumption, and ability to improve communication quality. In this paper, an IRS-assisted multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) communication system is considered. In order to improve the performance gain of the system, the user (UE) needs to send its channel state information (CSI) of several channels to the base station (BS) via feedback link. Therefore, the data volume and feedback overhead of CSI in this system will undoubtedly be much huger, as compared to the conventional MIMO systems. To address this problem, we propose an attention-based deep residual network named IARNet (Inception-Attention-Residual-Net) to compress and reconstruct the CSI with large data volume. The IARNet combines several sub-modules based on the traditional Inception network, such as the multi-convolutional feature fusion module, the hybrid attention module, and the residual module, etc. This hybrid structure can effectively compress and reconstruct the CSI of large data volumes. Simulation results show that with the warm-up training scheme, IARNet can significantly improve the reconstruction quality of CSI of large data volumes, as compared to two existing deep learning networks.

Keywords

Intelligent Reflecting Surface, Deep Learning, CSI Feedback

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着 5G 通信网络进入商业化阶段，为了获得更快和更可靠的数据传输，6G 通信技术已经处于研究状态，其中智能反射面(IRS, Intelligent Reflecting Surface)技术因其成本低、易部署、功耗低、可提升通信质量等特点被应用到各种无线通信系统。IRS 是一种有大量无源反射单元的表面，该表面的反射单元可以将入射信号进行被动反射，通过调整 IRS 的反射系数还可以进一步提高反射信号的传输质量。因为 IRS 十分轻巧，所以人们可以轻易地将其部署在建筑外墙、广告面板和楼顶等地方。在信号反射的过程中，由于 IRS 除了控制反射单元以外无需消耗额外能量的特点，因此 IRS 被业界广泛视为一种绿色、环保以及有前景的技术。基于上述优点，IRS 技术很好地契合了现阶段人们对 6G 的愿景，即智能、融合、绿色[1] [2] [3] [4] [5]。

深度学习是一种通过构造深层网络自动地提取出数据内在特征和规律的人工智能技术。自从 2012 年 Geoffrey Hinton 等人使用深度学习技术并以绝对优势获得了 ImageNet 图像识别比赛的冠军以来，越来越多的研究者参与对深度学习的研究并取得了巨大进展。最近研究表明，深度学习技术不仅在图像识别领域有杰出表现，而且在自然语言处理和图像压缩等领域也取得了不俗的成绩[6]-[18]。近年来有很多通信领域的研究者将深度学习技术应用在了通信相关领域，和传统的通信算法相比，深度学习在信道估计、信号检测和 CSI (CSI, Channel State Information)反馈等方向上获取了更好的表现。

针对 CSI 反馈开销过大的问题，文献[15]首次提出了使用深度学习技术将 CSI 进行压缩再重建，并提出了名为 CsiNet 的深度学习网络。相比于传统压缩感知的方法，CsiNet 有更好的重建质量和重建速度。

文献[16]在 CsiNet 的基础上引入了 Inception 模块,提出了多分辨率体系结构的网络:CRNet。相比于 CsiNet, CRNet 可以在网络参数变化不大的情况下进一步提升重建的质量。文献[17]在 CsiNet 的基础上引入了 Dense Block 模块,提出了有极致残差模块的网络:DS-NLCsiNet。相比于 CsiNet, DS-NLCsiNet 进一步提高了重建质量和恢复精读。文献[18]在 CsiNet 基础上引入了量化模块,提出了 QuanCsiNet。相比于 CsiNet, QuanCsiNet 可以进一步压缩反馈的 CSI。此外文献[18]在训练深度学习网络的时候还使用了基于真实信道的数据集,这进一步表明了基于深度学习的 CSI 压缩反馈确实是有效的。

但是现有的网络和工作都是在压缩和重建数据量较小的 CSI,数据量一般都不超过 2048 个 32 位浮点数。在 IRS 辅助的频分双工(FDD, Frequency Division Duplex)模式下的多输入多输出(MIMO, Multiple input Multiple Output)通信系统中采用正交频分复用(OFDM, Orthogonal Frequency Division Multiplexing)作为多载波的传输方案。该系统中下行链路反馈的 CSI 不仅包括基站端(BS, Base Station)到用户端(UE, User Equipment)的 CSI,还需要包括 BS 到 IRS 的 CSI 以及 IRS 到 UE 的 CSI,因此该系统的反馈开销将会是更加巨大的,同时使用深度学习将 CSI 进行压缩和重建的时候数据量也会大大增加。在本系统中压缩和重建的数据量超过了一般工作研究的 4 倍,达 8704 个 32 位浮点数。现有的网络在本系统中对数据量更大的 CSI 进行压缩重建的时候会出现重建质量低下的问题。因此需要针对数据量更大的 CSI 设计出一种新的深度学习网络来将 CSI 压缩和重建,以提升系统的重建质量。

本文针对 IRS 辅助的通信系统中反馈开销更加巨大的问题提出了一种新的深度学习网络 IARNet 以及基于热身法的模型训练方案[7]。IARNet 在传统卷积神经网络基础上采用了多卷积特征融合、混合注意力机制以及残差等模块。通过仿真发现:与现有的深度学习网络相比, IARNet 在基于热身法的模型训练方案加持下可以显著提高 CSI 重建质量,即使是在较低压缩比下 IARNet 仍能很好地将 CSI 恢复出来。本文的贡献总结如下:

