arXiv:2303.12653v3 [cs.IT] 2 Aug 2024

基于自监督混合深度学习的鲁棒毫米波波束成形

Fenghao Zhu、Bohao Wang、Zhaohui Yang、Chongwen Huang、Zhaoyang Zhang、George C. Alexandropoulos、Chau Yuen 和 M´erouane Debbah

摘要—近年来，大规模天线阵列波束成形得到了广泛应用，被认为是 5G 和即将到来的 6G 的重要组成部分。因此，人们采用了多种技术来提升其性能，例如深度学习、先进优化算法等。尽管在以往许多结合深度学习的研究场景中，其性能颇具吸引力，但当环境或数据集发生变化时，性能通常会迅速下降。因此，设计具有强鲁棒性的有效波束成形网络，是智能无线通信领域尚待解决的问题。本文提出了一种鲁棒的波束成形自监督网络，并在两种不同类型、涵盖多种场景的数据集上对其进行了验证。仿真结果表明，所提出的结合混合学习的自监督网络，在经典的 DeepMIMO 和新型 WAIR-D 数据集上均表现良好，且在各种环境下具有很强的鲁棒性。此外，我们还阐述了这一混合学习方法的原理，以说明其合理性，为将其应用于更多类型的数据集提供了指导。

索引词——**稳健波束成形、深度神经网络、混合学习、自监督学习。**

I. 引言

近年来，毫米波（mmWave）波束成形因其极宽的带宽和形成窄波束的能力而备受关注，这能够显著提高通信系统的吞吐量。窄波束能够改善波束成形的方向性，使能量更加集中，从而提高信噪比。过去几十年来，波束成形技术已得到深入研究 [1]。具体而言，最初提出了穷举搜索和自适应分层搜索来解决这一问题，但代价是训练开销和时延较大。主化最小化算法是一种有效的求取次优解的方法 [2], , 但通常需要精确的信道信息，这会导致巨大的训练开销。

机器学习（ML）擅长利用历史经验解决新的类似问题，先前的研究表明，基于 ML 的技术可用于预测毫米波波束 [3]。此外，近期一项研究提出了一个适用于多输入单输出（MISO）波束成形的通用框架 [4], , 但这些研究都无法迁移到不同场景，甚至无法适用于开放环境。这也是大多数基于 ML 的波束成形方法的缺陷。因此，构建一个稳健的波束成形框架

F. Zhu、B. Wang、Z. Yang、C. Huang 和 Z. Zhang 任职于中国杭州 310007 浙江大学信息科学与电子工程学院（电子邮箱：chongwenhuang@zju.edu.cn）。G. C. Alexandropoulos 任职于希腊雅典 15784 Panepistimioúpoli Ilissia，雅典国立卡波季斯特里亚大学信息学与电信系（电子邮箱：alexandg@di.uoa.gr）。C. Yuen 任职于新加坡南洋理工大学电气与电子工程学院。M. Debbah 任职于阿联酋阿布扎比 P O Box 127788 哈利法科技大学（电子邮箱：merouane.debbah@ku.ac.ae）。

是一个非常有意义且重要的课题，尤其是在该框架能够适用于不同类型的数据集时。首先，我们将介绍这两种不同类型的数据集。众所周知，DeepMIMO 是无线通信领域应用最广泛的开源数据集，它由三维射线追踪和虚拟场景生成技术构建而成[5]。由于其完整性，该数据集能够模拟最经典的场景。然而，其缺点也十分明显：它基于虚拟生成的场景，可能无法为真实通信提供准确的仿真。此外，该数据集对每种场景仅提供一种设置，因此训练得到的模型无法适应各种建筑设置。

最近，另一个名为 WAIR-D 的开源数据集被提出，以推动相关研究 [6]。它与 DeepMIMO 最大的区别在于，该数据集基于真实场景地图。此外，数据集的生成和保存方式也有所不同。WAIR-D 的主要特征可概括如下：

