arXiv:2303.00986v2 [eess.SP] 9 Nov 2023

少花钱，多获益：一种面向低密度导频大规模 MIMO 系统的双注意力信道估计网络

Binggui Zhou、Xi Yang、Shaodan Ma、Feifei Gao 和杨光华

摘要—为了充分发挥大规模多输入多输出（MIMO）系统的潜在优势，需要通过信道估计获得准确的信道状态信息（CSI）。然而，由于无线传播环境复杂且天线阵列规模庞大，大规模MIMO系统的精确信道估计极具挑战性，并且需要巨大的训练开销。为获得足够准确的 CSI，需要消耗大量时频资源，从而严重降低系统的频谱效率和能效。本文提出一种双注意力信道估计网络（DACEN），通过时间注意力模块和空间注意力模块联合学习大规模 MIMO 信道的时空域特征，从而利用低密度导频实现精确的信道估计。为进一步提高估计精度，我们提出一种参数-实例迁移学习方法，将训练数据集采集期间预先通过高密度导频学习到的信道知识迁移过来。实验结果表明，在不同导频密度设置和信噪比下，所提出的 DACEN 方法相比现有方法具有更优的信道估计性能。此外，采用所提出的参数-实例迁移学习方法后，DACEN 方法还能进一步提升性能，从而进一步证明了所提方法的有效性和优越性。

索引词—低开销信道估计、大规模 MIMO、注意力机制、迁移学习、深度学习

一、引言

MASSIVE 多输入多输出（MIMO）是第五代（5G）无线通信系统的基础技术，并将在第六代（6G） [1]中继续发挥重要作用。尽管大规模 MIMO 系统的大规模天线阵列能够提供较大的空间自由度、分集增益和阵列增益，从而大幅提升无线系统的频谱效率和能量效率 [2]，但要充分发挥这些优势，就需要通过信道估计准确获取信道状态信息（CSI）。

大规模 MIMO 系统的信道估计可以通过在发射信号中插入导频符号，再基于接收信号估计信道来实现。只要能够承担导频训练开销，并为导频符号分配足够的时频资源，就可以获得准确的信道估计值。[1] 信道估计方法大致可分为传统估计方法和基于深度学习的方法。传统估计方法 [3]–[7] 主要基于特定信道模型或信道分布的假设来估计信道。因此，传统方法需要具备合适的信道模型，或掌握有关信道分布的先验信息。随着天线阵列规模不断增大，由于信道维度扩大，传统方法的导频训练开销和计算复杂度也显著增加。为了支持实际的大规模 MIMO 应用，在降低导频开销的同时实现准确的信道估计极为重要，近期也因此受到广泛关注。现有的低开销信道估计方法主要基于大规模 MIMO 信道的稀疏性。其基本原理是利用信道稀疏性，将信道状态信息估计转化为对数量显著减少的信道参数（例如到达角 (AoA) 和离开角 (AoD)）的估计，从而实现低开销信道估计。例如，许多研究采用压缩感知

本工作部分得到国家自然科学基金委员会资助，项目编号为62171201、62261160650和62301221；部分得到广东省重大人才计划资助，项目编号为2019QN01S103；部分得到上海市浦江人才计划资助，项目编号为22PJ1403100；部分得到澳门特别行政区科学技术发展基金资助，项目编号为0087/2022/AFJ和SKL-IOTSC(UM)-2021-2023；部分得到澳门大学研究委员会资助，项目编号为MYRG2020-00095-FST；部分得到北京市自然科学基金资助，项目编号为L222002。（通讯作者：Shaodan Ma；Xi Yang。）

Binggui Zhou现任职于暨南大学智能系统科学与工程学院，中国珠海519070；同时任职于智慧城市物联网国家重点实验室以及澳门大学电气与计算机工程系，中国澳门999078（电子邮箱：binggui.zhou@connect.um.edu.mo）。

Xi Yang 任职于华东师范大学通信与电子工程学院、上海市多维信息处理重点实验室，中国上海 200241（电子邮箱：xyang@cee.ecnu.edu.cn）。

Shaodan Ma 任职于智慧城市物联网国家重点实验室及澳门大学电气与计算机工程系，中国澳门 999078（电子邮箱：shaodanma@um.edu.mo）。

Feifei Gao 任职于清华大学人工智能研究院、THUAI、智能技术与系统国家重点实验室、清华大学北京国家信息科学技术研究中心（BNRist）及自动化系，中国北京 100084（电子邮箱：feifeigao@ieee.org）。

杨光华任职于暨南大学智能系统科学与工程学院、粤港澳大湾区与“一带一路”国际智慧物流联合研究中心，中国珠海 519070（电子邮箱：ghyang@jnu.edu.cn）。

本工作的源代码可在此处获取https://github.com/bgzhou/DACEN/。

[1]尽管信道状态信息参考信号（CSI-RS）和解调参考信号（DM-RS）都可视为导频，但本文主要关注 CSI-RS。

© 2023年 IEEE。允许将本材料用于个人用途。对于任何其他用途，无论采用当前还是未来的媒体形式，包括为广告或促销目的重印或再版本材料、创作新的汇编作品、转售或再分发至服务器或列表，或在其他作品中重复使用本作品中任何受版权保护的组件，均须获得 IEEE 的许可。

(压缩感知，CS)，例如正交匹配追踪 (OMP) 算法[8], [9]，来估计 AoA 和 AoD，以实现低开销信道估计。文献 [10] 提出了一种混合信道估计算法，结合压缩感知技术和稀疏贝叶斯学习 (SBL) 来利用信道稀疏性。通过考虑大规模 MIMO 信道的分层稀疏性， [11] 提出了一种适用于大规模 MIMO 系统的低开销信道估计方案，并对导频开销需求进行了理论分析。然而，由于无线传播环境复杂多变，实际大规模 MIMO 系统中的无线信道难以通过解析方式建模。这些挑战阻碍了上述传统信道估计方法的应用，尤其是在部署大规模天线阵列时。

近年来，深度学习已应用于无线通信系统，以提升性能 [12]–[14]。例如，研究人员提出了各种深度神经网络，如全连接深度神经网络（FC-DNN） [15], 和卷积神经网络（CNN） [16]–[18], ，利用大量训练数据学习信道特性。此外，研究人员还提出了多种变体，如复数去噪卷积神经网络（CDnCNN） [19],、卷积盲去噪网络（CBDNet） [20], 、基于CNN的深度残差网络（CDRN） [21],等，以提高信道估计的准确性。基于深度学习的信道估计方法取得的成功表明，深度神经网络能够学习实际系统中复杂的信道特性。然而，随着天线阵列规模不断扩大，现有的基于深度学习的方法面临着新的挑战。首先，为了在有限的时频资源下保证系统的频谱效率，需要设计具有卓越学习能力的深度神经网络架构，使其能够仅利用少量导频进行学习。因此，这种能力需要采用更复杂的深度神经网络结构来挖掘信道特性。其次，天线规模和带宽的扩大进一步增加了现有深度神经网络架构的计算复杂度。在此前的基于深度学习的方法中，由于采用了大量天线和宽带宽，大规模MIMO信道通常以空频域表示 [16], [22],[23]。因此，提取这些多域中的潜在信道特征，对于获得准确的信道估计值至关重要。为此，现有的基于深度学习的方法通常通过一个深度神经网络模块提取多域特征。例如，[16], [17], [23] 提出使用一个CNN模块提取空频域特征。然而，不同域中的不同特征实际上对应着各域中不同的相关性，而这些相关性通常彼此无关。一个深度神经网络模块无法有效提取多域特征。因此，亟需一种由多个深度神经网络模块构成的先进多域特征提取器，以联合提取多域特征。

此外，通常来说，导频越多，获取的信道知识越丰富，信道估计性能也越好。因此，从高密度导频中（例如，在同一相干间隔内累积低密度导频，或仅在训练数据集采集期间分配特定时隙进行高密度导频训练）学习到的知识，如果能够迁移，将有助于提升低开销信道估计的估计性能。幸运的是，迁移学习（又称知识迁移）能够实现知识迁移。具体而言，迁移学习已成功应用于时变环境下提升信道估计精度，以克服快速信道波动 [24], [25]。因此，将迁移学习应用于低开销信道估计以提升性能，具有良好前景。

受此启发，我们首先提出一种双注意力信道估计网络（DACEN），通过联合学习大规模 MIMO 信道的多域特征，并利用特定的注意力模块学习领域特定特征，实现低开销信道估计。由于延迟路径的数量通常远少于子载波的数量，我们在 DACEN 中采用空时域信道表示，以降低其内存使用量和模型复杂度。具体而言，DACEN 由两个注意力模块组成，即时间注意力模块（TAM）和空间注意力模块（SAM）。TAM 旨在学习不同延迟路径之间的时间相关性，而 SAM 旨在学习不同天线之间 CSI 的相关性。借助 SAM 区分高度相关的天线，并据此主要关注数量有限的有效天线，可以在保持估计精度的同时显著降低计算复杂度。基于 DACEN，我们还提出一种参数-实例迁移学习方法，通过将在模型训练阶段从高密度导频中学习到的信道知识迁移至低密度导频场景，进一步提高估计精度。

本文的主要贡献总结如下：

1）我们提出了一种双注意力信道估计网络，通过低密度导频实现大规模 MIMO 系统的精确信道估计。通过利用 TAM 和 SAM 联合学习大规模 MIMO 信道的时空域特征，所提出的 DACEN 能够借助低密度导频获得精确的信道估计值。此外，通过利用大规模 MIMO 信道的空间相关性，SAM 的时间复杂度和空间复杂度远低于广泛采用的卷积神经网络。TAM 的有效性也验证了大规模 MIMO 系统的无线信道可能不满足广义平稳不相关散射（WSSUS）假设，并且具有不同延迟的多条路径彼此相关，这有利于实现低开销信道估计。

2) 我们提出一种基于 DACEN 的参数-实例迁移学习方法，通过迁移模型训练阶段从高密度导频中学习到的信道知识，进一步提高信道估计精度。需要注意的是，高密度导频可以通过在一个相干间隔内累积低密度

导频获得，也可以在训练数据集收集期间分配专门时隙，仅用于高密度导频训练。通过使用获得的高密度导频对 DACEN 进行预训练（参数迁移），并在训练 DACEN 时对低密度导频进行重新加权（实例迁移），仅使用低密度导频也能准确估计 CSI。

3）传输的导频信号遵循第三代合作伙伴计划（3GPP）5G 技术规范[26] 中关于信道状态信息参考信号（CSI-RS）配置的规定，以获取下行信道估计。此外，由于 CSI-RS 由基站向系统中的所有用户广播，因此，所提出的方法可以直接应用于多用户场景。在各种低密度导频设置下进行的大量数值实验验证了所提出方法的有效性和优越性。

本文其余部分的组织结构如下。在第 II 节中，我们介绍系统模型并构建信道估计问题。在第 III 节中，提出新颖的 DACEN 架构。为了进一步提高信道估计精度，我们在第 IV 节提出参数-实例迁移学习方法。在第 V 节中，给出实验结果和计算复杂度比较。最后，我们在第 VI 节总结本文。

符号说明：带下划线的大写粗体字母、大写粗体字母和小写粗体字母分别表示张量、矩阵和向量。$\mathbf{A}^{(i)}$ 表示张量 A 的第 i 个切片，$\mathbf{A}^{(i,:)}$ 表示矩阵 A 的第 i 行，$\mathbf{A}^{(i,j)}$ 表示矩阵 A 的第 i 行第 j 列元素，$\mathbf{a}^{(i)}$ 表示向量 a 的第 i 个元素。Im{A} 和 Re{A} 分别表示 $\mathbf{A} \in \mathbb{C}$ 的虚部和实部。$\hat{\mathbf{A}}$ 表示 A 的估计值。E(·)、Var(·) 和 $\|\cdot\|_2$ 分别表示期望、方差和 L2 范数。 $\delta(\cdot)$ 表示狄拉克 δ 函数，log(·) 表示以 10 为底的对数运算。 $\otimes$、 $<\cdot,\cdot>$、 $[\cdot,...,\cdot]$ 和 R(·) 分别表示哈达玛积、内积、拼接操作和重排操作。为进一步提高本文的可读性，表 I 列出了与一些关键符号相关的说明。

II. 系统模型与问题建模

A. 系统模型

本文考虑一个大规模 MIMO 系统的下行链路，该系统在基站（BS）端配备 $N_T \gg 1$ 根天线，在用户端（UE）配备 N_R 根天线。[2] 采用具有 N_c 个子载波的正交频分复用（OFDM）调制。第 i 个子载波上的接收信号，即$\mathbf{Y}^i \in \mathbb{C}^{N_R \times N_T}$，可表示为：

$$\mathbf{Y}^i = \mathbf{H}_f^i \mathbf{X}^i + \mathbf{N}^i, i = 1, \ldots, N_c, \tag{1}$$