1) 研究了在 IRS 辅助下的 MIMO 通信系统的 CSI 压缩及重建问题,并提出了相关系统模型。

2) 针对一般深度学习网络在大数据量 CSI 压缩重建过程中出现重建质量低下的问题,本文在传统卷积神经网络的基础上加入了多卷积特征融合、混合注意力机制以及残差等模块,提出了深度学习网络 IARNet,实验表明,在基于热身法的模型训练方案加持下可以显著提高 CSI 重建质量。

3) 进一步研究了基于热身法的学习率调整策略与三种传统的学习率调整策略在 1/8 压缩比下对网络的性能影响,实验表明:相比于传统方法基于热身法的学习率调整策略可以进一步提高 CSI 的重建质量,重建质量至少提升 24.9%。

2. 系统模型

本文研究 IRS 辅助下的 MIMO FDD 通信系统,它采用 OFDM 作为多载波的传输方案,系统模型如图 1 所示。

在该系统中,配置有 N_i 个反射单元的 IRS 辅助有 N_t 根天线的 BS 与有 N_r 根天线的 UE 进行通信,而 OFDM 的子载波数量则设置为 N_c 。那么 UE 处第 m 根天线在第 c 个子载波接收到的信号 $y_{m,c}$ 可以表达为:

$$y_{m,c} = (H_{\text{BI},m,c} \varphi H_{\text{IU},m,c}^H + H_{\text{BU},m,c}) v_c x_c + \delta_{m,c} \quad (1)$$

式中: $H_{\text{BI},m,c} \in C^{N_t \times N_i}$ 为 BS 与 IRS 之间的信道矩阵; $\varphi = \text{diag}(a\varphi_1, \dots, a\varphi_{N_i}) \in C^{N_i \times N_i}$ 为 IRS 的反射对角矩阵,其中 $a \in [0,1]$ 表示振幅系数, $\varphi_{n_i} \in [0, 2\pi]$, $n_i = 1, 2, \dots, N_i$ 表示相位系数; $H_{\text{IU},m,c} \in C^{N_t \times N_r}$ 为 BS 与 IRS 之间的信道矩阵; $H_{\text{BU},m,c} \in C^{N_t \times N_r}$ 为 BS 与 IRS 之间的信道矩阵; v_c 为第 N_c 个子载波的预编码向量; x_c 为下行链路中传输的数据符号; $\delta_{m,c}$ 为加性噪声;

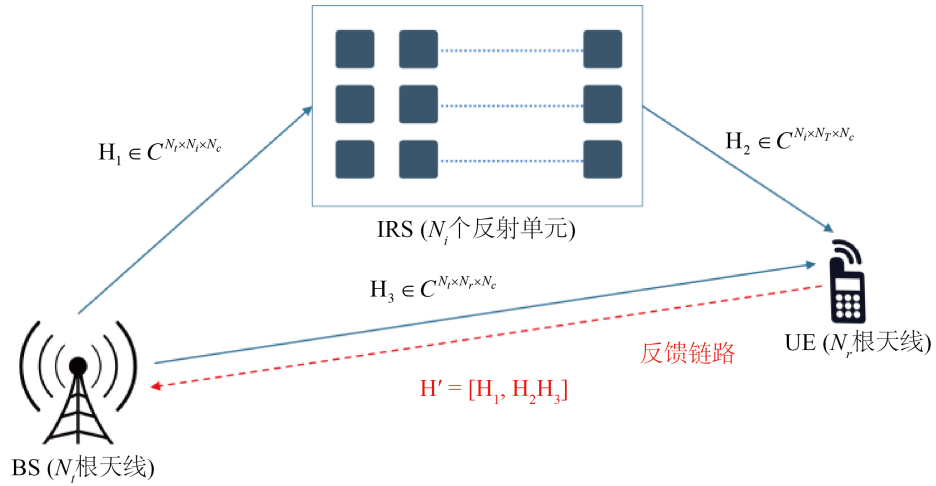


Figure 1. A IRS-assisted MIMO FDD communication system model
图1. IRS辅助下的MIMO FDD通信系统模型

BS 需要设计合理的预编码向量 v_c 才能消除用户间干扰进而提升通信质量。然而在 FDD 模式下，BS 需要获得精确的下行链路 CSI 才能对路预编码向量进行合理设计。在该系统中，完整维度的下行链路的 CSI H' 包括 BS 到 UE 信道的 CSI $H_1 \in C^{N_t \times N_r \times N_c}$ 、BS 到 IRS 信道的 CSI $H_2 \in C^{N_t \times N_r \times N_c}$ 、IRS 到 UE 信道的 CSI $H_3 \in C^{N_r \times N_r \times N_c}$ ，即 $H' = [H_1, H_2, H_3]$ 。完整的 CSI 数据如图 2 所示。