- 真实场景：从全球 40 多个大城市的真实场景地图中随机选取了 10000 个场景，示例见图。1，其中建筑物的布局信息直接从实际地图中提取。
- 灵活的参数：提供三维射线追踪参数，用户可以自由设置这些参数，从而生成无线人工智能任务所需的数据。
- 用户友好：为帮助用户快速上手，提供了数据生成、预处理、样例和训练代码。
- 适用于低内存计算机：可调整数据生成所用的线程数，以减少内存占用，并降低读取和训练的时间开销。

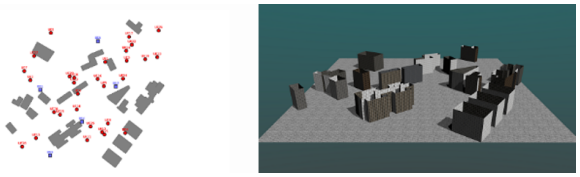


图 1. 示例环境及建筑物和基站（蓝色）/用户设备（红色）的位置俯视图（左），相同建筑物的鸟瞰图（右）

在这项工作中，我们将介绍一种基于这两类不同数据集的鲁棒波束成形自监督方案，该方案无需标签，所提出的波束成形网络即可通过训练，智能地学习如何优化频谱效率。最后，我们在各种场景下验证了所提方案的鲁棒性。主要贡献可总结如下：

- 数据集内的鲁棒性：我们设计了一种自监督波束成形神经网络，用于生成模拟移相器，而非数字移相器，这更符合毫米波的特点。面对不同的测试用例，该网络仍能输出经过优化的波束成形器，这表明该神经网络在特定数据集内具有鲁棒性。
- 跨数据集的鲁棒性：我们提出混合预训练，以获得更具鲁棒性的波束成形网络，使其能够适应不同的环境。此外，我们还阐述了混合学习的合理性原理，为将其应用于更多类型的数据集提供指导。
- 多个数据源的性能评估：我们在两类具有不同场景的数据集上验证了所提出的方案。仿真结果表明，采用混合学习的自监督网络在经典的 DeepMIMO 数据集和新型 WAIR-D 数据集上均表现良好，并且在各种环境下具有较强的鲁棒性。

II. 系统模型

本节首先介绍所研究的毫米波通信场景的系统模型，然后给出一个速率最大化优化问题。我们考虑基于多输入多输出的毫米波下行链路通信系统 [7], ，其中基站（BS）配备了一个由 N_t 个天线单元组成的阵列，以及 N_{RF} 条射频链。在接收端，有 N_r 个单天线移动用户。由于毫米波衰减严重，我们仅考虑基站与用户之间的视距（LOS）主导路径。我们将发送符号表示为$\mathbf{x} \in C^{N_{RF}\times 1}$, 其平均能量为 $E\left\{|\mathbf{x}\mathbf{x}^H|\right\} = I$。我们采用窄带块衰落信道模型，其中第 k-th个用户观测到的接收信号为

$$y = \mathbf{h}_k^H \mathbf{v}_k \mathbf{x} + n, \tag{1}$$

其中，$\mathbf{h}$ 是表示基站与用户之间毫米波信道的 $N_t \times 1$ 向量，$\mathbf{v} \in C^{N_t \times N_{RF}}$ 是波束成形矩阵。 n 是加性噪声，服从均值为零、协方差为 σ^2的圆对称复高斯分布。

关于信道模型，我们目前生成了两个数据集：DeepMIMO 和 WAIR-D。DeepMIMO 的时域信道矩阵 $\mathbf{h}_d$ 由 L 条信道路径组成，每条路径都有一个时延 $\tau_l \in R$, 以及方位角/俯仰角到达角（AoA） θ_l, φ_l。令 ρ 表示用户与基站之间的路径损耗， $p(\tau)$表示适用于 T_S间隔信号、在 τ 秒处取值的脉冲成形函数。因此，信道向量可表示为

$$\mathbf{h}_d = \sqrt{\frac{M}{\rho}} \sum_{l=1}^{L} \alpha_l p(dT_S - \tau_l)\mathbf{a}(\theta_l, \varphi_l), \tag{2}$$

