

工业雾霾环境下多模态语义辅助的毫米波波束预测方法

谢鑫, 杨迪寒, 王一凡, 王恒

(重庆邮电大学工业物联网与网络化控制教育部重点实验室, 重庆 400065)

摘要: 针对现有环境语义辅助的毫米波波束预测方法在工业雾霾环境下的精度下降问题, 提出了一种多模态语义辅助的波束预测方法。首先设计 Lite-DHNet 模块对雾霾图像进行端到端去雾重建, 然后将去雾图像与环境语义中的定位数据进行融合, 最后设计掩码图像特征提取网络和波束索引推理网络, 实现毫米波波束预测。实验结果表明, 所提方法能够以较低的模型开销实现高效的波束预测, Top-3 预测准确率最高可达 99.5%, 有效减轻了雾霾环境对波束预测精度的影响。

关键词: 毫米波通信; 波束预测; 深度学习; 环境语义

中图分类号: TN915.04

文献标志码: A

DOI:10.11959/j.issn.1000-436x.2026047

Multi-modal semantics aided mmWave beam prediction method in industrial haze environment

Xie Xin, Yang Dihan, Wang Yifan, Wang Heng

Key Laboratory of Industrial Internet of Things and Networked Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: To address the issue of decreased accuracy of existing environment semantics aided millimeter wave (mmWave) beam prediction methods in industrial haze environments, a multi-modal semantics aided beam prediction method was proposed. Firstly, a Lite-DHNet module was designed to perform end-to-end dehazing reconstruction on haze images. Then, the dehazed images were fused with localization data within the environment semantics. Finally, a masked image feature extraction network and a beam index inference network were designed to achieve mmWave beam prediction. Experimental results demonstrate that the proposed approach achieves high beam prediction accuracy with low computational overhead, and its Top-3 prediction accuracy reaches up to 99.5%, effectively mitigating the impact of haze environments on beam prediction accuracy.

Keywords: mmWave communication, beam prediction, deep learning, environment semantics

0 引言

随着工业 4.0 时代的到来, 越来越多的工业场景对数据的实时性和稳定性提出了更高要求, 而毫米波通信凭借其高带宽和低时延的特点, 已成为支

持工业无线系统实现超可靠低时延通信的关键技术之一^[1-2]。由于毫米波所具有的高频特性, 毫米波信号传输过程中的路径损耗较高, 因此通信系统通常需要配备大规模天线阵列, 以通过波束成形过程

收稿日期: 2025-11-12; 修回日期: 2026-02-11

通信作者: 王恒, wangheng@cqupt.edu.cn

基金项目: 国家自然科学基金资助项目(No.U23B2003, No.62402081); 重庆市自然科学基金资助项目(No.CSTB2024NSCQ-LZX0127, No.CSTB2025NSCQ-LZX0149); 重庆市教委科学技术研究基金资助项目(No.KJQN202400645)

Foundation Items: The National Natural Science Foundation of China (No.U23B2003, No.62402081), The Natural Science Foundation of Chongqing (No.CSTB2024NSCQ-LZX0127, No.CSTB2025NSCQ-LZX0149), The Science and Technology Research Program of Chongqing Municipal Education Commission (No.KJQN202400645)

产生具有强指向性的毫米波波束,从而保持复杂环境中稳定的视距连接^[3]。此外,为了适应通信目标位置变化引起的信号指向改变,通信系统需要不断扫描指导波束成形过程的大量波束向量^[4]。而这一波束扫描过程会导致通信系统产生大量的波束开销,引起较高的通信时延,从而难以支持如自动驾驶等具有高移动性和要求高实时性的工业无线通信应用^[5]。

传统的波束选择方法致力于根据通信环境的变化设计动态码本、利用信道稀疏性和压缩感知工具进行信道估计,或根据波束特性开发波束跟踪技术,从而降低波束开销^[6-8]。但此类方法所需的参数在实际通信中难以获取或无法完全避免波束扫描过程,因此其能够降低的波束开销较为有限。