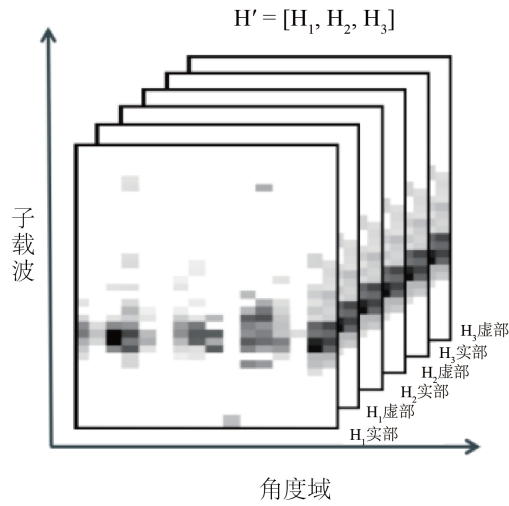


Figure 2. Schematic of the complete CSI data
图2. 完整的CSI数据示意图

因为时延扩展的有限性，CSI 中会有大量的 0 值，所以放入 IARNet 的 CSI 矩阵可以将 CSI 截断并只保留前 \tilde{N}_c 行的有效数据，阶段后的 CSI 可以表示为 $H = [H'_1, H'_2, H'_3]$ ；其中 $H'_1 \in C^{N_t \times N_r \times \tilde{N}_c}$ ； $H'_2 \in C^{N_t \times N_r \times \tilde{N}_c}$ ； $H'_3 \in C^{N_r \times N_r \times \tilde{N}_c}$ ；数据总量为 $N_t N_r \tilde{N}_c + N_r N_r \tilde{N}_c + N_t N_r \tilde{N}_c$ 个复数，其中 $N_t \times N_r$ 、 $N_r \times N_r$ 以及 $N_t \times N_r$ 分别表示为 H'_1 、 H'_2 以及 H'_3 的角度域长度， \tilde{N}_c 表示为阶段后的子载波数。将 CSI 截断后，IRS 辅助的 MIMO 系统的反馈量仍然是普通的 MIMO 系统反馈量的数倍。因此如果将直接将下行链路 CSI 进行反馈的话反馈开销将会是十分巨大的，系统难以承载，所以使用深度学习网络对数据进行压缩处理。

本文主要研究深度学习在 IRS 辅助下通信系统 CSI 压缩与恢复, 因此假设 UE 已经获得了反馈的所有 CSI, 即忽略信道估计误差, 同时假设 BS 也能完整地接收到 UE 反馈的所有信息。

本系统在 UE 和 BS 分别设置了编码器和译码器, UE 处的编码器可以将原始的 J 维的 CSI \mathbf{H} 压缩成 K 维向量 \mathbf{c} , 压缩比可以表示为: $\eta = K/J$, 其中 ($K < J$)。编码过程可以表示为:

$$\mathbf{c} = f_{en}(\mathbf{H}, \Theta_{en}) \quad (2)$$

式中: f_{en} 表示编码器; Θ_{en} 表示编码器的深度学习网络参数。

当 BS 收到完整的向量 \mathbf{c} 后, BS 端的译码器可以将其解码并重建 CSI, 该过程表示为:

$$\hat{\mathbf{H}} = f_{de}(\mathbf{c}, \Theta_{de}) \quad (3)$$

式中: $\hat{\mathbf{H}}$ 表示重建后的 CSI 矩阵; f_{de} 表示译码器; Θ_{de} 表示译码器的深度学习网络参数。

为了评估本系统的重建质量, 本系统使用归一化均方误差(NMSE, Normalized Mean Squared Error)作为判断标准, NMSE 可以评估原始 CSI 与重建后的 CSI 之间的误差, 这个值越小表示系统重建质量越佳, 因此本文主要的目标是通过优化系统模型最小化该值。其中 NMSE 定义为:

$$NMSE = \mathbf{E} \left\{ \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_2^2}{\|\mathbf{H}\|_2^2} \right\} \quad (4)$$

式中: $\|\cdot\|_2$ 表示 L2 范数。

3. IARNet 的结构

本文所提出的 IARNet 深度学习网络结构如图 3 所示, 其由 UE 处的编码器和 BS 处的解码器构成。IARNet 的输入是信道 CSI \mathbf{H} , \mathbf{H} 的具体尺寸为 $128 \times 16 \times 6$, 其中 128 表示角度, 16 表示截断后的子载波数, 6 表示三个 CSI 的虚数和实数。IARNet 的输出是重建后的 CSI $\hat{\mathbf{H}}$, 尺寸和 \mathbf{H} 一致。

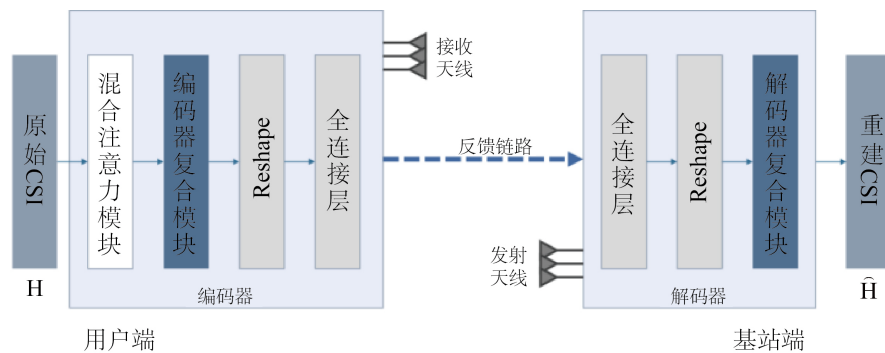


Figure 3. Architecture of IARNet

图 3. IARNet 的架构

在编码器侧。模型首先将 \mathbf{H} 放入混合注意力模块进行特征提取。该模块同时提取了 CSI 在空间和通道上的数据, 经过该模块后深度学习网络可以更专注于信息量大的数据, 提升数据权重, 反之降低信息量小的数据权重, 加强了特征表达能力。然后, 将数据放入编码器复合模块处理。该模块主要是借鉴了 Inception 网络的思想, 将多个尺寸卷积后的结果直接进行拼接处理, 这可以让拼接后的结果具有多维度特征信息的特点。此外该模块还引入了分组卷积的处理方法以降低训练参数。接着再利用混合注意力模块进行特征提取, 进一步加强了特征表达能力。接着, 将数据 Reshape 成长度为 $128 \times 16 \times 6 = 12288$ 的一