其中，$\mathbf{H}_f^i \in \mathbb{C}^{N_R \times N_T}$、$\mathbf{X}^i \in \mathbb{C}^{N_T \times N_T}$ 和 $\mathbf{N}^i \in \mathbb{C}^{N_R \times N_T}$ 分别表示空频域信道、由发射信号构成的对角矩阵，以及第 i 个子载波上的噪声。

[2]请注意，为了清楚地展示所提出的信道估计方法，我们以系统中的一个多天线用户设备为例。该方法可直接应用于单小区多用户场景。

通过拼接所有 N_c 子载波的空频域信道$\mathbf{H}_f^i$，得到的完整空频域信道$\mathbf{H}_f \in \mathbb{C}^{N_R \times (N_T \times N_c)}$ 可表示为：

$$\mathbf{H}_f = [\mathbf{H}_f^1, ..., \mathbf{H}_f^{N_c}], \tag{2}$$

可将其改写为张量形式：$\underline{\mathbf{H}}_f \in \mathbb{C}^{N_R \times N_T \times N_c}$。

B. 问题建模

假设基站发送 CSI-RS 符号（即导频符号），且用户设备处对应的接收导频信号为$\underline{\mathbf{Y}}_{rp} \in \mathbb{C}^{N_R \times N_T \times N_{rp}}$，$N_{rp}$ 是频域中接收导频的数量，则可使用基于 DNN 的信道估计器 F_f估计下行链路空频域信道：

$$\hat{\underline{\mathbf{H}}}_f^* = \arg\min_{\hat{\underline{\mathbf{H}}}_f} \mathcal{L}\left(\underline{\mathbf{H}}_f - \hat{\underline{\mathbf{H}}}_f\right),$$
$$\text{s.t. } [\mathrm{Im}\{\hat{\underline{\mathbf{H}}}_f\}, \mathrm{Re}\{\hat{\underline{\mathbf{H}}}_f\}] = F_f([\mathrm{Im}\{\underline{\mathbf{Y}}_{rp}\}, \mathrm{Re}\{\underline{\mathbf{Y}}_{rp}\}]), \tag{3}$$

其中，Im{·} 和 Re{·} 分别表示输入张量的虚部和实部。$\mathcal{L}$ 是一个损失函数，例如均方误差（MSE）损失函数，用于计算信道估计器的估计输出与期望输出之间的距离。

请注意， N_c 通常较大，例如， $N_c = 624$，这不可避免地会导致内存使用量和模型复杂度大幅增加，从而使式（3）的优化过程效率低下。为避免这一问题，一种可行的方法是仅从 $\mathbf{H}_f$ 中选取少量元素，构成简洁的空频域信道表示，然后基于该简洁信道表示训练网络。然而，这种操作会造成显著的信息损失，进而降低信道估计性能。

鉴于主要时延路径的数量通常远小于子载波数量 N_c（如图 1,[3] 所示，在示例信道快照中，主要时延路径少于 10 条，而对于 OFDM调制，我们在频域中有 N_c =624 个子载波），我们首先通过采用空时域信道$\mathbf{H}_t \in \mathbb{C}^{(N_R \times N_T) \times N_P}$ 进行频时转换，来应对较大的 N_c 所导致的高昂内存和模型复杂度：

$$\mathbf{H}_t = \mathrm{IFFT}_{N_{\mathrm{ifft}}}(\mathbf{H}_f')\mathbf{A}, \tag{4}$$

其中，$\mathrm{IFFT}_{N_{\mathrm{ifft}}}(\cdot)$表示沿频率维度进行的 N_{ifft}点快速傅里叶逆变换（IFFT），[4] $\mathbf{H}_f' \in \mathbb{C}^{(N_R \times N_T) \times N_c}$由$\mathbf{H}_f \in \mathbb{C}^{N_R \times N_T \times N_c}$重塑得到。$\mathbf{A} \in \mathbb{C}^{N_{\mathrm{ifft}} \times N_P}$ 是一个右截断矩阵，其定义如下：

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_{N_P} \\ \mathbf{0}_{N_{\mathrm{ifft}} - N_P, N_P} \end{bmatrix} \tag{5}$$

其中，$\mathbf{I}$ 为单位矩阵，$\mathbf{0}$ 为零矩阵， N_P 表示需要保留的$\mathrm{IFFT}_{N_{\mathrm{ifft}}}(\mathbf{H}_f')$条目数量。

[3]该时空信道表示的详细系统参数见第V节。

[4]与大量子载波带来的巨大内存使用量和模型复杂度相比，额外的逆快速傅里叶变换操作所需的计算复杂度可以忽略不计。

表 I：主要符号及其说明列表。

符号	说明	符号	说明
N_T	发射天线数	N_{RB}	系统的资源块总数
N_R	接收天线数	ρ	导频密度
N_c	子载波数	$\underline{\mathbf{Y}}_{pH} \in \mathbb{C}^{N_R \times N_T \times N_H}$	高密度接收导频信号（张量形式）
N_P	频时转换后需要保留的条目数量	$\underline{\mathbf{Y}}_{pL} \in \mathbb{C}^{N_R \times N_T \times N_L}$	低密度接收导频信号（张量形式）
$\underline{\mathbf{H}}_f \in \mathbb{C}^{N_R \times N_T \times N_c}$	空频域信道（张量形式）	s^{th}	余弦相似度阈值
$\underline{\mathbf{H}}_t \in \mathbb{C}^{N_R \times N_T \times N_P}$	空时域信道（张量形式）	$\mathbf{w} \in \mathbb{R}^{N_{DL_E}^{train} \times 1}$	所有训练样本的实例权重

$\mathbf{H}_t \in \mathbb{C}^{(N_R \times N_T) \times N_P}$ 可改写为如下张量形式$\underline{\mathbf{H}}_t \in \mathbb{C}^{N_R \times N_T \times N_P}$。由于路径损耗以及传播环境中散射体数量有限，发射信号穿过传播环境后，接收端只能观测到有限时延扩展内的少数传播路径。设信道采样频率为 f_s，假设最大路径时延为 N_P/f_s，这意味着前 N_P 项之后的元素可能具有非常小的路径增益。因此，式（ N_P 4）中的取足够大的值，就能保留信道的大部分乃至全部信息。[5] 还值得注意的是，由于时延扩展有限， N_P 与 N_c相比仍可足够小，从而有效降低内存使用量和模型复杂度。

然后，下行链路空频域信道估计问题可以转化为下行链路空时域信道估计问题，即:

$$\hat{\underline{\mathbf{H}}}_t^* = \arg\min_{\hat{\underline{\mathbf{H}}}_t} \mathcal{L}\left(\underline{\mathbf{H}}_t - \hat{\underline{\mathbf{H}}}_t\right),$$
$$\text{s.t.} \quad [\mathrm{Im}\{\hat{\underline{\mathbf{H}}}_t\}, \mathrm{Re}\{\hat{\underline{\mathbf{H}}}_t\}] = F_t([\mathrm{Im}\{\underline{\mathbf{Y}}_{rp}\}, \mathrm{Re}\{\underline{\mathbf{Y}}_{rp}\}]), \quad (6)$$

其中， F_t 是基于深度神经网络的信道估计器。与$\mathbf{H}_f$相比，使用$\mathbf{H}_t$ 作为标签来训练基于深度神经网络的信道估计器，可以显著减少内存使用量和模型参数，并避免传播信道出现严重的信息损失。值得一提的是， F_t 以接收导频信号（空频域表示）作为输入，并输出估计的空时域信道。因此， F_t 能够隐式地将空频域中的信号转换为空时域中的信道。

此外，正如我们之前所提到的，除了现有深度神经网络架构的复杂度显著增加之外，随着天线阵列规模不断扩大，大规模MIMO系统的导频训练开销也在迅速增加，尤其是频分双工模式（FDD）下的下行链路。通常，在FDD大规模MIMO系统中，下行信道估计的导频训练开销随基站天线数量线性增加，而这在实际场景中实际上是难以承受的。尽管已有一些技术可以用来降低导频训练开销，[27], [28], 研究低开销深度学习信道估计方法对于同时提高大规模MIMO系统的信道估计精度并降低导频训练开销至关重要。

[5] N_P 可根据预先收集的数据，对传播环境中的时延扩展进行粗略估计来确定。由于该模型部署在通信环境相对静态的用户设备上，传播路径的最大时延不会发生显著变化。因此，测试阶段也可以采用 N_P。

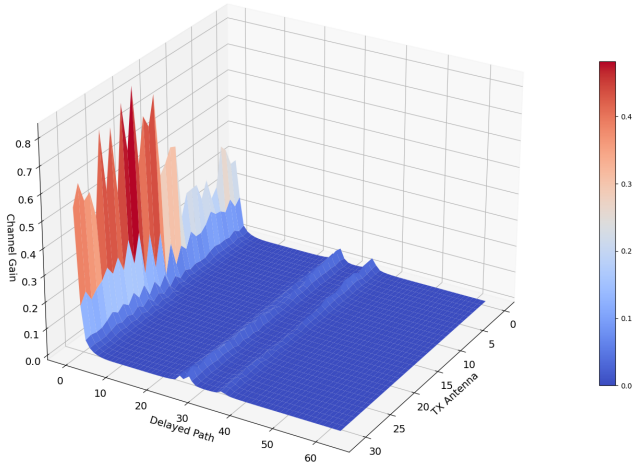

图 1：时空信道表示示意图。

需要注意的是，估计下行信道后，用户设备通常需要先对CSI进行预处理，再对其进行压缩和量化，最后将压缩后的CSI反馈给基站，以便基站进行复杂的信号处理或下行预编码。本文主要关注CSI反馈之前使用低密度导频进行精确信道估计的问题。[6]

C. 导频密度

在 5G 系统中，信道状态信息参考信号（CSI-RS）符号配置在资源块（RB）上，以支持信道估计 [26]。由于 CSI-RS 符号仅配置在一定比例的资源块上，我们将导频密度定义为 ρ：

$$\rho = N_{CSI-RS}/N_{RB}, \quad (7)$$

其中， N_{CSI-RS} 表示配置了 CSI-RS 符号的资源块数量，N_{RB} 表示系统中的资源块总数。为便于后续章节表述，我们采用两种对应于不同导频密度别名的 CSI-RS 配置:

1）**高密度导频，** $\rho_H = N_H/N_{RB}$: CSI-RS 符号配置在 N_H 个资源块上。高密度导频用于性能比较和迁移学习，并将在第IV节介绍。注意，根据 [26], 中的定义，导频密度通常设为1或0。5。因此，我们将 $\rho_H = 0.5$ （即 $N_H = 0.5N_{RB}$）设为用于比较的导频密度上限。

[6] 现有的CSI反馈方案通常假设已知完美下行信道状态信息 [29]–[31] ，或者依赖估计得到的下行信道的特征向量 [32]–[34]。因此，在进行CSI反馈之前，首先获得准确的下行信道估计至关重要。

2) **具有** $\rho_L = N_L/N_{RB}$ 的低密度导频：信道状态信息参考信号符号配置在 N_L 资源块上，且 $N_L < N_H$。

在上述两种配置下，我们分别将接收导频信号表示为$\mathbf{Y}_{pH} \in \mathbb{C}^{N_R \times N_T \times N_H}$和$\mathbf{Y}_{pL} \in \mathbb{C}^{N_R \times N_T \times N_L}$。

为了实现低开销信道估计，深度神经网络应设计为能够在低密度 CSI-RS 配置下执行信道估计。随着天线规模和带宽的增加，通常需要显著提高估计器的复杂度，才能获得准确的信道估计值。因此，对于大规模 MIMO 系统中的低开销信道估计，设计一种时间复杂度和空间复杂度都较低的有效网络架构至关重要。

从图 1所示的时空信道表示示例中可以看出，大规模 MIMO 信道中既存在相邻延迟路径之间的幅度相关性，也存在空间相关性。例如，索引 10 附近相邻延迟路径的信道增益高度相关。根据沿标记为“发射天线”的轴向变化的信道增益，也可以清楚地观察到空间域中的 CSI 相关性。图 1 还表明，时域特征（即延迟路径之间的相关性）与空间域特征（即不同天线处 CSI 之间的相关性）彼此无关。因此，可以预期，采用由多个专门设计的深度神经网络模块组成的先进多域特征提取器联合提取不同领域特征，其性能将优于使用单个深度神经网络模块提取多域特征，因为不同领域特征的提取通常需要采用不同的网络架构设计，以匹配各自的特性。在下一节中，我们提出 DACEN，通过空间注意力模块以及后续的时间注意力模块，联合学习多域特征。