为解决上述传统方法所面临的问题,近年来,研究人员提出了一系列通过深度学习进行波束预测的方法,通过在毫米波通信场景中加入额外的用于描述当前环境的传感器,如定位器、摄像头、雷达和激光雷达等,使用提取到的传感信息辅助完成高效的波束预测,此类方法在减少波束开销方面展现出了卓越潜力。具体而言,文献[9]提出了一种位置辅助的毫米波波束预测方法,通过提取安装在移动用户设备(user equipment, UE)端定位器中的定位信息,将其输入一个由3层隐藏层组成的多层感知机(multi-layer perceptron, MLP)网络中,最终输出对应的波束索引。Alrabeiah等^[10]提出了一种视觉辅助的毫米波波束预测方法,直接将安装在基站(base station, BS)端的摄像头拍摄的包含UE的图像输入ResNet中,完成对波束索引的推理和输出。除了视觉信息,Charan等^[11]额外利用定位信息,提出了一种视觉-位置辅助的毫米波波束预测方法,即将图像和定位信息输入一个修改后的ResNet和MLP网络中,对波束索引进行联合预测。文献[12]和文献[13]分别提出了激光雷达和雷达辅助的波束预测方法,通过获取安装在BS端的激光雷达或雷达数据,并将其输入构造的神经网络中,最终完成波束预测过程。此类方法通过对码本中唯一指示波束成形向量的波束索引进行直接预测,避免了波束扫描过程,有效减少了相关波束开销。文献[14-15]针对跨场景波束预测任务提出了迁移学习方案,首先基于已有场景的环境信息完成模型预训练,随后在新场景中仅对预训练模型进行轻量化微

调,即可适配新场景的波束预测需求。该方案有效降低了新场景下的训练计算开销,同时加快了模型的实际部署效率。此外,Sheng等^[16]提出了将波束预测问题构建为时间序列预测任务,通过历史观测信息并利用大型语言模型实现了一种高效且稳健的波束预测方法。

在上述方法中,为了有效处理视觉和雷达等表格化数据,需要构建ResNet等深度卷积神经网络(convolutional neural network, CNN)。然而,此类网络的推理过程通常会消耗大量的系统存储和计算资源,从而导致模型开销显著增加,进一步提高了通信系统的部署成本。此外,在矿厂、加工厂等多粉尘、水汽的工业场景中,易形成雾霾环境,从而导致摄像头的成像效果不佳,进一步降低此类基于视觉辅助的毫米波波束预测性能,影响工业无线通信质量。因此,如何设计一种波束预测方法,在降低波束预测模型开销并提高波束预测性能的同时,降低工业雾霾环境对波束预测的影响,具有重要的研究意义。

本文的主要工作如下。

1) 建立了雾霾环境中UE和BS间的毫米波波束预测模型,以及用于重建清晰图像的端到端图像去雾模型,并定义了基于多模态环境语义的毫米波波束预测任务和端到端图像去雾任务。

2) 为解决雾霾环境下的波束预测问题,提出了工业雾霾环境下多模态语义辅助的轻量级毫米波波束预测方法,通过构建轻量级端到端图像去雾网络(Lite-DHNet)、掩码图像特征提取网络和波束索引推理网络实现高效的波束预测。

3) 仿真结果表明,所提方法能够在工业雾霾环境下达到较高的波束预测精度,同时保持较低的训练开销。此外,在不同对比方法中,所提方法相较于现有方法在模型开销、推理时延以及波束预测精度等方面表现出更优异的性能。

1 系统模型

1.1 毫米波波束预测模型

考虑工业雾霾环境下BS和移动UE间的毫米波无线通信场景,如图1所示。装配有 M 元均匀线性阵列(uniform linear array, ULA)和摄像头的BS为装配有全向天线和定位器的单个移动UE提供毫米波频段下的无线通信服务。在通信过程中,BS采