维向量并将其输入到神经元数量为 $8704 \times \eta$ 的全连接层神经网络中进行压缩，其中 8704 表示 \mathbf{H} 的有效数据，为了对齐6个通道的尺寸， \mathbf{H}_3 需要补0处理，具体见第五章仿真部分。最后，通过反馈链路将压缩后的数据发送给BS端的解码器。

在解码器侧。模型首先将长度为 $8704 \times \eta$ 的一维向量 Reshape 成 \mathbf{H} 尺寸大小的矩阵。然后，用解码器复合模块将数据进行处理。该模块与编码器复合模块类似，但是更轻量化。其中该模块引入了比例残差的设计，相比一般的残差网络，网络性能更佳。最后，解码器输出重建后的CSI矩阵 $\hat{\mathbf{H}}$ 。

混合注意力模块主要由通道注意力模块和空间注意力模块两部分组成。通道注意力模块中，首先通过两种并行的平面平均池化和平面最大池化的处理，接着再将它们分别送入MLP神经网络之中，最后通过Sigmoid函数输出结果。通过通道注意力模块处理让模型关注到通道之间的关系并自动学习到不同通道特征的重要程度。空间注意力模块中，将通道注意力模块输出的数据作为输入，首先通过两个并行的通道平均池化和通道最大池化处理并将两个处理结果进行通道拼接，然后通过卷积操作将通道降为一维，最后进行Sigmoid激活函数处理和残差处理。通过通道注意力模块处理，模型会关注到同一通道上不同数据位置的关系并自动学习到不同空间特征的重要程度。混合注意力模块的结构如图4所示。混合注意力模块的输入为 $128 \times 16 \times 6$ 的特征，其中6表示为输入特征通道数；16表示特征的高度即子载波数；128表示特征的宽度即角度； 1×6 表示MLP神经网络的神经元个数； \otimes 表示矩阵乘法； \oplus 表示矩阵加法；

混合注意力模块同时关注到了通道信息和空间信息的重要关系，增大了有效通道和空间的权重，减少了无效通道和空间的权重，进而提升了网络性能。同时混合注意力模块还可以很轻易地集成到现有的深度学习网络架构上去，输入与输出的特征尺寸完全一致，这让网络的配置更加简单和便捷。

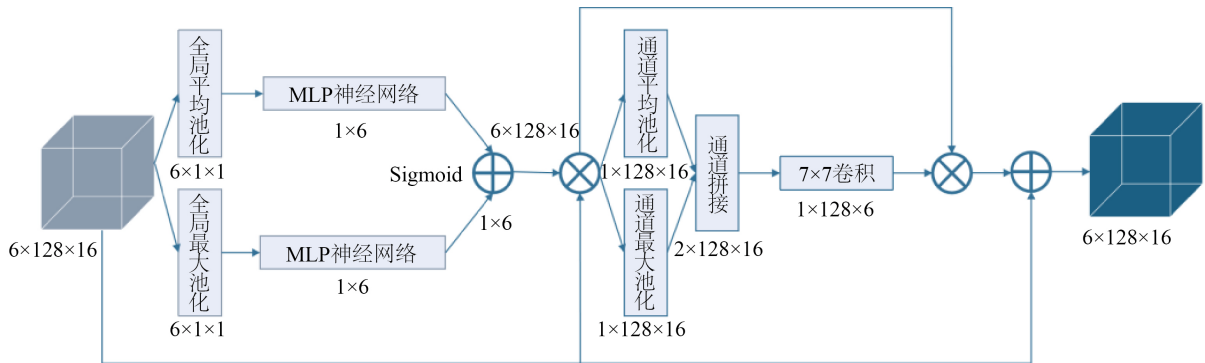


Figure 4. Architecture of hybrid attention module
图4. 混合注意力模块的架构

复合网络模块的结构

编码器复合模块及解码器复合模块如图5所示，其中每个方框附近的小数字表示此步的通道数(卷积核个数)，在数据处理的过程中CSI的长宽不变，即保持为 128×16 尺寸的数据，通道数随着卷积核变化。

文献[15]中已经证明基于 3×3 卷积和残差网络的CsiNet在信道压缩中的应用是有效的。但CsiNet是一种固定卷积尺寸的网络，固定尺寸卷积处理下的网络并不能很好地同时提取稀疏矩阵和密集矩阵的特征。如果想较好地同时提取稀疏矩阵和密集矩阵的特征，就需要考虑同时用不同尺寸卷积处理CSI。小尺寸的卷积处理(如 3×3 卷积)可以提取CSI更加精细的特征，在处理密集CSI的时候小尺寸的卷积有更好的效果。大尺寸的卷积处理(如 9×9 卷积)可以提供更大的感受视野，在处理稀疏CSI的时候这种卷积有更好的效果。因此在编码器复合模块和解码器复合模块中大量使用了多支路并行的多尺寸卷积处理，然后将不同支路上的结果在通道上直接拼接起来，这样可以将不同尺寸卷积处理下的结果进行多卷积特

征融合, 让输出拥有更加丰富的特征。特别是在编码器复合模块中, 为了更好地提取原始 CSI 的特征, 编码器复合模块在中同时使用了 3×3 卷积、 5×5 卷积、 7×7 卷积和残差的并行处理, 这将极大地丰富了输出特征。此外, 每个卷积模块进行卷积处理前都进行了一次批归一化处理。