III. 双注意力信道估计网络

为充分利用空时域信道的特性，本节提出 DACEN，分别通过 TAM 和 SAM 专门学习时间域特征和空间域特征。

A. 网络总体架构

如图2所示，所提出的 DACEN 主要由 N_{SA} 空间注意力模块和 N_{TA} 时间注意力模块组成，每个注意力模块均包含一个注意力层、一个前馈层，以及额外的残差连接和层归一化操作。TAM 用于学习不同延迟路径之间的时间依赖关系，而 SAM 则旨在提取空间域特征。在第一个 SAM、第一个 TAM 和输出之前，均设有线性层（即图2 (a)中的“L1”“L2”“L3”和“L4”）和重排操作（用标记为“R”的方框表示），将输入映射到适当的张量维度。特别是，“L1”和“L3”将原始张量映射到高维特征空间，从而实现更好的特征学习。“L1”和“L3”的输出分别记为$\mathbf{X}_{L1} \in \mathbb{R}^{(N_R \times N_T) \times d_{\text{model}}}$ 和$\mathbf{X}_{L3} \in \mathbb{R}^{(2 \times N_P) \times d_{\text{model}}}$。此外，正弦式正弦位置编码$\mathbf{P} \in \mathbb{R}^{(2 \times N_P) \times d_{\text{model}}}$ 与$\mathbf{X}_{L3}$ 形状相同，将其插入 TAM 之前，以便 TAM 学习延迟路径的位置信息。下面将详细介绍所提出的 DACEN 的基本组成部分。

B. 时间注意力模块

TAM 中的时间注意力层旨在发挥多头注意力的优势 [35]，该机制最初被提出用于提取句子中的长期依赖关系。多头注意力已扩展应用于时间序列等其他序列数据 [36] 以及图像块 [37]。由于相同的发射导频信号会穿过多个散射体并形成多条路径，我们提出利用多头注意力从这些路径中学习时间依赖关系，以促进精确的信道估计。将时间注意力层的输入记为$\mathbf{X}_T \in \mathbb{R}^{(2 \times N_P) \times d_{\text{model}}}$，其中原始复数表示的实部和虚部拼接在一起，且 d_{model} 表示输入表示维度。[7] 时间注意力层的运算表示为式（8）—（13）。具体而言，首先将输入投影为 N_h 组（注意力头的数量）查询、键和值矩阵，用于计算各注意力头的注意力。每个注意力头的输出记为$\mathbf{O}_i \in \mathbb{R}^{(2 \times N_P) \times d_v}$，可计算如下：

$$\mathbf{O}_i = \text{Attention}\left(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i\right), i = 1, 2, ..., N_h, \tag{8}$$

Attention(·)表示缩放点积注意力，其输入为查询矩阵$\mathbf{Q}_i \in \mathbb{R}^{(2 \times N_P) \times d_k}$、键矩阵$\mathbf{K}_i \in \mathbb{R}^{(2 \times N_P) \times d_k}$ 和值矩阵$\mathbf{V}_i \in \mathbb{R}^{(2 \times N_P) \times d_v}$。通过将时空域信道的高维特征表示映射，我们得到$\mathbf{Q}_i$、$\mathbf{K}_i$和$\mathbf{V}_i$，如下所示：

$$\mathbf{Q}_i = \mathbf{X}_T \mathbf{W}_i^Q, \tag{9}$$

$$\mathbf{K}_i = \mathbf{X}_T \mathbf{W}_i^K, \tag{10}$$

$$\mathbf{V}_i = \mathbf{X}_T \mathbf{W}_i^V, \tag{11}$$

其中$\mathbf{W}_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$、$\mathbf{W}_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$和$\mathbf{W}_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ 为可学习权重，且 $d_k = d_v = d_{\text{model}}/N_h$。缩放点积注意力 Attention(·)定义如下：

$$\text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d_k}}\right)\mathbf{V}_i, \tag{12}$$

其中，softmax(·)是一种激活函数。最后，将所有注意力头的输出拼接（用标记为“C”的方框表示），得到时间注意力层的输出$\mathbf{X}_{TA} \in \mathbb{R}^{(2 \times N_P) \times d_{\text{model}}}$：

$$\mathbf{X}_{TA} = [\mathbf{O}_1, \ldots, \mathbf{O}_{N_h}]\mathbf{W}^O, \tag{13}$$

其中 $\mathbf{W}^O \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$ 是一个可学习的权重矩阵。值得一提的是，通过将延迟路径的原始表示投影到不同的表示子空间中，如式（9），时间注意力层有望学习到比直接从原始表示中学习更丰富的特征

[7]表示维度通过超参数调优进行经验确定。

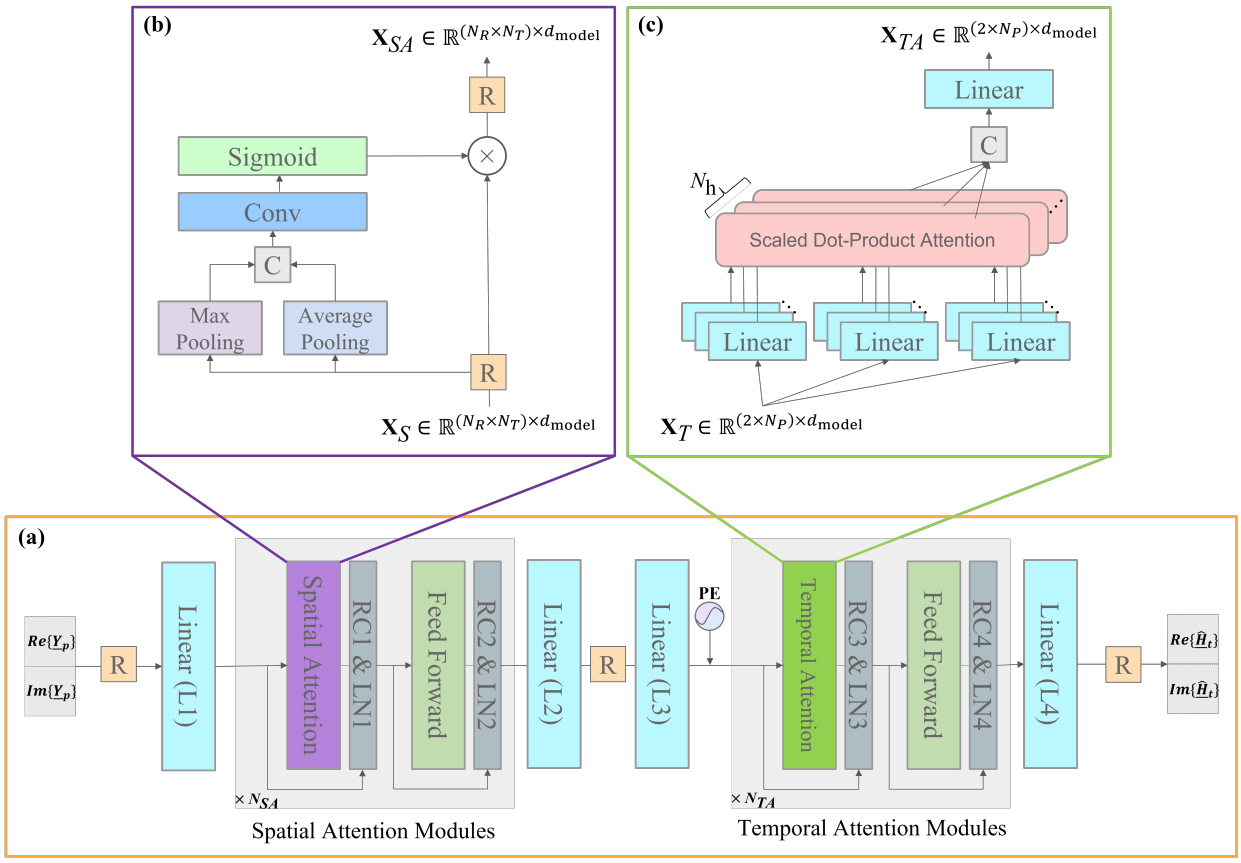

图 2：提出的双注意力信道估计网络（DACEN）。(a) DACEN 的整体网络架构。(b) 空间注意力机制。(c) 时间注意力机制。**RC**：残差连接。**LN**：层归一化。**PE**：位置编码。**C**：拼接操作。**R**：重排操作。

空间，因为直观而言，同时从不同视角观察一个对象时，我们可以获得更多信息。

归一化技术有助于稳定神经网络并缩短其训练时间 [38]，，因此我们在 DACEN 中采用层归一化。残差连接最初由 [39]提出，用于避免梯度消失/爆炸；DACEN 也采用了残差连接，以支持深层神经网络连接。具体而言，在注意力层和前馈层之后，分别应用两组层归一化与残差连接。前馈层由两个线性层及其间的一个 ReLU 激活组成，并紧接在第一个残差连接之后。分别将第一个层归一化、第一个残差连接、前馈层、第二个层归一化和第二个残差连接的输出记为$\mathbf{X}_{LN1} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$、$\mathbf{X}_{RC1} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$、$\mathbf{X}_{FF} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$、$\mathbf{X}_{LN2} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$和$\mathbf{X}_{RC2} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$，则有：

$$\mathbf{X}_{RC1} = \mathbf{X}_{TA} + \mathbf{X}_T, \tag{14}$$

$$\mathbf{X}_{LN1} = \frac{\mathbf{X}_{RC1} - \mathrm{E}_{-1}[\mathbf{X}_{RC1}]}{\sqrt{\mathrm{Var}_{-1}[\mathbf{X}_{RC1}] + \epsilon}} \otimes \mathbf{g}_{LN1} + \mathbf{b}_{LN1}, \tag{15}$$

其中，$\mathrm{E}_{-1}(\cdot)$和 $\mathrm{Var}_{-1}(\cdot)$分别表示沿输入矩阵最后一个维度计算期望和方差，而 $\otimes$ 表示哈达玛积。$\mathbf{g}_{LN1} \in \mathbb{R}^{1\times d_{\text{model}}}$和 $\mathbf{b}_{LN1} \in \mathbb{R}^{1\times d_{\text{model}}}$ 是可学习的仿射变换参数，ϵ 是一个用于保证数值稳定性的小数（例如 1e-5）。

前馈层的输出表示为

$$\mathbf{X}_{FF} = (\mathrm{ReLU}(\mathbf{X}_{RC1}\mathbf{W}_1 + \mathbf{b}_1))\,\mathbf{W}_2 + \mathbf{b}_2, \tag{16}$$

其中，ReLU(·) 是一种激活函数，$\mathbf{W}_1 \in \mathbb{R}^{d_{\text{model}}\times d_{\text{ff}}}$ 和$\mathbf{W}_2 \in \mathbb{R}^{d_{\text{ff}}\times d_{\text{model}}}$ 分别是第一层和第二层线性层的可学习权重，$\mathbf{b}_1 \in \mathbb{R}^{1\times d_{\text{ff}}}$和$\mathbf{b}_2 \in \mathbb{R}^{1\times d_{\text{model}}}$ 分别是第一层和第二层线性层的可学习偏置。d_{ff} 是前馈层的内部维度。经验表明，较大的 d_{model}（例如 512）和 $d_{\text{ff}} \geq d_{\text{model}}$ 能够实现更有效的特征学习。然而，为了限制前馈层的计算复杂度，我们在本文中设置 $d_{\text{ff}} = d_{\text{model}}$。然后，基于$\mathbf{X}_{FF}$，有：

$$\mathbf{X}_{RC2} = \mathbf{X}_{FF} + \mathbf{X}_{LN1}, \tag{17}$$

$$\mathbf{X}_{LN2} = \frac{\mathbf{X}_{RC2} - \mathrm{E}_{-1}[\mathbf{X}_{RC2}]}{\sqrt{\mathrm{Var}_{-1}[\mathbf{X}_{RC2}] + \epsilon}} \otimes \mathbf{g}_{LN2} + \mathbf{b}_{LN2}, \tag{18}$$

其中，$\mathbf{g}_{LN2} \in \mathbb{R}^{1\times d_{\text{model}}}$ 和$\mathbf{b}_{LN2} \in \mathbb{R}^{1\times d_{\text{model}}}$ 是可学习的仿射变换参数。

此外，根据（12），缩放点积注意力以等效方式关注不同的延迟路径，这

忽略了它们的位置信息。为了利用延迟路径的位置信息，我们在输入第一个 TAM 之前，将正弦位置编码添加到线性层“L3”的输出中，即$\mathbf{X}_{L3} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$，如图 2 所示。请注意，正弦位置编码与 $\mathbf{X}_{L3}$ 具有相同的形状，从而为处于不同位置的延迟路径提供唯一的位置信息。具体而言，正弦位置编码 $\mathbf{P} \in \mathbb{R}^{(2\times N_P)\times d_{\text{model}}}$的生成方式如下：

$$\mathbf{P}^{(p,2i)} = \sin\left(\frac{p}{\omega^{2i/d_{\text{model}}}}\right), \tag{19}$$

$$\mathbf{P}^{(p,2i+1)} = \cos\left(\frac{p}{\omega^{2i/d_{\text{model}}}}\right), \tag{20}$$