其中，$\mathbf{a}(\theta_l, \varphi_l)$ 是基站在到达角 θ_l, φ_l 处的阵列响应向量。相比之下，WAIR-D 具有相同的信道模型，但它是在大尺度上基于真实地图构建的，因此能够更准确地模拟真实场景。 k-th用户可按如下方式获得频谱效率 [8],

$$R_k(\mathbf{v}_k) = \log(1 + \frac{\|\mathbf{h}_k^H \mathbf{v}_k\|^2}{\sigma^2 + \sum_{j\neq k} \|\mathbf{h}_k^H \mathbf{v}_j\|^2}). \tag{3}$$

其中，波束成形向量满足功率约束$\sum_k \|\mathbf{v}_k^H \mathbf{v}_k\| \leq P$。信噪比（SNR）可定义为 10 lg($P/\sigma^2$)。由于此任务可表述为映射 $\mathbf{v}_k = \mathcal{V}(\mathbf{h}_k, P)$，因此优化问题可写为

$$\max_{\mathbf{v}_k} \mathbb{E}_{\mathbf{h},P}[\sum_k R_k(\mathcal{V}(\mathbf{h}_k, P))] \tag{4}$$

$$s.t. \sum_k \|\mathcal{V}(\mathbf{h}_k, P)^H \mathcal{V}(\mathbf{h}_k, P)\| \leq P, \tag{5}$$

其中，$\mathbb{E}_{\mathbf{h},P}[\cdot]$ 表示对信道样本取平均以及总功率约束。本文的主要思路是关注频谱效率如何随信噪比变化，并利用信道状态信息（CSI）预测最佳波束成形向量，尽管其估计并不完美。信道估计的准确性可通过导频噪声功率比（PNR）衡量；此外， [8] 已表明，深度神经网络在高PNR下可以逼近理论上界。为验证网络在不同数据集上的鲁棒性，同时降低估计偏差的影响，我们将PNR设为合适的20 dB。该场景中

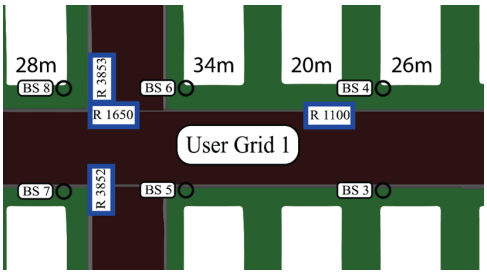


图 2. DeepMIMO 数据集中“O1”场景的俯视图。

我们在 DeepMIMO 中选择的是 O1 60。对于 WAIR-D 数据集，WAIR-D 的场景 1 有 30 个用户和 4 个基站，用户数量不足，因此我们选择 WAIR-D 的场景 2 进行实验。训练地图的详细信息分别如图。2 和图。3 所示。

相关研究 [9] 指出，此类波束成形任务中的多链路干扰有限，而将干扰纳入考量会需要更复杂的网络架构和更高的计算复杂度。此外，由于毫米波波束成形本身的特性，预计多链路干扰对预测性能的影响较小 [10]。因此，我们的仿真不考虑用户间的干扰，尽管我们的模型具有通用性，可用于多个用户。

III. 所提出的自监督波束成形器

本节介绍无监督机器学习架构，并探讨如何组合数据集以提高所提出网络架构的混合训练效率。该架构如图。4所示，主要由三个部分组成：输入块、特征提取块和输出块。这种设计能够捕捉各层之间的关系，并有效缓解过拟合，同时不对输入数据作特殊假设。随后，我们提出混合训练理论，以推导模型的鲁棒性，使模型能够适用于不同环境。具体内容将在以下小节中介绍。