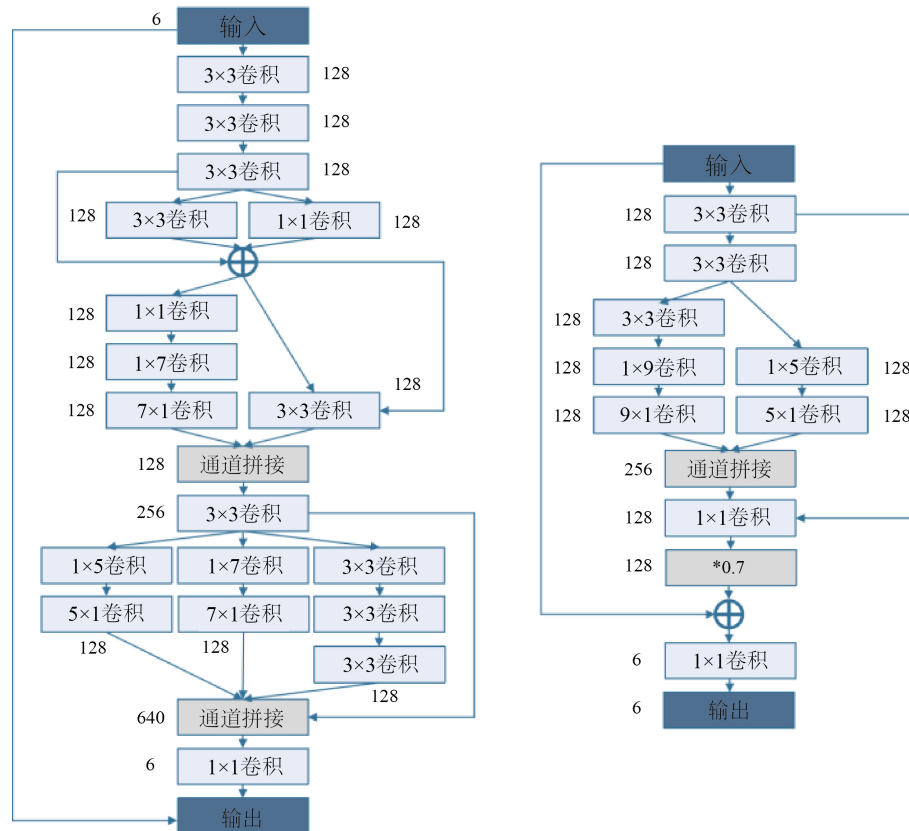


Figure 5. Composite module of encoder (left) and decoder (right)
图5. 编码器复合模块(左)及解码器复合模块(右)

为减少复合网络模块中的参数数量和运算复杂度, 模型还引入了分组卷积的处理方法。在分组卷积中通过以下步骤来分解 $M \times M$ 卷积。首先, 设置组数 g 并将原来的特征通道数平均分解成 g 组, 每个小组的特征通道数为原来的 $1/g$, 每一个小组的卷积核个数也为原来的 $1/g$, 保持长宽不变。然后, 每个小组进行 $M \times M$ 的卷积计算。最后, 将每个小组的结果进行通道拼接, 最终输出的特征尺寸不变。而且由于将标准的 $M \times M$ 卷积拆分成了更小规模的子运算, 这可以大幅度降低运算复杂度, 减少设备的运行要求。

此外, 为了解决梯度消失的问题, 提高系统的性能, 复合网络模块还加入了大量的残差网络。特别是在解码器复合模块的末端还利用了比例残差的网络, 即将主干的输出乘以一个小于 1 的系数(本网络采用 0.7), 调整主干的输出比例。经此设计, IARNet 的性能有进一步的提升。

4. 学习率及其调整策略

在深度学习模型的训练过程中, 模型训练方案对模型的最终呈现效果有着决定性的影响。在一些基于深度学习的信道压缩反馈研究中, 其模型训练方案是相对简要的, 没有针对特定的系统模型进一步优化模型训练方案。如在 CsiNet 和 DS-NLCsiNet 的文章中, 模型的 batch size、epochs 和初始学习率分别

直接设置为 200、1000 和 0.001，也没有设置学习率调整策略。这些文章都省略了对模型训练方案的介绍，特别是学习率及其调整策略上，而这恰恰是十分重要的。

如果学习率设置过高，虽然系统训练会加快，但是在采用梯度下降算法来寻找全局最优解的过程中，损失函数将不会收敛至全局最小值附近。如果学习率设置过小，虽然网络可以找到全局最优解，但是这会花费大量的训练时间并且很容易陷入局部最优解。另外学习率调整策略对系统训练也有很大的影响，一般采用的策略有固定法、步衰减法(Step Decay)法以及余弦衰减法(Cosine Decay)等。固定法需要多次试验才能找到较好的学习率，而且网络也很容易陷入局部最优解。衰减类的方法可以在较高的学习率上加速网络训练，然后在低学习率上寻找到全局最优解[19]。在 IARNet 的模型训练方案上采用基于热身法的余弦衰减学习率调整策略。由于深度学习网络在刚开始训练的时候非常不稳定，所以我们需要将初始的学习率设置得很低，这可以让深度学习网络缓慢地趋向于稳定。当网络趋于稳定的时候再升高学习率，这可以让网络可以快速地收敛，这个过程就称之为热身，热身完之后将采用余弦衰减的方法减少学习率。这样就可以让整个训练过程变得平稳、快速，同时也提高了网络性能。本方案的学习率调整策略可以表达为：

$$\begin{cases} \frac{1}{20}lr_s epoch & 0 < epoch < epoch' \\ \frac{1}{2}lr_s \left(1 + \cos\left(\frac{epoch - epoch'}{epochs}\right) \right) & epoch' < epoch < epochs \end{cases} \quad (5)$$