其中，$p \in [0, 2\times N_P - 1]$ 是沿 $2\times N_P$维度的索引，$i \in [0, d_{\text{model}}/2 - 1]$ 表示沿d_{model} 维度的索引，ω 是与延迟路径数量相关的超参数。之后，第一个 TAM 的输入记为$\mathbf{I}_{TA1}$，可表示为：

$$\mathbf{I}_{TA1} = \mathbf{X}_{L3} + \mathbf{P}, \tag{21}$$

而除第一个 TAM 外，其他 TAM 的输入均为其前一个 TAM 的输出。

TAM 的计算复杂度主要来自时间注意力层和前馈层。根据(8) - (13)和(16)，TAM 的计算复杂度可表示为：

$$\begin{aligned}\mathcal{O}(\text{TAM}) &= \mathcal{O}(3(2N_P)(d_{\text{model}})^2 + 2(2N_P)^2 d_{\text{model}} \\ &\quad + (2N_P)(d_{\text{model}})^2 + 2(2N_P)d_{\text{ff}}d_{\text{model}}) \\ &\overset{(a)}{=} \mathcal{O}(6(2N_P)(d_{\text{model}})^2 + 2(2N_P)^2 d_{\text{model}}),\end{aligned} \tag{22}$$

其中，(a)由 $d_{\text{ff}} = d_{\text{model}}$得出。

C. 空间注意力模块

由于天线阵列通常彼此靠近，且入射信号的角度扩展有限，大规模 MIMO 信道始终表现出空间相关性。此外，不同的天线间距会导致不同程度的空间相关性 [22]。因此，可以主要关注高度相关的天线，在获得良好估计精度的同时，大幅降低信道估计的计算复杂度。尽管卷积神经网络能够提取此类空间相关特征，但随着天线规模的增大，卷积神经网络层的计算复杂度会迅速增加，在实际通信系统中可能变得难以承受。为此，我们采用空间注意力机制 [40]并提出了一种强大且高效的空间域特征提取器，即 SAM。

SAM 中的空间注意力层定义如下。首先，将输入 $\mathbf{X}_S \in \mathbb{R}^{(N_R\times N_T)\times d_{\text{model}}}$ 重排为张量 $\mathbf{X}'_S \in \mathbb{R}^{N_R\times N_T\times d_{\text{model}}}$。然后，沿表示维度对 $\mathbf{X}'_S$进行最大池化和平均池化操作，得到两个特征图 $\mathbf{X}_S^{max} \in \mathbb{R}^{N_R\times N_T\times 1}$ 和 $\mathbf{X}_S^{avg} \in \mathbb{R}^{N_R\times N_T\times 1}$。接着，将这两个特征图拼接，并通过卷积层和 Sigmoid 激活函数处理以生成一个空间注意力图。随后，将空间注意力图与输入进行哈达玛积，得到输出 $\mathbf{X}''_{SA} \in \mathbb{R}^{N_R\times N_T\times d_{\text{model}}}$。最后，通过另一个重排操作，将$\mathbf{X}''_{SA}$重排为 $\mathbf{X}_{SA} \in \mathbb{R}^{(N_R\times N_T)\times d_{\text{model}}}$。具体而言，空间注意力层的操作可表示为(23)至(25)：

$$\underline{\mathbf{X}}'_S = \mathrm{R}(\mathbf{X}_S), \tag{23}$$

$$\underline{\mathbf{X}}''_S = \sigma\left(\mathrm{Conv}^{1\times 1}\left([\mathrm{MP}(\underline{\mathbf{X}}'_S), \mathrm{AP}(\underline{\mathbf{X}}'_S)]\right)\right) \otimes \underline{\mathbf{X}}'_S, \tag{24}$$

$$\mathbf{X}_{SA} = \mathrm{R}(\underline{\mathbf{X}}''_S), \tag{25}$$

其中，$\sigma(\cdot)$表示 Sigmoid 激活函数，$\mathrm{Conv}(\cdot)^{1\times 1}$ 是一个具有 1×1 滤波器的卷积层，MP(·)和 AP(·)分别表示最大池化和平均池化。

此外，为进一步增强 SAM 的学习能力，我们受 Transformer 架构启发，在空间注意力机制之后添加了一个前馈层 [35],，并在空间注意力机制和前馈层周围加入残差连接和层归一化操作。得益于这种专门设计的架构，SAM 能够在空间域中一步灵活地捕获大规模 MIMO 信道的全局和局部特征。此外，在相同条件下，与卷积神经网络相比，SAM 的复杂度更低，参数更少。SAM 的计算复杂度可计算如下：

$$\begin{aligned}\mathcal{O}(\text{SAM}) &= \mathcal{O}(2N_RN_T + N_RN_Td_{\text{model}} + 2N_RN_Td_{\text{ff}}d_{\text{model}}) \\ &\overset{(a)}{=} \mathcal{O}(2N_RN_T + N_RN_Td_{\text{model}} + 2N_RN_T(d_{\text{model}})^2),\end{aligned} \tag{26}$$

其中 (a) 源自 $d_{\text{ff}} = d_{\text{model}}$ 如第III节 B 所述。

作为比较，当采用大小为$K_S \times K_S$ （例如 3×3 或 7×7）的滤波器[8] 和表示维度 d_{model} 的标准二维卷积神经网络（即空间卷积层（SConv））进行空间域特征提取时，SConv 的计算复杂度为：

$$\mathcal{O}(\text{SConv}) = \mathcal{O}(N_RN_T(K_S)^2(d_{\text{model}})^2). \tag{27}$$

比较（27）与（26）可以发现，SAM 的计算效率高于 SConv。

D. 学习策略

给定用于训练的低密度导频样本及其对应的信道标签，即$D_L^{train} = \{ \quad _{pL\,i} \text{—} \; _{t\,i} \; | \underline{i} = (\mathbf{Y}_{\cdot}, \mathbf{H}_{\cdot}^{1\ldots}), \; , \; N_{D_L}^{train}\}$，所提出的 DACEN 可通过最小化均方误差（MSE）损失从头开始训练，如下一页顶部的式（28）所示，其中 $N_{D_L}^{train}$是用于训练的低密度样本数量。

从头训练 DACEN 的过程如图所示。3 (a)。考虑到通常使用更多导频有助于获得更准确的信道估计值，因此基于预先获取的接收的高密度导频 $\mathbf{Y}_{pH} \in \mathbb{C}^{N_R\times N_T\times N_H}$($N_H > N_L$)（例如，在

[8]滤波器尺寸通常设置为大于 1，以支持稀疏连接并保持空间局部性，最终确保卷积神经网络的性能。

$$F_L^{Scr} = \arg\min_{F_L} \frac{1}{N_{D_L}^{train}} \sum_{i=1}^{N_{D_L}^{train}} \|[\mathrm{Im}\{\underline{\mathbf{H}}_{t,i}\},\mathrm{Re}\{\underline{\mathbf{H}}_{t,i}\}] - F_L([\mathrm{Im}\{\mathbf{Y}_{\underline{pL},i}\},\mathrm{Re}\{\mathbf{Y}_{\underline{pL},i}\}])\|_2^2, \quad (28)$$

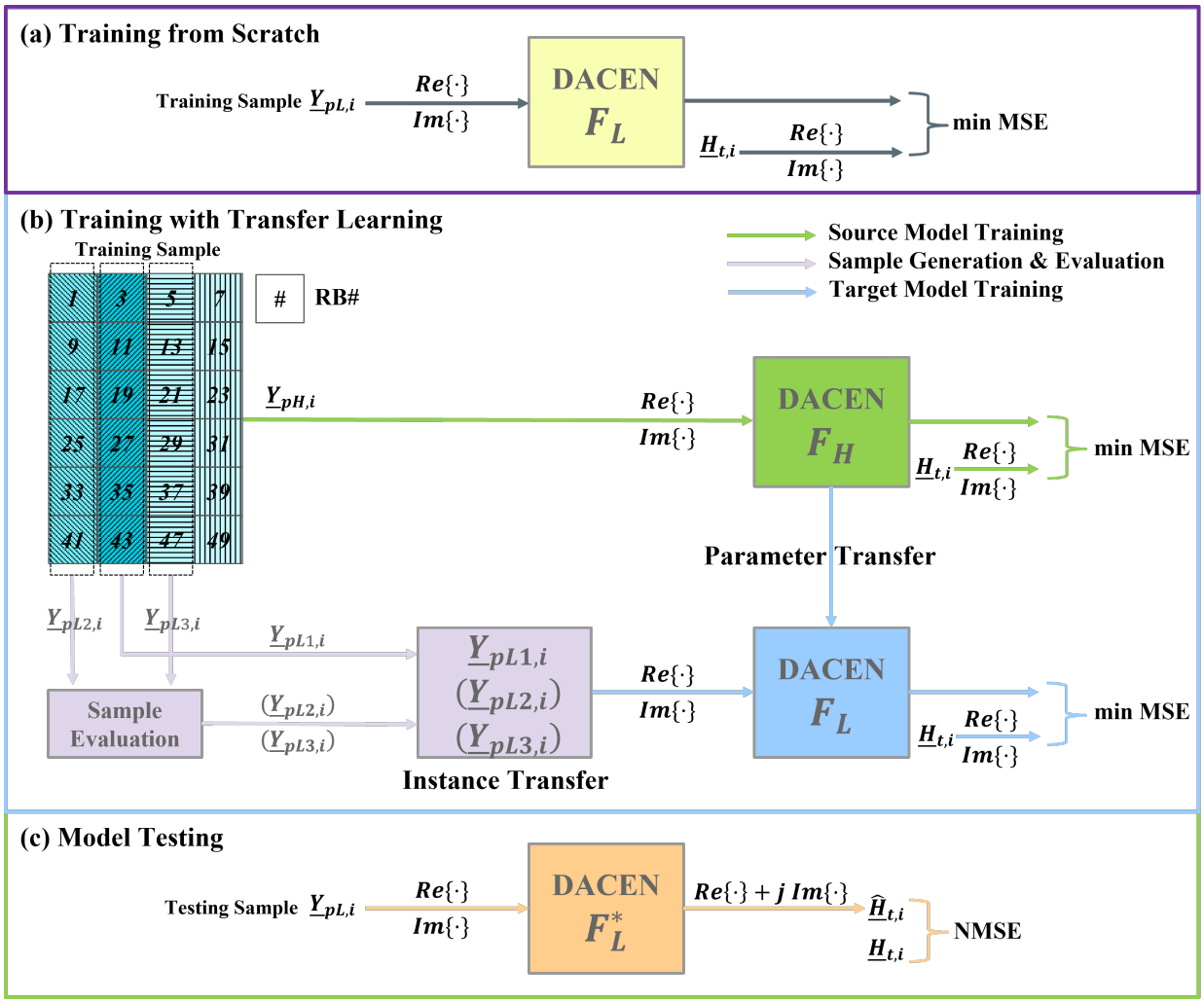

图 3：DACEN 的训练和测试过程。(a) 从头开始训练 DACEN。(b) 使用参数-实例迁移学习方法训练 DACEN。(c) 使用低密度导频测试 DACEN。

数据集收集期间分配特定时隙进行高密度导频训练），采用迁移学习训练 DACEN 有望提升其估计性能。下一节中，我们提出一种参数-实例迁移学习方法，以进一步提升 DACEN 的估计性能。

IV. 面向低开销信道估计的参数-实例迁移学习

通常，使用更多导频进行信道估计，比使用较少导频能够获得更准确的信道估计值，因为更多导频能让我们学习到更多信道知识。因此，我们利用迁移学习，将训练数据集采集期间预先获取的高密度导频所学到的信道知识迁移过来，进一步提高我们提出的基于 DACEN 的信道估计方法的准确性。本节首先简要介绍迁移学习的背景和常见的迁移学习方法，随后介绍所提出的参数-实例迁移学习方法。

A. 迁移学习

机器学习方法的基本假设是，训练数据和测试数据均从同一领域采样，这意味着两个数据集处于相同的特征空间中，并具有相同的分布 [41], [42]。然而，在许多现实场景中，这一假设可能并不成立。因此，迁移学习旨在通过将知识从源任务迁移到目标任务来改进学习器，以适应基本假设不成立的情况 [41]。假设一个源域S、相应的源任务 $\mathcal{T}_S$ 和一个学习器 F_S 用于 $\mathcal{T}_S$。一个

从源域 S 收集的训练数据集（记为 $\mathcal{D}_S$）可用于训练学习器 F_S，以完成源任务 $\mathcal{T}_S$。当数据分布发生变化时，会出现一个新的数据领域及其相应任务，分别记为T 和 $\mathcal{T}_T$。从目标域 T 重新收集大型训练数据集可能代价高昂，甚至无法实现。因此，针对目标任务 $\mathcal{T}_T$ 的学习器 F_T 只能利用少量目标域数据，甚至在完全没有目标域数据 T 的情况下获得。针对这类情况，迁移学习（也称为知识迁移）旨在通过从源域数据 $\mathcal{D}_S$ 迁移知识，改进用于 $\mathcal{T}_T$ 的学习器 F_T。