用循环前缀长度为 D 和子载波数量为 K 的正交频分复用方式不断向移动UE传输数据。

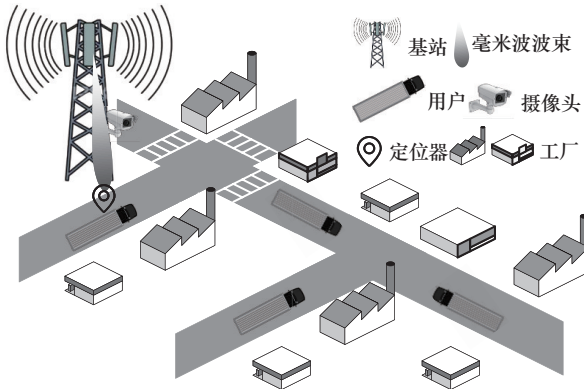


图1 工业毫米波无线通信场景

在 t 时刻,第 k 个子载波上的毫米波传输信道向量 $\mathbf{h}_k[t] \in \mathbb{R}^{M \times 1}$ 可采用几何信道建模,可表示为

$$\mathbf{h}_k[t] = \sum_{d_c=0}^{D_c-1} \sum_{\ell=1}^L \alpha_\ell e^{-j\frac{2\pi k}{K}d_c} p(d_c T_s - \tau_\ell) \mathbf{a}(\theta) \quad (1)$$

其中, L 表示信道路径的数量, α_ℓ 表示第 ℓ 条信道的复增益, τ_ℓ 表示第 ℓ 条信道的时延, θ 表示第 ℓ 条信道路径中到达信号的方位角度, T_s 表示采样的时间间隔(假设最大时延低于 $D_c T_s$), $p(\cdot)$ 表示脉冲整形滤波器, $\mathbf{a}(\theta)$ 表示ULA的阵列导向矢量,表示为

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \left[1, e^{-j\frac{2\pi d_a}{\lambda} \sin(\theta)}, \dots, e^{-j(M-1)\frac{2\pi d_a}{\lambda} \sin(\theta)} \right]^T \quad (2)$$

其中, d_a 表示天线单元间的间距, λ 表示信号波长,为了避免信号混叠,通常取 $d_a = \frac{\lambda}{2}$ 。

同时,考虑该子载波中存在高斯白噪声 $v_k \in \mathcal{CN}(0, \sigma^2)$,并假设BS发射的下行信号为 x ,则 t 时刻UE接收到的第 k 个子载波上的下行信号 $y_k[t]$ 可表示为

$$y_k[t] = \mathbf{h}_k^T[t] \mathbf{f}_i[t] x + v_k[t] \quad (3)$$

其中, $\mathbf{f}_i \in \mathbb{R}^{M \times 1}$ 表示波束成形向量。

信号的可达传输率是衡量无线通信质量的关键指标,在 t 时刻UE基于波束成形向量 $\mathbf{f}_i[t]$ 的信号可达传输率可表示为

$$\text{Rate}_{f_i[t]} = \frac{1}{K} \sum_{k=1}^K \text{lb} \left(1 + \frac{P_k(x)}{\sigma^2} |\mathbf{h}_k^T[t] \mathbf{f}_i[t]|^2 \right) \quad (4)$$

其中, $P_k(x)$ 表示传输信号 x 的功率。

BS需要在包含 Q 个波束成形向量 $\mathbf{f}_i \in \mathbb{R}^{M \times 1}$ 的预定义码本 $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^Q$ 中尽可能地选择能够提供较高信号可达传输率的波束成形向量,用于指示ULA的波束成形过程,实现波束对准。其中,能够最大化 t 时刻信号可达传输率的波束成形向量即最优波束向量,其由码本中向量的索引 $i[t]$ 唯一指示,即最优波束索引 $i^*[t]$,可表示为

$$i^*[t] = \underset{i \in \{1, \dots, Q\}}{\text{argmax}} \text{Rate}_{f_i[t]} \quad (5)$$

为了获得式(5)所示的最优波束索引 $i^*[t]$,现有的基于视觉辅助的波束预测方法通常通过ResNet等CNN实现图像特征的直接提取和波束索引的预测输出,然而此类方法的推理过程需要消耗大量的系统存储和运算资源,产生较大的模型开销和硬件成本,限制了其在真实工业无线通信场景中的应用。