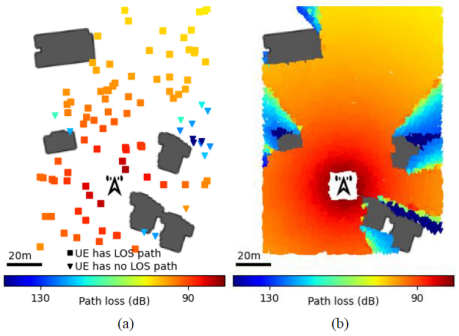


图 3. WAIR-D 场景 2 案例 1 的俯视图。(a) 显示全部 10,000 个用户设备中的 100 个；(b) 显示全部 10,000 个用户设备

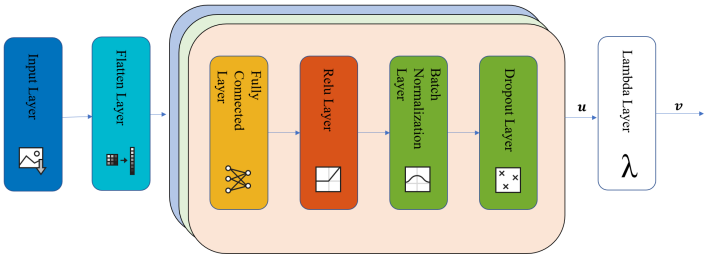


图 4. 波束成形网络的架构。

A. 网络架构

WAIR-D 的 CSI 通常以图像形式提供，以节省存储空间；图像的 x 轴和 y 轴分别表示载波数和天线数，颜色深度则表示信道频率响应（CFR）。DeepMIMO 的 CSI 以复矩阵形式提供，我们可以将其实部和负部拆分后拼接，组成图像。因此，我们可以采用图像作为输入形式。经过展平层后，输入数据变为一维，便可由下一层处理。第二个模块主要由三个较小的模块组成，其中全连接层主要用于提取特征，其后依次为 ReLU 层、批归一化层和丢弃层。批归一化层可以将偏置分布归一化为标准分布。这样一来，梯度会增大，学习收敛速度会加快，梯度消失问题也会得到缓解。相关研究表明，批归一化层应置于 ReLU 层和丢弃层之间 [11]，因此我们采用这一顺序设计，以最大限度地提高鲁棒性。选择三个较小模块的原因是，我们通过实验发现，这样能够在复杂度和性能之间取得最佳平衡。仿真中的详细参数见表 I。

最后一个丢弃层输出 $2N_tN_rN_{RF}$ 个实数，对应于 N_r 个复向量，而$\mathbf{u}=[\mathbf{u}_1,\mathbf{u}_2\cdots,\mathbf{u}_{N_r}]$是其组合。在 $\sum_k\|\mathbf{v}_k^H\mathbf{v}_k\| \le P$ 的约束下，λ 层利用生成的复向量 $\mathbf{u}$ 输出最终的波束成形向量 $\mathbf{v}=[\mathbf{v}_1,\mathbf{v}_2\cdots,\mathbf{v}_{N_r}]$ 作为

$$\mathbf{v}_k=\sqrt{\frac{P}{\sum\limits_{1\le j\le N_r}\|\mathbf{u}_j\|^2}}\mathbf{u}_k. \tag{6}$$

作为自监督学习的特点，我们在训练过程开始前不需要

表 I 网络详细信息

层名称	输出尺寸	参数
输入层	2×64	0
全连接层 1	320×1	41280
批归一化层 1	320×1	640
全连接层 2	320×1	102720
批归一化层 2	320×1	640
全连接层 3	128×1	41088
批归一化层 3	128×1	256
Lambda 层	64×1	0

标签，只需要信道状态信息和功率 P。由于我们直接输出波束成形向量和传输速率，因此训练损失为平均传输速率的相反数。我们采用随机梯度下降法 (SGD) 来逼近最大传输速率，损失函数的形式为

$$\text{Loss}=-\frac{1}{\mathrm{N_r M}}\sum_{\mathrm{i=1}}^{\mathrm{M}}\sum_{\mathrm{k=1}}^{\mathrm{N_r}}\mathrm{R_{k,i}}(\mathbf{v_k}). \tag{7}$$