根据“迁移什么”，迁移学习中有四种典型方法 [41]，包括实例迁移方法、参数迁移方法、特征表示迁移方法和关系知识迁移方法。

1)实例迁移方法。假设源域中的某些数据有助于完成目标任务，实例迁移方法旨在重新利用这些数据来训练目标任务的学习器，通常采用实例重加权和重要性采样。

2) 参数迁移方法。参数迁移方法假设源任务和目标任务可能共享学习器的某些参数或超参数的先验分布。因此，可将要迁移的知识编码到共享参数或先验中。在深度迁移学习中，广为人知的预训练方法就是一种典型的参数迁移方法。

3) 特征表示迁移方法。特征表示迁移方法旨在将知识直接编码为适用于目标任务的良好特征表示。换言之，这种特征表示应具有域不变性，从而有望提升目标任务的性能。

4)关系知识迁移方法。关系知识迁移方法旨在关系域之间进行知识迁移，其依据是关系域中的数据之间存在某些相似关系，而要迁移的知识直观而言就是数据之间的关系。

据我们所知，参数迁移方法已通过从先前经历过的环境中迁移环境知识，应用于时变环境下的信道预测 [24]，[25]，而其他方法尚未广泛用于信道估计。由于特征表示迁移方法需要专门的特征选择流程，且关系知识迁移方法仅适用于关系域内的知识迁移，因此无法直接用于改进低开销信道估计。相比之下，考虑到可以从更多导频中学习到更多信道知识，参数迁移方法和实例迁移方法有望通过预训练，并利用预先获取的高密度导频扩充更多训练数据，充分迁移信道知识。因此，我们提出参数-实例迁移学习方法，以提高基于 DACEN 的信道估计方法。

B. 所提出的参数-实例迁移学习方法

值得注意的是，所提出的参数-实例迁移学习方法兼具参数迁移和实例迁移的优势。一方面，使用高密度导频训练的 DACEN（记为 F_H^*）的参数存储了有关信道的有效知识。将 F_H^* 的参数用作低密度导频 DACEN 训练的初始参数，可以迁移嵌入这些参数中的知识。另一方面，从网络训练的角度来看，通过从高密度导频样本生成相邻的低密度导频样本来进行数据增强，可以丰富原有的低密度导频样本，从而进一步提升基于 DACEN 的低开销信道估计性能。

为了利用高密度导频的优势，在模型训练阶段，基站向用户设备发送高密度导频，从而生成高密度导频样本。考虑由 N_{D_H} 个高密度导频样本组成的源数据集 $\mathcal{D}_H$，以及使用 $\mathcal{D}_H$ 进行信道估计的任务，将其作为源任务 $\mathcal{T}_H$。与接收导频样本相对应的时空域信道表示用作训练 DACEN 的标签。记 $\mathbf{Y}_{pH,\,i} \in \mathbb{C}^{N_R \times N_T \times N_H}$ 为 $\mathcal{D}_H$ 中第 i 个高密度导频样本，$\mathbf{H}_{t,\,i}$ 为相应的信道标签。目标数据集 $\mathcal{D}_L$ 包含低密度导频样本，使用 $\mathcal{D}_L$ 进行信道估计的任务是目标任务 $\mathcal{T}_L$。模型训练阶段结束后，将训练完成的目标模型部署到用户设备上。此时，基站向用户设备发送低密度导频，随后用户设备利用部署的模型，根据接收到的低密度导频估计下行信道。

采用所提出的参数-实例迁移学习方法，模型训练过程如图所示。3 （b）。所提出的参数-实例迁移学习方法包括三个阶段，即源模型训练阶段、样本生成与评估阶段以及目标模型训练阶段。在源模型训练阶段，我们使用高密度导频样本训练用于信道估计的源模型。我们预计，模型参数能够学习并存储更加完整的知识。在样本生成与评估阶段，我们通过选择相邻导频样本，从高密度导频样本中采样得到低密度导频样本。随后，基于余弦相似度准则，对获得的低密度导频样本进行评估和排除。通过评估和排除过程，可以避免引入过多的不相似样本以及连续的负迁移问题 [43]。在目标模型训练阶段，我们最终使用生成的低密度导频样本以及从源模型迁移的参数，训练用于低开销信道估计的目标模型。所提出的迁移学习算法的伪代码见算法1。下面我们将根据算法1，详细阐述所提出的参数-实例迁移学习方法。

阶段 1：源模型训练（第 1 行至第 5 行）。给定包含高密度导频的源数据集 $\mathcal{D}_H$

样本及对应的信道标签 $\{(\mathbf{Y}_{pH,i},\mathbf{H}_{t,i})|i=1,...,N_{D_H}\}$，我们将其划分为训练集、验证集和测试集。利用训练子集 $\{\mathbf{Y}_{pH,i}|i=1,...,N_{D_H}^{train}\}$ 和 $\mathcal{F}_H$ 的 $\mathbf{H}_{t,i}$，首先通过最小化均方误差损失来训练 DACEN。注意，F_H 预期能够达到较高的估计精度，最优的 F_H 可由下一页顶部的式（29）确定，其中 $\|\cdot\|_2$表示 L2 范数，$N_{D_H}^{train}$表示用于训练的样本数量。随后，将训练完成的 F_H^* 骨干网络参数用作良好的初始化，以训练 F_L，实现 $\mathcal{T}_L$低开销信道估计。

阶段 2：样本生成与评估（第 6 行至第 22 行）。首先，初始化一个空的目标数据集 $\mathcal{D}_L$，用于存放生成的低密度导频样本。然后，从 $\mathbf{Y}_{pH,i}\in\mathbb{C}^{N_R\times N_T\times N_H}$ 中选取 N_L 个元素 $\{\mathbf{Y}_{pH,i}^{(r)}|r=r_0,r_0+I,...,r_0+I*(N_L-1)\}$，即 $\mathbf{Y}_{pH,i}^{(r_0)},\mathbf{Y}_{pH,i}^{(r_0+I)},...,\mathbf{Y}_{pH,i}^{(r_0+I*(N_L-1))}$，生成一个低密度导频样本 $\mathbf{Y}_{pL1,i}\in\mathbb{C}^{N_R\times N_T\times N_L}$，其中 r_0 是起始资源块的索引，I 是相邻两个元素之间的索引间隔。注意，每个 N_L 元素都表示一个资源块中的接收导频信号。生成的样本及其对应的信道标签构成目标数据集 $\mathcal{D}_L$，该数据集也分别划分为训练子集$\mathcal{D}_L^{train}$（包含 $N_{D_L}^{train}$ 个样本）、验证子集 $\mathcal{D}_L^{val}$和测试子集 $\mathcal{D}_L^{test}$。具体而言，将 $\mathcal{D}_L^{train}$复制到一个新集合 $\mathcal{D}_{L_E}^{train}$中，以存放扩展样本来训练 F_L。考虑到相邻资源块在频域上的相干性，可以得到两个可能与 $\mathbf{Y}_{pL1,i}$相关的额外低密度导频样本，分别为 $\{\mathbf{Y}_{pH,i}^{(r)}|r=r_0-1, r_0+I-1,...,r_0+I*(N_L-1)-1\}$ 和 $\{\mathbf{Y}_{pH,i}^{(r)}|r=r_0+1,r_0+I+1,...,r_0+I*(N_L-1)+1\}$。将这两个样本记为 $\mathbf{Y}_{pL2,i}\in\mathbb{C}^{N_R\times N_T\times N_L}$ 和 $\mathbf{Y}_{pL3,i}\in\mathbb{C}^{N_R\times N_T\times N_L}$。为避免将不相关数据纳入训练所带来的负面影响，还需评估 $\mathbf{Y}_{pL1,i}$、$\mathbf{Y}_{pL2,i}$ 和 $\mathbf{Y}_{pL3,i}$ 之间的相关性。以评估 $\mathbf{Y}_{pL2,i}$ 和 $\mathbf{Y}_{pL1,i}$ 之间的相似度为例，首先将 $\mathbf{Y}_{pL2,i}$和 $\mathbf{Y}_{pL1,i}$ 分别重新排列为矩阵 $\mathbf{V}_{pL2,i}\in\mathbb{C}^{N_L\times(N_R\times N_T)}$ 和 $\mathbf{V}_{pL1,i}\in\mathbb{C}^{N_L\times(N_R\times N_T)}$。随后，通过逐一匹配其 N_L 对行向量来评估相似度。将所有 N_L 对向量的中间余弦相似度向量记为 $\mathbf{a}_{12,i}\in\mathbb{R}^{N_L\times 1}$，$\mathbf{a}_{12}$ 的第 j 个元素表示第 j 对向量 $(\mathbf{V}_{pL2,i}^{(j)},\mathbf{V}_{pL1,i}^{(j)})$ 的余弦相似度：

$$\mathbf{a}_{12,i}^{(j)}=cossim(\mathbf{V}_{pL2,i}^{(j)},\mathbf{V}_{pL1,i}^{(j)})=\frac{|<\mathbf{V}_{pL2,i}^{(j)},\mathbf{V}_{pL1,i}^{(j)}>|}{\|\mathbf{V}_{pL2,i}^{(j)}\|_2\|\mathbf{V}_{pL1,i}^{(j)}\|_2}, \tag{30}$$

其中，$cossim(\cdot)$表示余弦相似度，$<\cdot,\cdot>$表示内积。接下来，我们对所有 N_L 向量对的余弦相似度取平均，得到最终余弦相似度得分：

$$\mathbf{s}_{12}^{(i)}=\frac{1}{N_L}\sum_{j=1}^{N_L}\mathbf{a}_{12,i}^{(j)}, \tag{31}$$

其中，$\mathbf{s}_{12}=[\mathbf{s}_{112},...,\mathbf{s}_{12}^{N_{D_L}^{train}}]^T\in\mathbb{R}^{N_{D_L}^{train}\times 1}$是所有 $N_{D_L}^{train}$样本对的余弦相似度向量。样本 $\mathbf{Y}_{pL2,i}$ 一旦满足条件，就会被添加到扩展训练子集 $\mathcal{D}_{L_E}^{train}$ 最终分数 $\mathbf{s}_{12}^{(i)}$ 大于或等于预定义阈值s^{th}（例如 0.8），并将最终分数记录为其实例权重。对所有高密度导频样本执行样本生成过程，并对训练子集 $\mathcal{D}_H^{train}$中的所有高密度导频样本执行评估过程。

第3阶段：目标模型训练（第23行 - 第27行）。给定生成的低密度导频样本$\{\mathbf{Y}_{pL,i}|i=1,.....,N_{D_{L_E}}^{train}\}$（包括$\mathbf{Y}_{pL1,i}$、$\mathbf{Y}_{pL2,i}$或$\mathbf{Y}_{pL3,i}$）$\mathcal{D}_{L_E}^{train}$中的对应信道标签$\{\mathbf{H}_{t,i}|i=1,.....,N_{D_L}^{train}\}$[9] 以及记录的实例权重 $\mathbf{w}$，随后即可基于从**阶段 1** 迁移而来的参数初始化来训练所提出的 DACEN，具体可表示为下一页顶部的（32），其中$N_{D_{L_E}}^{train}$为扩展训练子集 $\mathcal{D}_{L_E}^{train}$中的样本数。注意，$\mathbf{Y}_{pL1,i}$ 的实例权重为 1。

值得注意的是，如图 3 （c）所示，在模型测试阶段，训练好的 DACEN网络仅接收低密度导频样本，以获取信道估计值。

V. 实验

本节首先介绍实验设置，包括系统配置、超参数设置和评估指标等，用于数值评估。随后，我们在不同的导频密度设置下，将所提出的方法与现有方法的性能进行比较。最后，我们通过消融研究进一步验证 SAM 和 TAM 的有效性。

A. 实验设置

为验证所提出的基于DACEN的低开销信道估计方法的有效性，我们首先利用MATLAB 5G工具箱 [44]和[45]，基于簇状时延线（CDL）MIMO信道模型构建仿真数据集。考虑一个由配备32根天线的基站服务的单小区系统。该系统采用OFDM，时频资源包含52个资源块，每12个子载波和14个OFDM 符号构成一个资源块。单小区系统中随机分布有2,000个用户设备，其中1,800个用户设备用于训练与验证，因此被配置为接收高密度导频；其余200个用户设备在不同的信噪比设置和导频密度设置下接收导频，以充分比较所提出的基于DACEN的低开销信道估计方法与基线方法。每个用户设备配备4根天线，速度为15 km/h。基站以10分贝（dB）的信噪比向训练集和验证集中的用户设备通过26个资源块发送信道状态信息参考信号，并记录连续十帧中十个均匀分布的时隙所接收的高密度导频。因此，为训练（16,200个样本）和验证（1,800个样本）提供了来自训练集和验证集内1,800个用户设备的18,000个高密度接收导频样本，其导频密度为 $\rho_H=26/52$。而在测试中，给定特定的信噪比和导频密度设置，使用测试集中200个用户设备的2,000个接收导频样本进行性能评估。信噪比设置、导频密度