波束预测方法的核心原理是BS根据UE的实时位置,动态选择最优的波束成形向量,因此若能提取通信环境中对于波束预测具有关键作用的位置语义,同时滤除额外的环境因素,实现两者的显式解耦,即可有效简化预测网络,降低模型开销。基于此,本文利用摄像头拍摄的图像信息和定位器获得的位置信息,提取其中的多模态环境语义 $\mathbf{S}[t] = \{\mathbf{m}[t], \mathbf{b}[t], \mathbf{g}[t]\}$,其中 $\mathbf{m}[t]$ 和 $\mathbf{b}[t]$ 分别为当前 t 时刻UE的掩码图像和边界框向量,表示UE的相对空间方位,即相对位置环境语义, $\mathbf{g}[t]$ 为当前 t 时刻UE的位置向量,表示UE的绝对地理坐标,即绝对位置环境语义。相对位置环境语义聚焦UE与BS的局部空间关系,掩码图像通过分割凸显UE在图像中的区域,边界框向量量化UE在BS摄像头中的方位偏移,两者共同表征UE在BS视场中的相对姿态。绝对位置环境语义聚焦UE的全局地理坐标,表征UE在工业场景中的宏观位置,在存在遮挡的情况下辅助修正波束指向偏差。两类环境语义协同构建了局部和全局的多维度空间描述,为波束索引的精准映射提供了完整的空间信息支撑。

基于多模态环境语义的毫米波波束预测任务即是基于多模态环境语义 $\mathbf{S}[t]$,通过映射函数 f_θ 实时估计当前 t 时刻的毫米波波束索引 $\hat{i}[t]$,该过程可表示为

$$f_\theta: \mathbf{S}[t] \rightarrow \hat{i}[t] \quad (6)$$

其中, θ 表示构成映射函数的相关参数。

1.2 端到端图像去雾模型

在工业无线通信场景中，雾霾是影响视觉辅助波束预测的主要因素之一，而大气中粒子的光散射作用是使观测图像产生雾气的主要原因。大气散射模型（atmospheric scattering model, ASM）是一种用于描述大气中光传播及其与尘埃、水汽等颗粒物相互散射作用的物理模型，能够较好地模拟实际的雾霾效果，广泛应用于基于深度学习的图像去雾领域。具体而言，设大气光常量值为 A ， $I \in \mathbb{R}^{C \times H \times W}$ 表示摄像头拍摄到的雾霾图像， $J \in \mathbb{R}^{C \times H \times W}$ 表示清晰图像， z 表示图像中的某一个像素位置， C 、 H 和 W 分别表示图像的通道数、高度和宽度，则图像去雾领域的ASM可以表示为

$$I(z) = J(z)t(z) + A(1 - t(z)) \quad (7)$$

其中， $t(z) = e^{-\beta d_i(z)}$ 表示透射率图，决定了光在穿过大气后最终到达摄像头传感器的比例， β 表示大气散射系数， d_i 表示摄像头的景深距离。

根据式(7)可知，清晰图像可通过获得参数 $t(z)$ 和 A 的值，并通过式(8)进行重建。

$$J(z) = X(z)I(z) - X(z) + b_{ias} \quad (8)$$

其中，参数 $t(z)$ 和 A 封装在一个新的如式(9)所示的参数 X 中。

$$X(z) = \frac{I(z) - A}{t(z)} + (A - b_{ias}) \quad (9)$$

其中， b_{ias} 是值为1的默认偏差项。

端到端图像去雾的任务是估计参数 $X \in \mathbb{R}^{C \times H \times W}$ ，并基于映射函数 f_X ，去除雾霾图像 I 中的雾霾，重

建清晰图像 \hat{J} ，该过程可表示为

$$f_X: I \rightarrow \hat{J} \quad (10)$$

2 基于多模态语义辅助的雾霾环境毫米波波束预测

为了完成毫米波波束预测任务和端到端图像去雾任务，本文提出了如图2所示的工业雾霾环境下多模态语义辅助的毫米波波束预测架构。

在多模态语义辅助的毫米波波束预测方法应用于实际雾霾场景的过程中，首先需要获取BS摄像头拍摄的包含用户的雾霾图像，并通过 Lite-DHNet 完成清晰图像端到端重建，而后基于去雾图像进行用户目标检测，提取用户的边界框向量，进而生成其掩码图像，作为相对位置环境语义。同时通过定位器获取用户的位置向量，作为绝对位置环境语义。在毫米波波束预测过程中，通过掩码图像特征提取网络处理掩码图像，并将其输出与对应的边界框向量和位置向量合并，输入波束索引推理网络，输出最优波束索引，完成毫米波波束预测。值得注意的是，多模态数据的正确对应是实现高精度波束预测至关重要。使用毫米波扫描的时间作为锚点，并使用协调世界时间戳动态选择时间上最接近毫米波扫描时间的GPS和相机数据样本，可以有效地降低不同模态传感器之间异步的不利影响。