式中 lr_s 表示初始学习率； $epochs$ 表示训练的总轮次；前 $epoch'$ 个轮次的学习率采用热身的调整策略， $epoch'$ 轮次后的学习率采用余弦衰退的调整策略。学习率整个调整过程如图 6 所示。

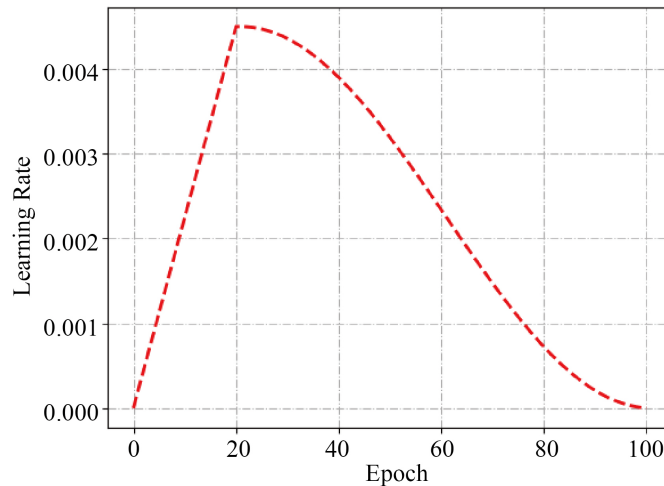


Figure 6. Learning rate adjustment strategy of cosine decay based on the warm-up method

图6. 基于热身法的余弦衰减学习率调整策略

5. 仿真结果

仿真过程中 BS 的天线数设置为 4，IRS 的反射单元数设置为 32，UE 的天线数设置为 4，子载波数设置为 128。仿真采用 COST 2100 信道模型[20]在中心频率为 5.3GHz 的频带的室内场景下生成数据集。其中 BS 到 IRS 之间信道用发射天线数为 4、接收天线为 32 的信道替代，IRS 到 UE 之间信道用发射天线

数为 32、接收天线为 4 的信道模型替代。然后将 BS 到 IRS、IRS 到 UE 以及 BS 到 UE 的 CSI 在通道上拼接起来。因为时延扩展的有限性，CSI 中会有大量的 0 值，所以将 CSI 截断并只保留前 16 行的有效数据。然后将截断后 CSI 中的 6 个通道的特征尺寸补齐，即 BS 到 UE 的 CSI 由 16×16 用 0 值扩展至 128×16 。6 个通道保持 128×16 的尺寸，其中有效数据为 8704，其余数据为对齐 CSI 补的 0 值，压缩比是按照有效数据 8704 与压缩后的数据量计算。补齐后 CSI 的尺寸为： $128 \times 16 \times 6$ 。其中 128 表示为角度；16 为子载波；6 分别为 BS 到 IRS 的 CSI、IRS 到 UE 的 CSI 以及 BS 到 UE 的 CSI 的虚数和实数。

使用 COST 2100 模型生成 10 万个数据集，然后按照 4:1 的比例将数据集分成训练集和测试集。模型训练时，采用均方误差(MSE, Mean Squared Error)作为系统的损失函数，使用 Adam 算法[21]更新参数。batch size 设置为 150，epoch 设置为 100，初始学习率设置为 0.0045 并使用基于热身法的模型训练方法，*epoch'* 设置为 20。

本文比较了 IARNet 与 CRNet 与 CsiNet 在不同压缩比下的性能表现，结果如表 1 所示，加粗表示为该压缩比下的最佳性能表现。仿真结果表明，IARNet 在大数据量 CSI 压缩场景下对比其他基于深度学习的 CSI 重建算法有更好的性能表现，即使是在 1/32 的压缩比下 IARNet 仍能将 CSI 较好地重建起来，这主要得益于 IARNet 采用了多卷积特征融合、混合注意力机制以及比例残差等方法并进行了联合优化，在优化过程中尽量避免不必要的计算开销，这让网络保持性能的同时也更轻量化。

Table 1. Comparison of NMSE performance of IARNet with CRNet and CsiNet
表1. IARNet与CRNet和CsiNet算法的NMSE性能比较

压缩比	深度学习网络	NMSE
1/2	CsiNet	1.0889
	CRNet	0.5529
	IARNet	0.0273
1/4	CsiNet	1.0894
	CRNet	0.5537
	IARNet	0.0281
1/8	CsiNet	1.0902
	CRNet	0.5625
	IARNet	0.0680
1/16	CsiNet	1.0910
	CRNet	0.6172
	IARNet	0.1740
1/32	CsiNet	1.0919
	CRNet	0.6578
	IARNet	0.3841
1/64	CsiNet	1.2300
	CRNet	0.8266
	IARNet	0.6076

图 7 是 IARNet、CRNet 和 CsiNet 在压缩比为 1/8 下系统的 NMSE 随着 epoch 变化的曲线。由图 7 所示，在训练过程中，CsiNet 的 NMSE 始终保持较高的水平，NMSE 从最初的 25.112 收敛至 1.0902，这表明数据集超出了系统的学习能力，CsiNet 无法学习和重建大数据量的 CSI。CRNet 的 NMSE 从最初的

3.11 可以收敛至 0.5625，但是 33 个 epoch 后系统的 NMSE 基本不变，这表明 CRNet 可以学习和重建部分 CSI 的数据，但是重建质量不佳。IARNet 的 NMSE 从最初的 0.9912 可以收敛至 0.0680，NMSE 曲线在整个训练过程中都趋于下降，在 90 个 epoch 后逐渐平稳，这表明 IARNet 可以很好地学习和重建大部分的 CSI 数据，重建质量佳。