M 表示训练批次的大小，而$R_{k,\ i}$表示 i-th 个样本中第 k-th 个用户的频谱效率，且每个样本包含 N_r 个用户的信道数据。

B. 混合训练理论

在本节中，我们提出一种用于波束成形的混合训练方法，并给出解释该方法的理论。我们假设训练数据由 K 种数据集组成，总计包含n 个样本，每个数据集所占的比例为 q_k，而 p_k 是某个特定原始数据集对应的信道状态信息。显然，$\sum_{k=1}^{K}q_k=1$，且数据集 k 中的信道向量数量为 nq_k。因此，该混合数据集可定义为$p=\sum_{k=1}^{K}q_kp_k$，且$\mathbf{q}=(q_1,q_2\cdots,q_K)$。我们假设神经网络的参数为 $\mathbf{\Theta}$；输入的信道状态信息为$\mathbf{h}$，输出的频谱效率为 R。$\ell(\mathbf{\Theta};\mathbf{h}\ R)$是训练样本损失，用于评估预测值与实际最大值之间的差距，尽管在自监督学习中不需要实际最优值，且无法直接获取。因此，最优参数的搜索可表示为

$$\mathbf{\Theta}(n,\mathbf{q})_{optimal}=\underset{\mathbf{\Theta}}{\operatorname{argmin}}\,\mathbb{E}_{n,\mathbf{q}}[\ell(\mathbf{\Theta};\mathbf{h},R)]. \tag{8}$$

由于我们只关注 n 和$\mathbf{q}$ 对训练损失的影响，因此可以构造一个函数来评估额外的平均训练损失，

$$L(n,\mathbf{q})=\mathbb{E}_{n,\mathbf{q}}[\ell(\mathbf{\Theta};\mathbf{h},R)]-\{\mathbb{E}_{n,\mathbf{q}}[\ell(\mathbf{\Theta};\mathbf{h},R)]\}_{min}. \tag{9}$$

早期研究表明，如果$\mathbf{q}$ 固定，$L(n,\mathbf{q})$与 $log(n)$之间存在对数线性关系 [12]，，因此额外损失可表示为

$$\log(L(n,\mathbf{q}))\approx\alpha(\mathbf{q})\log(n)+C(\mathbf{q}). \tag{10}$$

此外，我们发现，在特定任务中，$\alpha(\mathbf{q})$是一个常数，不受数据组成的影响 [13]，，这使 n 与$\mathbf{q}$相互解耦，帮助我们专注于比例参数$\mathbf{q}$。当n 固定时，$\alpha(\mathbf{q})\log(n)$变为常数，因此额外的平均训练损失可以仅用 $C(\mathbf{q})$来表示。在

与此同时，我们提出了一个有理函数 [14] 来拟合 $C(\mathbf{q})$，写作

$$C_{\lambda}(\mathbf{q})=\sum_{i=1}^{d}\left(\sum_{k=1}^{K}\lambda_{ik}q_k\right)^{-1}, \tag{11}$$

其中，λ_{ik} 是由该模型确定的参数，d 是信道向量$\mathbf{h}$中的元素数量。该公式已通过实证和理论方法得到有效验证，因此我们可以独立研究$\mathbf{q}$ 对训练损失的影响。但求得参数 λ_{ik} 的过程并非那么直接。我们首先假设损失 ℓ 二阶可微（若网络能够反向传播，则该条件成立），且 n 固定且足够大。因此，我们可以定义如下两个变量：

$$\sum\nolimits_{k}=\mathbb{E}_{p_k}[\nabla(\ell(\boldsymbol{\Theta};\mathbf{h},R))]. \tag{12}$$

$$\sum\nolimits^{*}=\mathbb{E}_{p^*}[\nabla(\ell(\boldsymbol{\Theta};\mathbf{h},R))]. \tag{13}$$

此处，$\nabla(\ell(\boldsymbol{\Theta};\mathbf{h}\ R))$表示 $\ell(\boldsymbol{\Theta};\mathbf{h}\ R)$的海森矩阵，其中$\mathbf{h}$的各个元素为变量。而 p^* 表示按一定比例混入的测试数据集的信道状态信息。因此，$C(\mathbf{q})$可近似表示为 [14]

$$C(\mathbf{q})\approx Tr\left(\sum\nolimits^{*}\left(\sum_{k}\left(q_k\sum\nolimits_{k}\right)\right)^{-1}\right), \tag{14}$$