2.1 基于图像去雾的多模态环境语义提取

在多模态环境语义提取过程中，为了避免雾霾环境对相机成像效果和后续毫米波波束预测的影响，本文采用深度学习的方式，通过构建如图2所

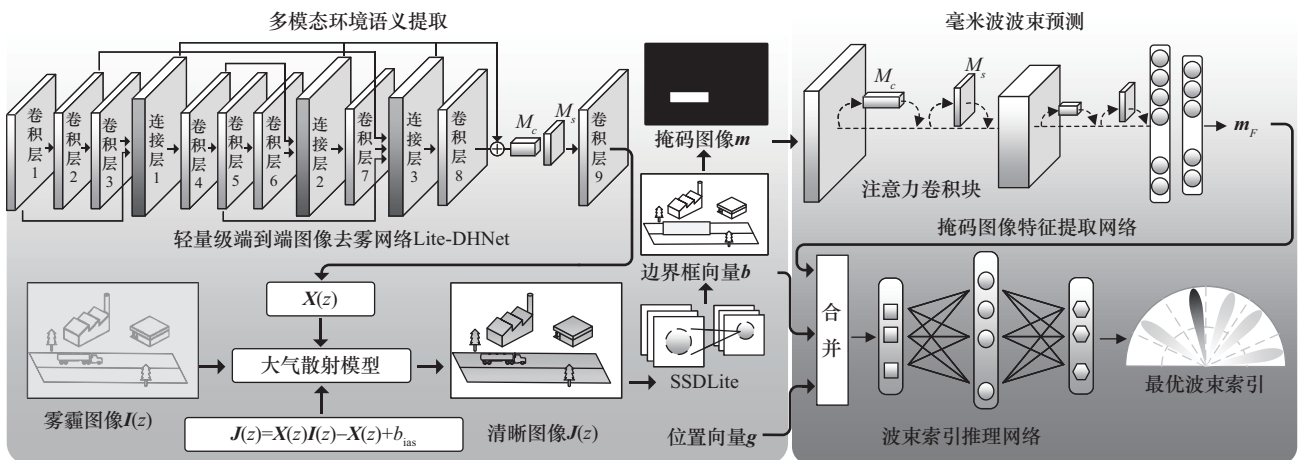


图2 工业雾霾环境下多模态语义辅助的毫米波波束预测架构

示的 Lite-DHNet, 并使用雾霾图像和清晰图像共存的数据集 $Z = \{(\mathbf{I}_p, \mathbf{J}_p)\}_{p=1}^P$ 进行训练, 以拟合映射函数 f_X , 其中 P 表示该数据集中可用的样本总数。具体而言, 为了满足模型训练时对足量雾霾样本和清晰样本对的需求, 雾霾图像 \mathbf{I}_p 通过从 $[0.6, 1.0]$ 和 $\{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$ 中随机选择参数 A 和 β 的值, 进一步基于式(7)生成雾霾图像, 所使用的数据集由 27 256 对清晰图像和雾霾图像样本组成。深度学习的优化目标是 minimized 给定样本的清晰图像和对应去雾图像间的均方误差损失, 该过程可表示为

$$f_X^* = \arg \min_{f_X} \frac{1}{P} \sum_{p=1}^P \left\| \mathbf{J}_p - f_X(\mathbf{I}_p) \right\|_2^2 \quad (11)$$

Lite-DHNet 以深度可分离卷积为基础, 加入跳层拼合和残差连接操作, 同时引入注意力机制, 从而在较低模型开销下实现对参数 \mathbf{X} 的准确估计, 进一步重建清晰图像 \mathbf{J} 。