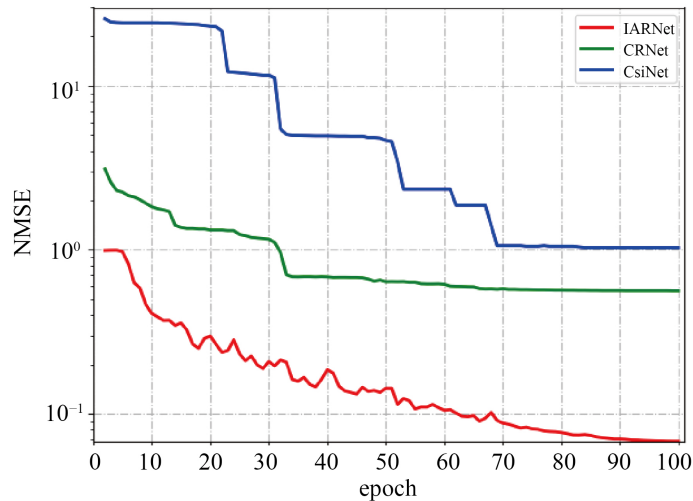


Figure 7. NMSE variation curves of three networks at 1/8 compression ratio during the training process

图7. 训练过程中三种网络在1/8压缩比下的NMSE变化曲线

图8是IARNet在压缩比为1/8下系统的测试集损失函数和训练集的损失函数随着epoch变化的曲线。由图7可见，测试集损失函数和训练集的损失函数随着epoch增加而逐渐减少，测试集损失函数在训练集损失函数附近波动。训练结束后，训练集的损失函数可以收敛至0.03附近，训练集的损失函数可以收敛至0.05附近，这说明IARNet有轻微过拟合，但是整体上可以忽略不计。

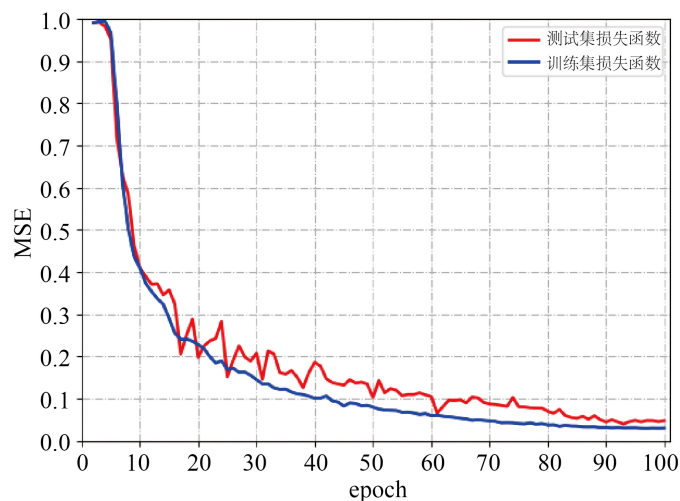


Figure 8. Comparison of the loss functions of the training and test sets

图8. 训练集和测试集的损失函数对比

此外本文还比较了四种学习率调整策略配合其最佳初始学习率在压缩比为 1/8 情况下对 IARNet 性能的影响。第一种是本方案采用基于热身法的余弦衰减方案，最佳初始学习率为 0.0045。第二种方案是固定法，最佳初始学习率为：0.0002。第三种方案是步衰减法，最佳初始学习率为：0.001。第四种方案是余弦衰减法，最佳初始学习率为 0.0003。NMSE 随着训练过程推进的仿真变化曲线如图 9 所示。由图 9 仿真结果表明，相比于传统的学习率调整策略，本方案采用的方法可以让系统有更好的重建性能，NMSE 收敛至更低的水平，可达 0.0683。而固定法、步衰减法和余弦衰减法训练后的系统分别收敛至 0.1572、0.0903 和 0.1822。相比传统的方法基于热身法学习率调整策略的重建质量至少提升 24.9%。

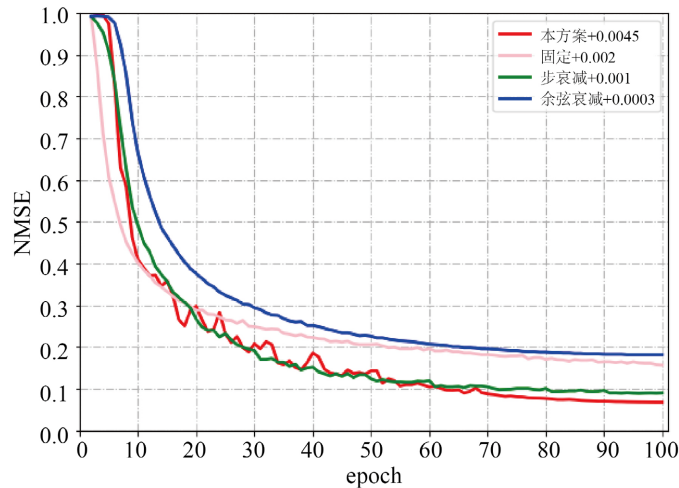


Figure 9. NMSE variation curves of four learning rate adjustment strategies during training

图9. 训练过程中四种学习率调整策略的NMSE变化曲线

另外表 2 比较了 IARNet 与 CRNet 和 CsiNet 在不同压缩比下的参数量，M 表示百万。IARNet 除了在 1/64 压缩比下参数量高于 CRNet 和 CsiNet 以外，其他压缩比下的参数量均小于 IARNet 和 CRNet，因此 IARNet 可以较易地部署在各类设备里面，节省设备存储空间。