其中，Tr 表示矩阵的迹。这表明，$C(\mathbf{q})$不仅与训练数据集有关，也与验证数据集有关。若$\sum^{*}$可对角化为 $P^{-1}\sum^{*}P=D^{*}$，且$\sum_{k}$也可通过相同的正交矩阵 P 近似对角化为 $P^{-1}\sum_{k}P=D_k+O_k$，并且 O_k足够小，则将（12）和（13）代入（14），即可通过以下公式求得参数 λ_{ik} ：

$$C(\mathbf{q})\approx\sum_{i=1}^{d}\frac{1}{\sum_{k=1}^{K}q_k\frac{D_{k,ii}}{D^{*}_{ii}}}+o\left(\|\sum_{k=1}^{K}q_kO_k\|_F\right). \tag{15}$$

$D_{k,\ ii}$ 表示 D_k对角线上的第 i-th 个元素。现在，我们得到式（11）中的 λ_{ik}，其等于 $D_{k}\frac{ii}{D^{*}_{ii}}$，这可以通过上述计算海森矩阵并进行矩阵对角化来实现。如果神经网络是在比例为$\mathbf{q}$的混合数据集上训练的，则可按照所示步骤推导出式（15），因此我们可以将模型的实际性能与 $C(\mathbf{q})$进行比较。需要注意的是，如果不满足对角化要求，仍可通过计算海森矩阵得到式（14），从而获得 $C(\mathbf{q})$。

IV. 性能评估结果

本节展示所提出的自监督混合深度学习方法的性能评估计算机仿真结果。如第II节所述，我们聚焦于单用户情形，仿真中两个数据集的参数与表II相同。我们假设 WAIR-D n 指的是 WAIR-D 的场景-2 和情形 n，而 DeepMIMO n 指的是 DeepMIMO 的 O1 60 中第 1000 行至第 1300 行之间的 n-th 100θ 个用户。因此，每次训练所用的训练数据集由 WAIR-D 1 和 DeepMIMO 1 混合组成，共计 1000 个用户。测试数据集选自 WAIR-D 2 至 WAIR-D 5 以及 DeepMIMO 2 至 DeepMIMO 5。模型使用 TensorFlow 2.8 在配备 EPYC 75F3 中央处理器和 RTX 3090 图形处理器的 Ubuntu 20.04 系统上训练 10000 个 epoch，学习率初始化为 0.001，批量大小为 256。首先，我们将展示所提方案在两个不同数据集上的鲁棒性，然后评估比例的影响，并与理论分析进行比较。

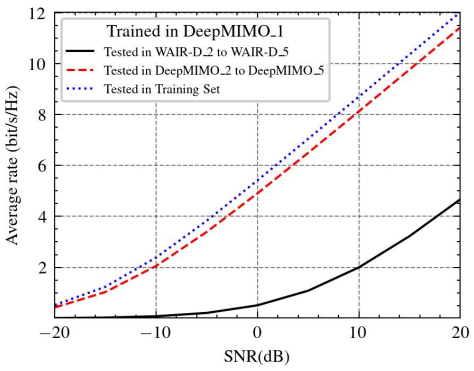


图5. 仅在 DeepMIMO 1 上训练。

表II
系统参数

参数	DeepMIMO	WAIR-D
AP	BS3	BS0
用户	1000	
载波频率	60GHz	
射频链	1	
天线数量	64	
天线间距	$\lambda/2$	
带宽	50 MHz	
路径数量	5	
信号功率	1dbm	
PNR	20dB	

图. 5 和图. 6 表明，波束成形方案仅使用一种数据集进行训练，但分别在两种不同的数据集中进行测试。结果显示，在相同的场景设置下，使用不同训练数据集的两个系统具有相近的系统容量。然而，结果也表明，只有在同一种数据集上训练和测试时，波束成形模型才能取得良好效果，这意味着波束成形模型不具备适应环境变化的鲁棒性。