具体而言, Lite-DHNet 利用深度可分离卷积替代传统卷积作为网络的基础架构, 采用深度可分离卷积层对输入的雾霾图像 $\mathbf{I} \in \mathbb{R}^{C \times H \times W}$ 进行特征提取, 输出通道数为 C_{out} 。首先, 通过 C 个大小为 $1 \times d_h \times d_w$ 的深度可分离卷积核为雾霾图像 $\mathbf{I} \in \mathbb{R}^{C \times H \times W}$ 的每个通道单独进行卷积操作, 完成空间滤波, 并输出特征 $\mathbf{I}_{\text{dw}} \in \mathbb{R}^{C \times H' \times W'}$ 。而后, 通过 C_{out} 个大小为 $C \times 1 \times 1$ 的逐点卷积核为特征 \mathbf{I}_{dw} 的每个通道进行卷积操作, 完成通道变换, 最终输出特征 $\mathbf{I}_{\text{pw}} \in \mathbb{R}^{C_{\text{out}} \times H'' \times W''}$ 。相比于普通卷积, 深度可分离卷积通过解耦对输入特征的空间滤波和通道变换操作, 有效降低了网络参数数量和计算量, 同时保持了较好的特征表达能力。

Lite-DHNet 在各深度可分离卷积层间加入了拼接层和残差连接, 即拼接层 1 沿通道方向对深度可分离卷积层 1 和 3 进行连接, 得到输出 $\mathbf{O}_{\text{link1}}$; 拼接层 2 沿通道方向对深度可分离卷积层 4 和 6 进行连接, 得到输出 $\mathbf{O}_{\text{link2}}$; 拼接层 3 沿通道方向对深度可分离卷积层 2、5 和 7 进行连接, 得到输出 $\mathbf{O}_{\text{link3}}$; 深度可分离卷积层 8 的输出 \mathbf{O}_{con8} 和 $\mathbf{O}_{\text{link1}}$ 、 $\mathbf{O}_{\text{link2}}$ 及 $\mathbf{O}_{\text{link3}}$ 通过残差连接, 得到输出 \mathbf{O}_{Res} , 该过程可表示为

$$\mathbf{O}_{\text{Res}} = \mathbf{O}_{\text{con8}} + \text{Resize}_{16 \rightarrow 56}(\mathbf{O}_{\text{link1}}) + \text{Resize}_{32 \rightarrow 56}(\mathbf{O}_{\text{link2}}) + \mathbf{O}_{\text{link3}} \quad (12)$$

其中, $\text{Resize}_{16 \rightarrow 56}(\cdot)$ 和 $\text{Resize}_{32 \rightarrow 56}(\cdot)$ 表示将特征维度分别从 16 维和 32 维映射至 56 维的操作。通过拼合操作, 可在避免网络出现低层次信息丢失

的同时, 有效保留网络各层次之间的差异化特征, 通过残差连接, 可增强网络各层之间的特征传递和融合, 提高网络的训练效率。

Lite-DHNet 加入了注意力层, 实现了通道注意力和空间注意力机制^[17]。注意力层包含通道注意力模块 M_c 和空间注意力模块 M_s , 将残差连接的输出 $\mathbf{O}_{\text{Res}} \in \mathbb{R}^{C'' \times H'' \times W''}$ 输入通道注意力模块 M_c 后, 首先分别通过窗口大小为 $H'' \times W''$ 的最大池化层和平均池化层输出 $\mathbf{F}_{\text{max}} \in \mathbb{R}^{C'' \times 1 \times 1}$ 和 $\mathbf{F}_{\text{avg}} \in \mathbb{R}^{C'' \times 1 \times 1}$, 而后将 \mathbf{F}_{max} 和 \mathbf{F}_{avg} 输入一个 MLP 网络中, 该网络的输入层和输出层均包含 C'' 个神经元, 隐藏层则包含 $\frac{C''}{r}$ 个神经元, 其中 r 表示降低比率, 最后将该网络的两个输出进行加和, 得到 \mathbf{O}_{Res} 的通道注意力 $M_c(\mathbf{O}_{\text{Res}})$, \mathbf{O}_{Res} 逐元素乘 $M_c(\mathbf{O}_{\text{Res}})$ 进一步得到应用通道注意力后的输出 \mathbf{O}'_{Res} , 该过程可表示为

$$\mathbf{O}'_{\text{Res}} = M_c(\mathbf{O}_{\text{Res}}) \otimes \mathbf{O}_{\text{Res}} \quad (13)$$