[9]请注意，$\mathbf{Y}_{pL1,i}$、$\mathbf{Y}_{pL2,i}$和$\mathbf{Y}_{pL3,i}$ 共享相同的信道标签$\mathbf{H}_{t,i}$。

$$F_H^* = \arg\min_{F_H} \frac{1}{N_{D_H}^{train}} \sum_{i=1}^{N_{D_H}^{train}} \|[\mathrm{Im}\{\underline{\mathbf{H}}_{t,i}\}, \mathrm{Re}\{\underline{\mathbf{H}}_{t,i}\}] - F_H([\mathrm{Im}\{\underline{\mathbf{Y}}_{pH,i}\}, \mathrm{Re}\{\underline{\mathbf{Y}}_{pH,i}\}])\|_2^2, \tag{29}$$

$$F_L^{TL} = \arg\min_{F_L} \frac{1}{N_{D_{L_E}}^{train}} \sum_{i=1}^{N_{D_{L_E}}^{train}} \mathbf{w}^{(i)} \|[\mathrm{Im}\{\underline{\mathbf{H}}_{t,i}\}, \mathrm{Re}\{\underline{\mathbf{H}}_{t,i}\}] - F_L([\mathrm{Im}\{\underline{\mathbf{Y}}_{pL,i}\}, \mathrm{Re}\{\underline{\mathbf{Y}}_{pL,i}\}])\|_2^2, \tag{32}$$

算法1 参数-实例迁移学习方法 **输入：** $\mathcal{D}_H$：源数据集，包含高密度导频样本及相应的信道标签，即 $\{(\mathbf{Y}_{pH,i}, \mathbf{H}_{t,i}) | i = 1, ..., N_{D_H}\}$；$\mathcal{T}_H$，$\mathcal{T}_L$：源任务和目标任务；$F_H$，$F_L$：用于源任务和目标任务的DACEN网络；$N_H^{itr}$，$N_L^{itr}$：训练的最大迭代次数，以及 N_L：一个低密度导频样本中的资源块数量；I：从高密度导频样本中采样两个相邻元素的间隔；r_0：低密度导频的起始资源块编号；s^{th}：纳入样本的余弦相似度阈值 **阶段 1：源模型训练** 1：初始化：将 $\mathcal{D}_H$ 划分为训练、验证和测试子集：$\mathcal{D}_H^{train}$、$\mathcal{D}_H^{val}$、$\mathcal{D}_H^{test}$；$m \leftarrow 0$；随机初始化；2：重复 3：利用 $\mathcal{D}_H^{train}$ 中的样本，按照式（29）更新 F_H 4：$m \leftarrow m + 1$ 5：直至 $m \geq N_H^{itr}$ **阶段 2：样本生成与评估** 6：初始化：初始化一个空的目标数据集 $\mathcal{D}_L$，以包含生成的低密度导频样本 7：对于 $i = 1, ..., N_H$ 执行 8：从 $\mathbf{Y}_{pH,i}$ 中采样 $\{\mathbf{Y}_{pH,i}^{(r)} | r = r_0, r_0 + I, ..., r_0 + I * (N_L - 1)\}$，以形成低密度导频样本 $\mathbf{Y}_{pL1,i}$；9：将 $(\mathbf{Y}_{pL1,i}, \mathbf{H}_{t,i})$ 插入 $\mathcal{D}_L$；10：结束循环 11：将 $\mathcal{D}_L$ 划分为训练、验证和测试子集：$\mathcal{D}_L^{train}$（包含 $N_{D_L}^{train}$ 个样本）、$\mathcal{D}_L^{val}$ 和 $\mathcal{D}_L^{test}$；12：对于 $i = 1, ..., N_{D_L}^{train}$ 执行 13：从 $\mathbf{Y}_{pH,i}$ 中采样 $\{\mathbf{Y}_{pH,i}^{(r)} | r = r_0 - 1, r_0 + I - 1, ..., r_0 + I * (N_L - 1) - 1\}$ 以及 $\{\mathbf{Y}_{pH,i}^{(r)} | r = r_0 + 1, r_0 + I + 1, ..., r_0 + I * (N_L - 1) + 1\}$，以形成 $\mathbf{Y}_{pL2,i}$ 和 $\mathbf{Y}_{pL3,i}$；14：对于执行 16：将 $\mathbf{Y}_{pLk,i}$ 和 $\mathbf{Y}_{pL1,i}$，重排为 $\mathbf{V}_{pLk,i}$，$k = 2, 3$ 15：根据式（30）和（31）计算最终余弦相似度得分 $\mathbf{s}_{1k}^{(i)}$ 17：若 $\mathbf{s}_{1k}^{(i)} \geq s^{th}$，则 18：将 $(\mathbf{V}_{pLk,i}, \mathbf{H}_{t,i})$ 插入 $\mathcal{D}_{L_E}^{train}$ 19：结束条件 20：结束循环 21：结束循环 22：

第3阶段：目标模型训练 23：初始化：基于最优源模型 F_H^* 初始化 F_L，$m \leftarrow 0$ 24：重复 25：根据（32）使用 $\mathcal{D}_{L_E}^{train}$ 中的样本进行更新；26：$m \leftarrow m + 1$ 27：直到 $m \geq N_L^{itr}$ **输出：** 最优目标模型 F_L^{TL}

设置以及其他关键系统参数列于表II中。

多个天线端口之间用于信道测量的一个资源块内信道状态信息参考信号的配置如图4所示。 32 个资源元素被划分为16组，在每组中，仅有两个发射端口发送参考信号（即和 $[+1, -1]$，分别对应）。信道状态信息参考信号配置的详细信息请参见 [26]。

请注意，测得的空频域信道表示对应于624 个频域子载波。通过逆快速傅里叶变换操作（4）可得到（$N_P = 64$），时空信道表示在时域信道响应中仅有 64 个条目，约为1/10 空频域信道表示中子载波数量的这一比例。因此，以时空信道表示作为训练基于深度神经网络的信道估计器的标签，可以大幅减少内存使用量和模型复杂度。

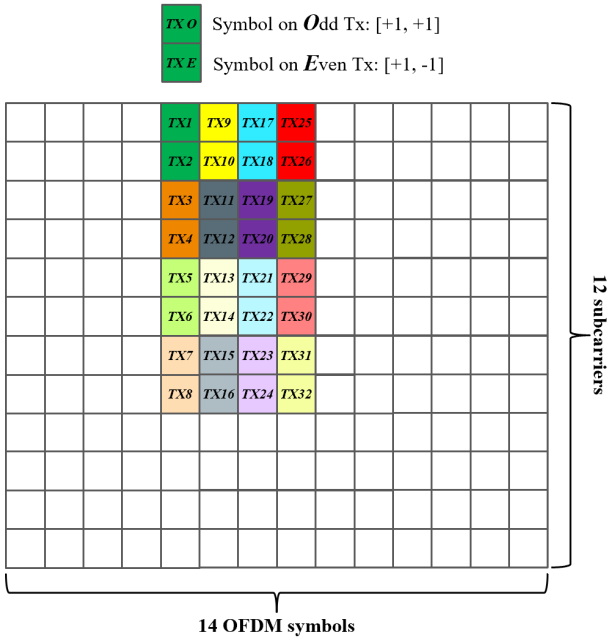

图 4：一个资源块内的信道状态信息参考信号配置。

为提高所提出的 DACEN 的通用性，采集来自不同场景下分布式用户设备的数据样本

[10]簇的时延分布定义与CDL-E信道模型相同 [44]。

表 II：系统参数设置。

参数	值
信道模型	CDL
簇数量[10]	14
第一簇中的莱斯 K 因子（分贝）	-10
信道采样率	61.44 兆赫
FFT 大小	1024
载波频率	3.5 吉赫兹
子载波间隔	60 kHz
用户设备速度	15 km/h
N_{RB}	52
N_T	32
N_R	4
N_c	624
训练集中的用户设备数量	1、800
测试集中的用户设备数量	200
信噪比设置（分贝）	5、10、15、20、25
导频密度设置	26/52, 6/52, 4/52, 2/52

表 III：所提出的 DACEN 的最终超参数设置。

超参数	设置
表示维度 d_{model}	512
前馈维度 d_{ff}	512
注意力头数 N_h	2
SA模块数量 N_{SA}	8
TA模块数量 N_{TA}	8
PE超参数 ω	10000
批次大小 N_{bs}	256
初始学习率 η	$6e-5$

训练集中的数据用于训练模型。这基于以下观察。首先，相邻的用户设备共享一些散射体，而位于不同位置的用户设备可能具有相同的移动模式。利用来自多个位置的数据对这些物理相关性进行建模，有助于提高深度学习方法的泛化能力。其次，训练过程是模型学习将接收导频映射到信道矩阵的过程。利用来自具有多样场景的分布式用户设备的训练数据，可能会简化学习过程，并使模型具有更好的泛化能力。

DACEN使用PyTorch实现 [46]，，并使用Adam优化器 [47] ，通过两张英伟达3090 GPU显卡训练DACEN。所提出的基于DACEN的方法的超参数设置如表III所示。采用以分贝为单位的归一化均方误差（NMSE）来评估信道估计性能，其定义如下：

$$\text{NMSE (dB)} = 10 * \log\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|\underline{\mathbf{H}}_{t,i} - \underline{\hat{\mathbf{H}}}_{t,i}\|_2^2}{\|\underline{\mathbf{H}}_{t,i}\|_2^2}\right), \quad (33)$$

其中 N 是样本数量。

我们还通过浮点乘加运算次数（“FLOPs”）和参数数量（“Params”），分别评估所提出的 DACEN 在时间复杂度和空间复杂度方面的计算复杂度。

B. 所提出的基于 *DACEN* 的信道估计方法的有效性

将所提出的基于DACEN的信道估计方法与几种基线方法和最先进（SOTA）在不同信噪比和导频密度设置下的信道估计方法进行了比较。具体而言，我们将所提出的基于DACEN的方法与传统的基于最小二乘（LS）估计的方法 [48] 以及基于线性最小均方误差（LMMSE）估计的方法 [49]进行了比较。对于基于LS的信道估计方法，首先利用传统的LS算法对具有已知导频符号的子载波进行估计，然后通过插值方法获得所有子载波的估计值。而对于基于LMMSE的方法，则获取两个维纳滤波矩阵，分别用于利用相邻天线之间以及相邻子载波之间的相关性。随后，基于LS估计值和 [48] 除传统方法外，还比较了所提出方法与基于深度学习的信道估计方法的性能，包括基于全连接深度神经网络（FC-DNN）的信道估计方法 [15],基于卷积神经网络（CNN）的信道估计方法 [16],CDRN[21],以及跳跃连接注意力（SC-attention）网络[50],的性能。需要注意的是，为了进行公平比较，我们还对基于深度学习的基线方法的超参数进行了调优，以获得其最佳性能。

图 5 至图 8 展示了不同信噪比和导频密度设置下的归一化均方误差结果。请注意，图 6 至图 8 中的 DACEN（标记为“DACEN”）是使用低密度导频从头训练的，而“DACEN+TF”则按照算法1，采用所提出的迁移学习算法进行训练。为了更好地将所提出的方法与其他基于深度学习的方法进行比较，表 IV 展示了它们的计算复杂度。[11] 从这些结果可以看出，在特定导频密度下，所提出的基于 DACEN 的方法在所有信噪比条件下均优于其他方法。具体而言，采用 $\rho_L = 2/52$的所提出的基于 DACEN 的方法，其性能优于采用高密度导频的传统 LS 和 LMMSE 方法。此外，采用 $\rho_L = 6/52$ 的所提出的基于 DACEN 的方法，其性能也与采用 $\rho_H = 26/52$的 FCDNN 和 CNN 估计器相当。因此，这进一步证实了所提出的 DACEN 能够降低导频开销。与当前最先进的方法（即 CDRN 和 SC注意力网络）相比，所提出的基于 DACEN 的方法在采用高密度导频时，归一化均方误差性能分别提升 3 分贝和 1 分贝；采用 $\rho_L = 2/52$时，归一化均方误差性能则分别提升 3 分贝和 0.6 分贝。此外，从表 IV 可以看出，DACEN 的计算复杂度远低于 CDRN 和 SC注意力网络，这证明了所提出的基于 DACEN 的方法具有较高的效率。

‘DACEN+TF’在三种低密度导频设置下实现了相应的归一化均方误差性能，即 $\rho_L = 6/52$, ρ_L =4/52，以及 $\rho_L = 2/52$。根据算法1，我们分别为 $\rho_L = 6/52$、 $r_0 = 5$、$I = 5$ 和 $N_L = 4$ 设置 $r_0 = 1$、 $I = 4$ 和 $N_L = 6$ ，并为 $\rho_L = 4/52$设置 $r_0 = 9$、 $I = 8$ 和$N_L = 2$ ，为 $\rho_L = 2/52$ 设置相应参数。余弦相似度阈值 s^{th} 设为 0.9。DACEN通过知识迁移，在训练过程中受益于具有 $\rho_H = 26/52$ 的高密度