在图. 7 中，我们评估了所提出的混合训练技术的性能，其中两个数据集按相同比例混合，同时保持用户总数不变。令人惊讶的是，如图. 5 和图. 6 所示，在两个不同数据集上的测试结果并未出现显著的性能损失，即所有测试情形都实现了相近的最大速率。因此，我们得出结论：所提出的自监督混合深度学习方案对不同数据集具有很强的鲁棒性。此外，该方案还可以扩展到更一般的场景，即存在多种数据集时，只需按一定比例从这些数据集中获取训练数据并输入神经网络即可。

在图. 8中，我们评估了比例$\mathbf{q}$ 对所提出的混合训练性能的影响，并将其与理论计算结果 log($C(\mathbf{q})$) 进行了比较。具体而言，当我们将这两个

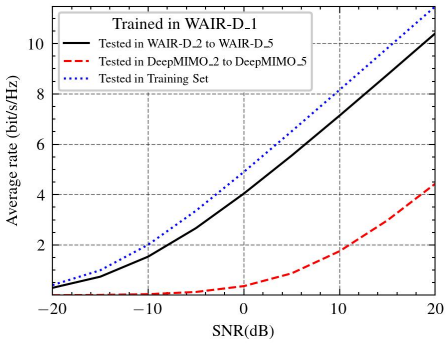


图 6. 仅使用 WAIR-D 1 进行训练。

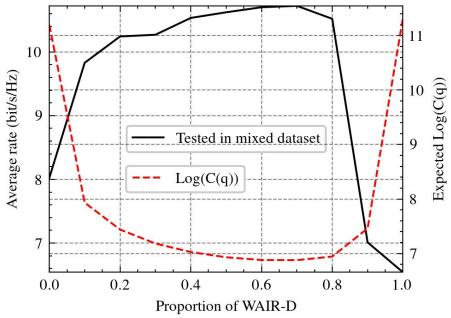


图 8. 使用一半混合数据集和 log(C(q)) 进行评估时，WAIR-D 占比的影响。

数据集的训练样本数量限制为 1000，同时将 WAIR-D 1 的比例从 0 按 0.1 的步长增加到 1 时，我们发现，当 WAIR-D 的比例达到 70% 时，模型输出的平均速率最高。这意味着，如果从 WAIR-D 获取 70% 的数据、从 DeepMIMO 获取 30% 的数据，所提出的方案在两种场景下都能达到最高效率。类似地，我们通过理论函数 (14) 和 (15) 绘制了 log(C(q)) 曲线。可以看出，该曲线呈 U 形，并且在 WAIR-D 的比例为 70% 时达到最小值，这与平均速率性能很好地反向吻合。此外，图. 8 还表明，数据比例过小或过大都会导致性能失衡并产生偏差。这意味着，我们只需最小化 C(q)，即可确定不同训练数据集的最佳比例，从而确保训练好的模型在所有数据集上都具有最佳鲁棒性。

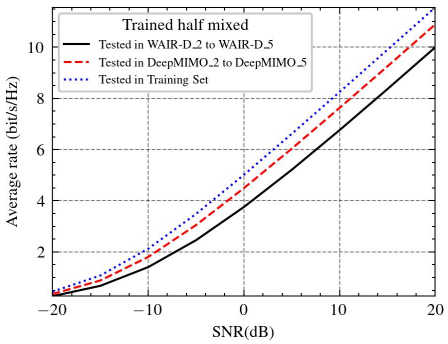


图 7. 使用一半混合数据集进行训练。

V. 结论

本文提出了一种利用自监督混合深度学习的毫米波通信鲁棒波束成形方法。通过大量仿真实验，我们发现，所提出的方案不仅在单个数据集内具有鲁棒性，在不同数据集上也能良好运行。此外，我们还进行了理论分析以解释该方法的有效性，分析结果也与实验结果高度吻合。所提出的自监督波束成形方案还可以扩展至更多类型的数据集，并具有很强的鲁棒性。然而，如何设计最佳网络架构以及各层的具体作用等更有趣的问题，仍在探索之中。