将 \mathbf{O}'_{Res} 输入空间注意力模块 M_s 后, 首先沿通道方向分别计算 \mathbf{O}'_{Res} 的最大值和平均值, 得到 $\mathbf{F}'_{\text{max}} \in \mathbb{R}^{1 \times H'' \times W''}$ 和 $\mathbf{F}'_{\text{avg}} \in \mathbb{R}^{1 \times H'' \times W''}$, 而后将 \mathbf{F}'_{max} 和 \mathbf{F}'_{avg} 沿通道方向拼接, 得到 $[\mathbf{F}'_{\text{avg}}; \mathbf{F}'_{\text{max}}] \in \mathbb{R}^{2 \times H'' \times W''}$, 并对其进行值为 $\frac{d-1}{2}$ 的填充操作, 最后通过一个大小为 $2 \times d \times d$ 的卷积核对其进行特征提取, 得到 \mathbf{O}'_{Res} 的空间注意力 $M_s(\mathbf{O}'_{\text{Res}})$, \mathbf{O}'_{Res} 逐元素乘 $M_s(\mathbf{O}'_{\text{Res}})$ 进一步得到应用空间注意力后的输出 \mathbf{O}_{ResA} , 该过程可表示为

$$\mathbf{O}_{\text{ResA}} = M_s(\mathbf{O}'_{\text{Res}}) \otimes \mathbf{O}'_{\text{Res}} \quad (14)$$

通过实现通道注意力和空间注意力机制, 可以帮助网络学习如何关注特定的对于参数 \mathbf{X} 的估计具有重要作用的特征, 分别确定“关注什么”和“关注哪里”。

将雾霾图像 \mathbf{I} 输入 Lite-DHNet, 重建清晰图像 \mathbf{J} , 完成端到端图像去雾任务, 而后通过轻量级目标识别网络 SSDLite^[18] 提取去雾图像中的用户边界框坐标 $\mathbf{b}' = [x_r, y_r, x_l, y_l]$, 其中 x_r 、 y_r 、 x_l 和 y_l 分别表示边界框的左上点横纵坐标和右下点横纵坐标。边界框坐标 \mathbf{b}' 经过进一步计算后得到用户边界框向量 $\mathbf{b} = [c_x, c_y, w, h]$, 其中 c_x 和 c_y 分别表示边界框中心点的横纵坐标, w 和 h 分别表示边界框的宽度和高度。基于边界框坐标 \mathbf{b}' 生成用户掩码图像 $\mathbf{m}' \in \mathbb{R}^{1 \times H \times W}$ 。因为波束指向仅取决于 UE 的位置,

图像中的冗余信息易使模型学习伪特征,降低预测精度,所以将用户边界框以内的区域变为白色,其余部分则为黑色,以此生成二值化的掩码图像。此外,由于UE相对位置语义侧重区域方位,将掩码图像的尺寸缩小后能够在不显著降低波束预测精度的同时减少计算量、加速训练,故将其尺寸进行缩小,得到 $\mathbf{m} \in \mathbb{R}^{1 \times 32 \times 32}$,最终完成表示相对位置环境语义的掩码图像 \mathbf{m} 和边界框向量 \mathbf{b} 的提取。

同时,通过部署在移动用户上的定位器获取其定位坐标,表示为位置向量 $\mathbf{g} = [l_a, l_o]$,其中 l_a 和 l_o 分别表示移动用户当前所处位置的纬度和经度坐标,完成表示绝对位置环境语义的位置向量 \mathbf{g} 的提取。