[11]表IV 表示测试阶段输入为低密度导频且 $\rho_L = 6/52$ 时的模型复杂度。

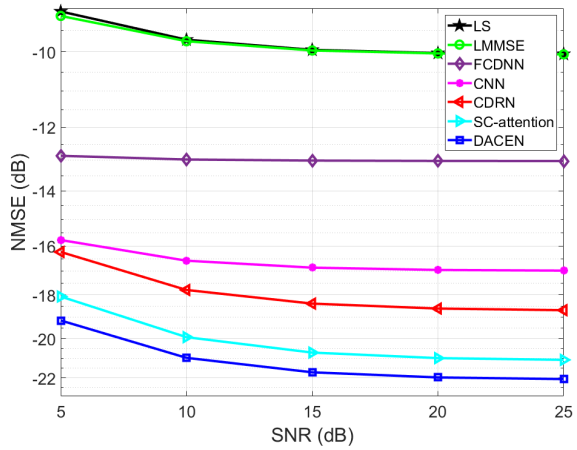

图 5：$\rho_H = 26/52$条件下不同信噪比的归一化均方误差性能。

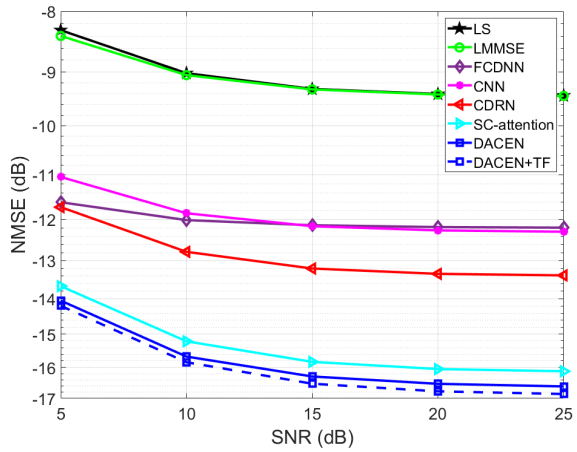

图 6：$\rho_L = 6/52$条件下不同信噪比的归一化均方误差性能。

导频，从而最多可获得 0.25 分贝，0.6 分贝，以及 0.6 分贝的额外性能提升，相较于从头开始训练。如表IV所示，值得强调的是，所提出的迁移学习算法在测试阶段不会增加模型复杂度，因此不会给用户设备带来额外负担。

一旦在多个用户设备上部署所提出的 DACEN，该方法即可直接应用于单小区多用户场景。基站可通过下行广播模式向小区内所有用户发送信道状态信息参考信号，使所有用户均能接收并解调该信号。在这种情况下，由于信道状态信息参考信号采用下行广播，不同用户之间不存在干扰。因此，尽管我们的仿真结果基于单用户场景，但在一定程度上反映了多用户多输入多输出性能。

C. 消融研究

为进一步评估 TAM 和 SAM 的有效性与计算效率，我们进行了消融研究。在固定导频密度 $\rho_L = 6/52$ 的情况下，我们首先

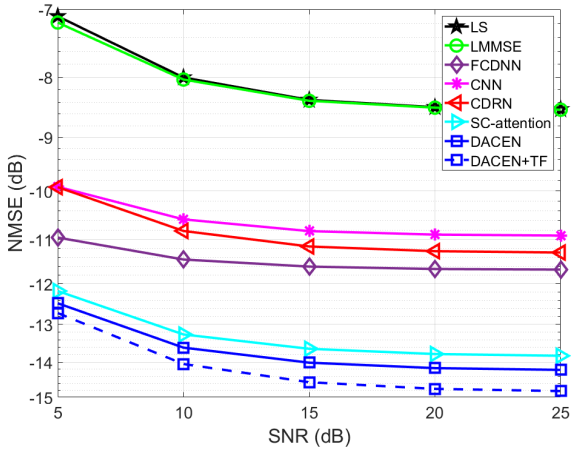

图 7：使用$\rho_L = 4/52$时不同信噪比下的归一化均方误差性能。

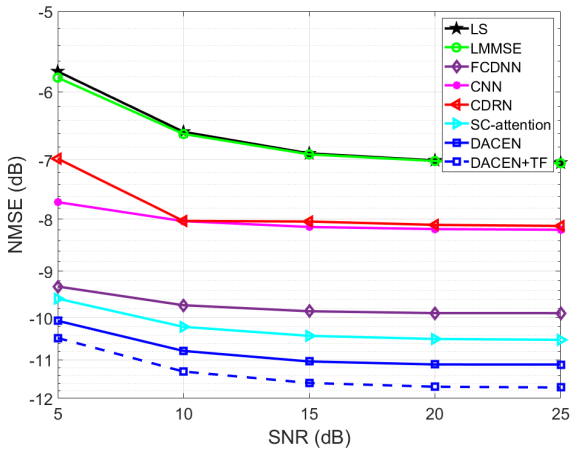

图 8：使用$\rho_L = 2/52$时不同信噪比下的归一化均方误差性能。

表 IV：所提出的方法与其他基于深度学习的方法的计算复杂度比较。

方法	计算复杂度	
	时间复杂度（浮点运算次数）	空间复杂度（参数）
全连接深度神经网络	19点00米	18。90 M
CNN	2.50 千兆符号/秒	19.52 M
CDRN	3.40 G	26.58 米
SC注意力	3.74 G	29。21 M
DACEN	2.31 G	17.01 兆
DACEN+TF	2.31 吉	17.01 兆

分别从 DACEN 中移除所有空间注意力模块和所有时间注意力模块。此外，我们将 DACEN 中的空间注意力模块替换为二维卷积层（即滤波器尺寸为 $K_S \times K_S = 3 \times 3$ 的空间卷积层（SConv））。此外，我们将 DACEN 中的时间注意力模块替换为一维卷积层（即滤波器尺寸为 $K_T = 3$ 的时间卷积层（TConv））。为简便起见，得到的四种改进深度神经网络模型分别记为**‘不含空间注意力模块’**、**‘不含时间注意力模块’**、**‘空间卷积-时间注意力模块’**和**‘空间注意力模块-时间卷积’**。实验结果如图 9 所示，这些改进深度神经网络模型的计算复杂度分析则列于

表 V。此外，TAM、TConv、SAM 和 SConv 的逐层计算复杂度见表VI。

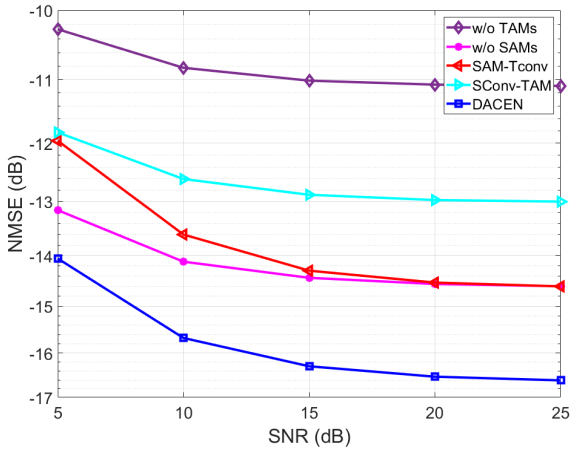

图9：使用 $\rho_L = 6/52$ 时，改进的深度神经网络模型在不同信噪比下的归一化均方误差性能。

从图9可见，移除所有空间注意力模块（‘**无空间注意力模块**’）或所有时间注意力模块（‘**无时间注意力模块**’）都会导致性能明显下降，这表明部署时间注意力模块和空间注意力模块的必要性和重要性。用 SConv 替换空间注意力模块，或用 TConv 替换时间注意力模块后，‘**SConv-TAM**’和‘**空间注意力模块-时间卷积**’的表现也都不如所提出的 DACEN。这进一步凸显了所设计的空间注意力模块和时间注意力模块的有效性。

根据表V 和表VI，尽管时间注意力模块的计算复杂度接近 TConv 的两倍，但它能够学习不同延迟路径的时间依赖关系，并带来显著的性能提升（即，与‘空间注意力模块-时间卷积’相比，DACEN 平均可获得约**2** 分贝增益）空间注意力模块的估计性能也优于 SConv（即，与‘3分贝增益）相比，DACEN 平均可获得约**SConv-TAM**’），而计算复杂度仅约为 SConv 的 $1/5$，由此证明了空间注意力模块的有效性和计算效率。

表V：消融研究中改进的深度神经网络模型的计算复杂度比较。

方法	计算复杂度	
	时间复杂度（浮点运算次数）	空间复杂度（参数）
DACEN	2.31 G 17.01M	
TAM	1.76 吉高氏	12.74 M
SAM	0.55 吉	4。27 M♀♀♀♀
空间卷积-时间注意力模块	3.96 G	29.92 M
空间注意力模块-时间卷积	1.35 G	10.57 兆

六、结论

本文提出了一种基于 DACEN 的信道估计方法，适用于采用低密度导频的大规模 MIMO 系统。DACEN 由两个注意力模块组成，即 TAM 和 SAM，用于联合学习大规模 MIMO 信道的时空域特征。TAM 旨在学习不同延迟路径之间的时间依赖关系

表 VI：TAM、TConv、SAM 和 SConv 的逐层计算复杂度。

方法	计算复杂度	
	时间复杂度（浮点运算次数）	空间复杂度（参数）
TAM	218。47 米	1.58 M
TConv	100.73 M英寸	786.94 k
SAM	67。18 M	525.31 千克
SConv	302.06 M M	2.36 M

，而 SAM 旨在学习不同天线之间 CSI 的相关性。通过利用大规模 MIMO 信道的空间相关性，与卷积神经网络相比，SAM 的时间复杂度和空间复杂度都低得多。这对大规模 MIMO 系统具有重要价值，尤其是在部署越来越大的天线阵列时。此外，为进一步提高信道估计精度，我们还提出了一种基于 DACEN 的参数实例迁移学习方法，将模型训练阶段从高密度导频中学到的信道知识迁移到低密度导频场景。基于各种导频密度设置的实验结果表明，所提出的联合空时域特征提取网络行之有效，且所提出的基于 DACEN 的低开销信道估计方法具有优越性。

VII. 致谢

我们感谢 IMT-2020(5G)、中国信通院和 OPPO 提出了降低当前 5G 系统导频密度这一研究问题。本研究部分工作在 SICC 完成，该中心由澳门大学的 SKL-IOTSC 提供支持。

参考文献

[1] T. L. Marzetta，“大规模 MIMO：简介”，贝尔实验室技术期刊，第 20 卷，第 11–22 页，2015年。[2] E. G. Larsson、O. Edfors、F. Tufvesson 和 T. L. Marzetta，“面向下一代无线系统的大规模 MIMO”，*IEEE* 通信杂志，第 52 卷，第 2 期，第 186–195 页，2014年2月。[3] H. Nguyen-Le 和 T. Le-Ngoc，“用于 OFDM 传输的导频辅助 CFO 与双选择性信道联合估计”，*IEEE* 广播汇刊，第 56 卷，第 4 期，第 514–522 页，2010年12月。[4] X. Wang、P. Ho 和 Y. Wu，“具有抑制特征的 OFDM系统的鲁棒信道估计与码间干扰消除”，*IEEE* 通信领域精选期刊，第 23 卷，第 5 期，第 963–972 页，2005年5月。[5] P. Chen 和 H. Kobayashi，“OFDM系统的最大似然信道估计与信号检测”，载于 *2002* 年 *IEEE* 国际通信会议（*ICC*）论文集，2002年4月，第 1640–1645 页。[6] N. Geng、X. Yuan 和 L. Ping，“OFDM系统的双对角 LMMSE 信道估计”，*IEEE* 信号处理汇刊，第 60 卷，第 9 期，第 4734–4746 页，2012年9月。[7] L. Huang、J. W. M. Bergmans 和 F. M. J. Willems，“通过角度域处理实现低复杂度的基于 LMMSE 的 MIMO-OFDM 信道估计”，*IEEE* 信号处理汇刊，第 55 卷，第 12 期，第 5668–5680 页，2007年12月。[8] J. Lee、G.-T. Gil 和 Y. H. Lee，“毫米波通信中混合 MIMO 系统基于正交匹配追踪的信道估计”，*IEEE* 通信汇刊，第 64 卷，第 6 期，第 2370–2386 页，2016年6月。[9] X. Wu、G. Yang、F. Hou 和 S. Ma，“毫米波 FDD 大规模 MIMO 系统的低复杂度下行信道估计”，*IEEE* 无线通信快报，第 8 卷，第 4 期，第 1103–1107 页，2019年8月。[10] X. Wu、X. Yang、S. Ma、B. Zhou 和 G. Yang，“UPA 辅助毫米波大规模 MIMO 物联网系统的混合信道估计”，*IEEE* 物联网期刊，第 9 卷，第 4 期，第 2829–2842 页，2022年2月。