参考文献

[1] E. Bj¨ornson、M. Bengtsson 和 B. Ottersten，“最优多用户发射波束成形：一个难题及其简单的解结构 [讲义],” *IEEE Signal Processing Magazine*，第 31 卷，第 4 期，第 142–148 页，2014年。[2] A. Arora、C. G. Tsinos、B. Shankar Mysore R、S. Chatzinotas 和 B. Ot- tersten，“大规模天线阵列模拟波束成形的主化最小化算法”，载于 *2019 IEEE*全球信号与信息处理会议（*GlobalSIP*），2019年，第 1–5 页。[3] I. Chafaa、R. Negrel、E. V. Bel mega 和 M. Debbah，“利用6 GHz以下频段信道进行毫米波波束转向的自监督深度学习”，*IEEE Transactions on Wireless Communications*，第 21 卷，第 10 期，第 8803–8816 页，2022年。[4] J. Kim、H. Lee、S.-E. Hong 和 S.-H. Park，“通用多输入单输出波束成形的深度学习方法”，*IEEE Wireless Communications Letters*，第 9 卷，第 11 期，第 1894–1898 页，2020年。[5] A. Alkhateeb，“DeepMIMO：用于毫米波和大规模MIMO应用的通用深度学习数据集”，*arXiv* 预印本 *arXiv:1902.06435*，2019年。[6] Y. Huangfu、J. Wang、S. Dai、R. Li、J. Wang、C. Huang 和 Z. Zhang，“WAIR-D：无线人工智能研究数据集”，2022年。 [在线]。链接：https://arxiv.org/abs/2212 02159 。[7] T. Gong、I. Vinieratou、R. Ji、C. Huang、G. C. Alexandropoulos、L. Wei、Z. Zhang、M. Debbah、H. V. Poor 和 C. Yuen，“全息MIMO通信：理论基础、使能技术与未来方向”，2022年。 [在线]。链接：https://arxiv.org/abs/2212 01257 。[8] T. Lin 和 Y. Zhu，“利用深度学习进行大规模天线阵列波束成形设计”，*IEEE Wireless Communications Letters*，第 9 卷，第 1 期，第 103–107 页，2020年。[9] M. Alrabeiah 和 A. Alkhateeb，“利用6 GHz以下频段信道进行毫米波波束与阻塞预测的深度学习”，*IEEE Transactions on Communications*，第 68 卷，第 9 期，第 5504–5518 页，2020年。[10] H. Zhao、R. Mayzus、S. Sun、M. Samimi、J. K. Schulz、Y. Azar、K. Wang、G. N. Wong、F. Gutierrez 和 T. S. Rappaport，“纽约市建筑物内及周边地区28 GHz毫米波蜂窝通信的反射和穿透损耗测量”，载于 *2013 IEEE*国际通信会议（*ICC*），2013年，第 5163–5167 页。[11] X. Li、S. Chen、X. Hu 和 J. Yang，“通过方差偏移理解丢弃法与批归一化之间的不协调”，载于 *2019 IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），2019年，第 2677–2685 页。[12] J. Hestness、S. Narang、N. Ardalani、G. Diamos、H. Jun、H. Kianinejad、M. M. A. Patwary、Y. Yang 和 Y. Zhou，“深度学习的缩放规律在经验上可预测”，2017年。 [在线]。链接：https: // arxiv.org/abs/1712 00409 。[13] J. S. Rosenfeld、A. Rosenfeld、Y. Belinkov 和 N. Shavit，“跨尺度泛化误差的构造性预测”，载于 国际学习表征会议，2020年。 [在线]。链接：https://openreview.net/forum?id= ryenvpEKDr [14] T. Hashimoto，“多个数据源下的模型性能缩放”，载于 第*38* 届国际机器学习会议论文集，Proceedings of Machine Learning Research 系列，M. Meila 和 T. Zhang 编，第 139 卷。PMLR，2021年7月18–24日，第 4107–4116 页。[在线]。链接：https://proceedings.mlr.press/v139/hashimoto21a. html