Table 2. Comparison of the number of parameters of IARNet with CRNet and CsiNet at different compression ratios

表2. IARNet与CRNet和CsiNet在不同压缩比下的参数量比较

	压缩比	1/2	1/4	1/8	1/16	1/32	1/64
参数量	CsiNet	67.22M	33.63M	16.82M	16.82M	4.26M	2.21M
	CRNet	67.13M	33.57M	16.79M	16.79M	4.21M	2.11M
	IARNet	40.95M	24.17M	15.78M	15.78M	3.91M	2.86M

6. 结论

本文针对 IRS 辅助的通信系统中反馈开销更加巨大的问题提出了一种新的深度神经网络 IARNet。该网络在传统的 Inception 网络结构上结合了多卷积特征融合、混合注意力机制以及残差等子模块，这种混合结构可以有效地将大数据量的 CSI 进行压缩重建。计算机仿真结果显示 IARNet 在 IRS 通信系统中对大数据量的 CSI 进行压缩重建有更好的表现，并且基于热身法的模型训练方案优于传统的固定法和衰减类法。

参考文献

- [1] 朱政宇, 王梓暄, 徐金雷, 王忠勇, 王宁, 郝万明. 智能反射面辅助的未来无线通信: 现状与展望[J/OL]. 航空学报, 1-16. <http://kns.cnki.net/kcms/detail/11.1929.v.20210225.1111.008.html>, 2021-02-25.
- [2] Cui, M., Zhang, G. and Zhang, R. (2019) Secure Wireless Communication via Intelligent Reflecting Surface. *IEEE Wireless Communications Letters*, **8**, 1410-1414. <https://doi.org/10.1109/LWC.2019.2919685>
- [3] Zhang, D., Wu, Q., Cui, M., Zhang, G. and Niyato, D. (2021) Throughput Maximization for IRS-Assisted Wireless Powered Hybrid NOMA and TDMA. *IEEE Wireless Communications Letters*, **10**, 1944-1948. <https://doi.org/10.1109/LWC.2021.3087495>
- [4] Jiang, M., Li, Y., Zhang, G. and Cui, M. (2021) Achievable Rate Region Maximization in Intelligent Reflecting Surfaces-Assisted Interference Channel. *IEEE Transactions on Vehicular Technology*, **70**, 13406-13412. <https://doi.org/10.1109/TVT.2021.3120308>
- [5] Li, Y., Jiang, M., Zhang, G. and Cui, M. (2021) Achievable Rate Maximization for Intelligent Reflecting Surface-Assisted Orbital Angular Momentum-Based Communication Systems. *IEEE Transactions on Vehicular Technology*, **70**, 7277-7282. <https://doi.org/10.1109/TVT.2021.3089021>
- [6] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.
- [7] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2015) Rethinking the Inception Architecture for Computer Vision. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 2818-2826. <https://doi.org/10.1109/CVPR.2016.308>
- [9] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2016) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv preprint arXiv:1602.07261.
- [10] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- [11] Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H., Wu, H. and Wang, H. (2019) ERNIE 2.0: A Continual Pre-training Framework for Language Understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 8968-8975. <https://doi.org/10.1609/aaai.v34i05.6428>
- [12] Li, M., Zuo, W., Gu, S., Zhao, D. and Zhang, D. (2017) Learning Convolutional Networks for Content-weighted Image Compression. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 3214-3223. <https://doi.org/10.1109/CVPR.2018.00339>
- [13] Kim, D., Kim, M., Kwon, G. and Kim, D.-S. (2019) Progressive Face Super-Resolution via Attention to Facial Landmark. arXiv preprint arXiv:1908.08239.
- [14] 于恒, 梅红岩, 许晓明, 贾慧萍. 基于深度学习的图像压缩算法研究综述[J]. 计算机工程与应用, 2020, 56(15): 15-23.
- [15] Wen, C., Shih, W. and Jin, S. (2018) Deep Learning for Massive MIMO CSI Feedback. *IEEE Wireless Communications Letters*, **7**, 748-751. <https://doi.org/10.1109/LWC.2018.2818160>
- [16] Lu, Z., Wang, J. and Song, J. (2020) Multi-Resolution CSI Feedback with Deep Learning in Massive MIMO System. 2020 *IEEE International Conference on Communications*, Dublin, 7-11 June 2020, 1-6. <https://doi.org/10.1109/ICC40277.2020.9149229>
- [17] Yu, X., Li, X., Wu, H. and Bai, Y. (2020) DS-NLCsiNet: Exploiting Non-Local Neural Networks for Massive MIMO CSI Feedback. *IEEE Communications Letters*, **24**, 2790-2794. <https://doi.org/10.1109/LCOMM.2020.3019653>
- [18] Guo, J., *et al.* (2020) AI Enabled Wireless Communications with Real Channel Measurements: Channel Feedback. *Journal of Communications and Information Networks*, **5**, 310-317. <https://doi.org/10.23919/JCIN.2020.9200894>
- [19] 蒋文斌, 彭晶, 叶阁焰. 深度学习自适应学习率算法研究[J]. 华中科技大学学报(自然科学版), 2019, 47(5): 79-83.
- [20] Liu, L., *et al.* (2012) The COST 2100 MIMO channel model: Channel feedback. *IEEE Wireless Communications*, **19**, 92-99. <https://doi.org/10.1109/MWC.2012.6393523>
- [21] Kingma, D.P. and Ba, J. (2017) Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.