[11] G. Wunder、S. Stefanatos、A. Flinth、I. Roth 和 G. Caire，“多用户宽带大规模 MIMO 的低开销分层稀疏信道估计”，IEEE《无线通信汇刊》，第 18 卷，第 4 期，第 2186–2199 页，2019年4月。[12] X. Chen、C. Deng、B. Zhou、H. Zhang、G. Yang 和 S. Ma，“利用超分辨率网络实现大规模MIMO系统的高精度 CSI反馈”，IEEE 无线通信快报，第 11 卷，第 1 期，第 141–145 页，2022年。[13] Q. Xue、Y.-J. Liu、Y. Sun、J. Wang、L. Yan、G. Feng 和 S. Ma，“通过联邦强化学习实现超密集毫米波网络中的波束管理：一种智能且安全的方法”，IEEE《认知通信与网络汇刊》，第 1–13 页，2022年。[14] H. He、C.-K. Wen、S. Jin 和 G. Y. Li，“用于MIMO检测的模型驱动深度学习”，IEEE 信号处理汇刊，第 68 卷，第 1715、2020 1702 页。[15] H. Ye、G. Y. Li 和 B.-H. Juang，“深度学习在OFDM系统信道估计与信号检测中的作用”，IEEE 无线通信快报，第 7 卷，第 1 期，第 114–117 页，2018年2月。[16] P. Dong、H. Zhang、G. Y. Li、I. S. Gaspar 和 N. NaderiAlizadeh，“毫米波大规模MIMO系统的基于深度 CNN 的信道估计”，IEEE《信号处理精选主题期刊》，第 13 卷，第 5 期，第 989–1000 页，2019年9月。[17] M. Soltani、V. Pourahmadi、A. Mirzaei 和 H. Sheikhzadeh，“基于深度学习的信道估计”，IEEE《通信快报》，第 23 卷，第 4 期，第 652–655 页，2019年4月。[18] M. Honkala、D. Korpi 和 J. M. J. Huttunen，“DeepRx：全卷积深度学习接收机”，IEEE《无线通信汇刊》，第 20 卷，第 6 期，第 3925–3940 页，2021年6月。[19] T. Zeng、J. Li、M. Hu、S. Hou 和 Q. Zhang，“利用复数DnCNN实现更高性能的信道估计”，IEEE《通信快报》，第 24 卷，第 1 期，第 198–201 页，2020年1月。[20] Y. Jin、J. Zhang、B. Ai 和 X. Zhang，“利用卷积盲去噪网络进行毫米波大规模MIMO的信道估计”，IEEE《通信快报》，第 24 卷，第 1 期，第 95–98 页，2020年1月。[21] C. Liu、X. Liu、D. W. K. Ng 和 J. Yuan，“智能反射面辅助的多用户通信中的信道估计深度残差学习”，IEEE《无线通信汇刊》，第 21 卷，第 2 期，第 898–912 页，2022年2月。[22] Z. Liu、L. Zhang 和 Z. Ding，“通过深度学习突破大规模MIMO通信中的信道估计瓶颈”，IEEE《无线通信》，第 27 卷，第 5 期，第 104–111 页，2020年10月。[23] P. Jiang、C.-K. Wen、S. Jin 和 G. Y. Li，“MIMO-OFDM系统的基于双 CNN 的信道估计”，IEEE 通信汇刊，第 69 卷，第 9 期，第 5859–5872 页，2021年9月。[24] Y. Yang、F. Gao、Z. Zhong、B. Ai 和 A. Alkhateeb，“基于深度迁移学习的 FDD大规模MIMO系统下行信道预测”，IEEE 通信汇刊，第 68 卷，第 12 期，第 7485–7497 页，2020年12月。[25] W. Alves、I. Correa、N. Gonz´alez-Prelcic 和 A. Klautau，“低分辨率毫米波MIMO中特定站点的深度迁移学习信道估计”，IEEE 无线通信快报，第 10 卷，第 7 期，第 1424–1428 页，2021年7月。[26] 3GPP，“新空口；物理信道与调制”，第三代合作伙伴计划（3GPP），技术规范（TS）38.211 V17.4.0，2023年1月。[27] S. Noh、M. D. Zoltowski、Y. Sung 和 D. J. Love，“大规模 MIMO 系统的信道估计导频波束图案设计”，IEEE《信号处理精选主题期刊》，第 8 卷，第 5 期，第 787–801 页，2014年10月。[28] Z. Gao、L. Dai、W. Dai、B. Shim 和 Z. Wang，“面向 FDD大规模MIMO 的基于结构化压缩感知的时空联合信道估计”，IEEE 通信汇刊，第 64 卷，第 2 期，第 601–617 页，2016年2月。[29] C.-K. Wen、W.-T. Shih 和 S. Jin，“大规模 MIMO CSI反馈的深度学习”，IEEE 无线通信快报，第 7 卷，第 5 期，第 748–751 页，2018年10月。[30] J. Guo、C.-K. Wen、S. Jin 和 G. Y. Li，“用于大规模 MIMO CSI反馈的基于卷积神经网络的多速率压缩感知：设计、仿真与分析”，IEEE《无线通信汇刊》，第 19 卷，第 4 期，第 2827–2840 页，2020年4月。[31] ——，“大规模 MIMO 系统中基于深度学习的 CSI反馈综述”，IEEE 通信汇刊，第 70 卷，第 12 期，第 8017–8045 页，2022年12月。[32] 3GPP，“新空口；数据物理层过程”，第三代合作伙伴计划（3GPP），技术规范（TS）38.214 V16.1.0，2020年4月。[33] W. Liu、W. Tian、H. Xiao、S. Jin、X. Liu 和 J. Shen，“EVCsiNet：3GPP链路级信道下基于特征向量的 CSI反馈”，IEEE 无线通信快报，第 10 卷，第 12 期，第 2688–2692 页，2021年。

[34] M. Chen、J. Guo、C.-K. Wen、S. Jin、G. Y. Li 和 A. Yang，“基于深度学习的大规模 MIMO 隐式CSI反馈”，IEEE 通信汇刊，第70卷，第2期，第935–950页，2022年2月。[35] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser 和 I. Polosukhin，“Attention is All You Need”，载于《神经信息处理系统进展》第31届年会论文集（NIPS）2017，2017年5月，第5998–6008页。[36] B. Zhou、G. Yang、Z. Shi 和 S. Ma，“用于COVID-19预测的可解释时间注意力网络”，《应用软计算》，第120卷，文章编号108691，2022年5月。[37] A. Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit 和 N. Houlsby，“一张图像胜过16×16个词：用于大规模图像识别的Transformer”，载于2021年国际学习表征会议（ICLR）论文集，2021年。[在线]。网址：https://openreview.net/forum?id=YicbFdNTTy [38] J. L. Ba、J. R. Kiros 和 G. E. Hinton，“层归一化”，arXiv预印本 arXiv:1607.06450，2016年7月。[39] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，载于2016年IEEE计算机视觉与模式识别会议（CVPR）论文集，2016年6月，第770–778页。[40] S. Woo、J. Park、J.-Y. Lee 和 I. S. Kweon，“CBAM：卷积块注意力模块”，载于2018年欧洲计算机视觉会议（ECCV）论文集，2018年，第3–19页。[41] S. J. Pan 和 Q. Yang，“迁移学习综述”，IEEE《知识与数据工程汇刊》，第22卷，第10期，第1345–1359页，2010年10月。[42] K. Weiss、T. M. Khoshgoftaar 和 D. Wang，“迁移学习综述”，《大数据期刊》，第3卷，文章编号9，2016年。[43] W. Zhang、L. Deng、L. Zhang 和 D. Wu，“负迁移综述”，IEEE/CAA《自动化学报》，第1–25页，2022年。[44] 3GPP，“0.5至100吉赫兹频段的信道模型研究”，第三代合作伙伴计划（3GPP），技术报告（TR）38.901 V17.0.0，2022年3月。[45] MathWorks，“MATLAB 5G工具箱”，2023年2月。 [在线]。网址：https://ww2.mathworks.cn/en/products/5g.html [46] A. Paszke、S. Gross、F. Massa、A. Lerer、J. Bradbury、G. Chanan、T. Killeen、Z. Lin、N. Gimelshein、L. Antiga、A. Desmaison、A. K¨opf、E. Z. Yang、Z. DeVito、M. Raison、A. Tejani、S. Chilamkurthy、B. Steiner、L. Fang、J. Bai 和 S. Chintala，“PyTorch：一种命令式风格的高性能深度学习库”，载于2019年神经信息处理系统进展（NeurIPS）论文集，2019年，第8024–8035页。[47]D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，载于2015年国际学习表征会议（ICLR）论文集，2015年5月。 [在线]。网址：http://arxiv.org/abs/1412.6980 。Y. Li，“具有多发射天线的OFDM系统的最优训练序列”，载于2000年IEEE全球通信会议论文集，第3卷，2000年11月，第1478–1482页，第3卷。[49] Y. Choi、J. H. Bae 和 J. Lee，“面向OFDM系统的低复杂度二维LMMSE信道估计”，载于2015年IEEE第82届车辆技术会议（VTC2015-Fall）论文集，2015年9月，第1–5页。[50] M. Liu、X. Li、B. Ning、C. Huang、S. Sun 和 C. Yuen，“双RIS辅助大规模MIMO系统的基于深度学习的信道估计”，IEEE 无线通信快报，第12卷，第1期，第70–74页，2023年1月。[48]



Binggui Zhou 分别于2018年和2021年获得中国珠海暨南大学工学学士学位，以及中国澳门澳门大学理学硕士学位。目前，他正在澳门大学攻读电气与计算机工程博士学位。他还担任中国珠海暨南大学智能系统科学与工程学院的研究助理。他的研究兴趣包括人工智能、智能无线通信以及时空数据挖掘。



Xi Yang 于2013年、2016年和2019年分别获得中国南京东南大学的理学学士、工学硕士和博士学位。2020年7月至2022年7月，她在中国澳门大学智慧城市物联网国家重点实验室从事博士后研究。她目前是中国上海华东师范大学通信与电子工程学院的紫江青年学者。她目前的研究兴趣包括超大孔径阵列（ELAA）系统、毫米波通信以及无线通信系统原型设计。



杨光华于2006年获得香港大学电子与电气工程博士学位。2006年至2013年，他在香港大学担任博士后研究员和研究助理。自2017年4月起，他加入暨南大学，目前任智能系统科学与工程学院正教授。他的研究兴趣主要涵盖人工智能与通信领域。



Shaodan Ma 于1999年和2002年分别获得南开大学（中国天津）的理学和经济学双学士学位，以及电子工程工学硕士学位，并于2006年获得香港大学（中国香港）的电气与电子工程博士学位。2006年至2011年，她在香港大学担任博士后研究员。自2011年8月起，她任职于澳门大学，目前为教授。她的研究兴趣包括阵列信号处理、收发机设计、定位、感知与通信一体化、毫米波通信和大规模 MIMO。她曾担任多个会议的专题研讨会联合主席，包括IEEE国际通信会议（IEEE ICC）2021年、2019年和2016年会议，IEEE/CIC国际通信大会（IEEE/CIC ICCC）2019年会议，IEEE全球通信大会（IEEE GLOBECOM）2016年会议等。目前，她担任IEEE《无线通信汇刊》、IEEE通信汇刊、IEEE通信快报以及《通信与信息网络学报》的编辑。



Feifei Gao （M’09-SM’14-F’20）于2002年获得中国西安交通大学工学学士学位，2004年获得加拿大安大略省汉密尔顿麦克马斯特大学理学硕士学位，2007年获得新加坡新加坡国立大学博士学位。自2011年起，他加入中国北京清华大学自动化系，现任副教授。

高教授的研究兴趣包括通信信号处理、阵列信号处理以及人工智能辅助通信。他已撰写或合著了150多篇经同行评审的IEEE期刊论文和150多篇IEEE会议论文集论文，这些论文在谷歌学术上的引用次数超过16000次。Gao教授曾担任《IEEE通信汇刊》、《IEEE无线通信汇刊》、《IEEE信号处理选题期刊》（首席客座编辑）、《IEEE认知通信与网络汇刊》、《IEEE信号处理快报》（高级编辑）、《IEEE通信快报》（领域编辑）、《IEEE无线通信快报》和《中国通信》的编辑。他还曾担任2019年IEEE通信会议（ICC）、2018年IEEE春季车辆技术会议（VTC）、2015年IEEE通信会议（ICC）、2014年IEEE全球通信会议（GLOBECOM）和2014年IEEE秋季车辆技术会议（VTC）的专题研讨会联合主席，并担任50多个IEEE会议的技术委员会委员。