2.2 基于特征提取和索引推理的毫米波波束预测

在毫米波波束预测过程中,为了实现基于多模态环境语义 $\mathbf{S} = \{\mathbf{m}, \mathbf{b}, \mathbf{g}\}$ 的毫米波波束预测,本文采用深度学习的方式,通过构建如图2所示的掩码图像特征提取网络和波束索引推理网络,使用多模态传感数据和通信数据共存的 DeepSense 6G^[19] 数据集 $\mathcal{D} = \{(\mathbf{S}_u, i_u^*)\}_{u=1}^U$ 进行训练,以拟合映射函数 f_ϕ ,其中 U 表示该数据集中可用的样本总数。深度学习的优化目标是最大化给定样本准确预测的可能性,该过程可表示为

$$f_\phi^* = \operatorname{argmax}_{f_\phi} \prod_{u=1}^U \mathbb{P}(i_u^* = \hat{i}_u | \mathbf{S}_u) \quad (15)$$

其中, $\mathbb{P}(\cdot)$ 表示映射函数 f_ϕ 准确预测的可能性。

掩码图像特征提取网络旨在从图像数据中提取有关波束预测的关键特征,因此其以小尺寸卷积层和简单全连接层为基础,同时引入注意力机制,从而在轻量级模型开销下实现对掩码图像特征 \mathbf{m}_F 的准确提取。

具体而言,掩码图像特征提取网络由两个基于注意力的卷积块 (attention-based convolutional block, AbCB) 和全连接网络组成,其中 AbCB 由一层卷积层、一层池化层,以及包含通道注意力模块 M_c 和空间注意力模块 M_s 的卷积块注意力模块 (convolutional block attention module, CBAM) 构成,负责提取表格类型的掩码图像中的关键特征。考虑第一个 AbCB 对掩码图像 \mathbf{m} 的特征提取过程,首先通过卷积层对 \mathbf{m} 进行卷积操作,得到特征 \mathbf{F}'_m ,而后通过池化层对特征 \mathbf{F}'_m 进行最大池化操作,得到特征 \mathbf{F}''_m ,进一步通过通道注意力模块 M_c 得到 \mathbf{F}'''_m 的通道注意力 $M_c(\mathbf{F}'''_m)$,将其应用于 \mathbf{F}'''_m 后得到 \mathbf{F}'_A ,

最后通过空间注意力模块 M_s 得到 \mathbf{F}'_A 的空间注意力 $M_s(\mathbf{F}'_A)$,将其应用于 \mathbf{F}'_A 后得到 \mathbf{F}_A 。全连接网络负责对 AbCB 输出的数据进行特征的进一步提取,最终输出数值型的掩码图像特征 \mathbf{m}_F 。掩码图像特征提取网络相关参数如表 1 所示。

表 1 掩码图像特征提取网络相关参数

网络架构	卷积核/池化窗口大小	卷积核/神经元个数	步长数
卷积层 1+ReLU 函数	5 × 5	6	1
最大池化层 1	2 × 2	—	2
CBAM 层 1	—	—	—
卷积层 2+ReLU 函数	5 × 5	16	1
最大池化层 2	2 × 2	—	2
CBAM 层 2	—	—	—
全连接层 1+ReLU 函数	—	120	—
全连接层 2+ReLU 函数	—	84	—
全连接层 3+ReLU 函数	—	14	—

将用户掩码图像 \mathbf{m} 输入掩码图像特征提取网络,输出特征 $\mathbf{m}_F \in \mathbb{R}^{14 \times 1}$ 后,将其与对应的边界框向量 $\mathbf{b} \in \mathbb{R}^{4 \times 1}$ 和位置向量 $\mathbf{g} \in \mathbb{R}^{2 \times 1}$ 在空间维度的第一个维度上通过拼接操作进行融合,得到向量 $\mathbf{S}' \in \mathbb{R}^{20 \times 1}$ 。

不同于掩码图像特征提取网络中处理图像数据的需求,波束索引推理网络旨在从数值型数据 \mathbf{S}' 中推理最终的毫米波波束索引 \hat{i} ,因此由全连接网络组成,其相关参数如表 2 所示。

表 2 波束索引推理网络相关参数

网络架构	神经元个数
输入层	20
隐藏层 1+ReLU 函数	128
隐藏层 2+ReLU 函数	256
隐藏层 3+ReLU 函数	512
隐藏层 4+ReLU 函数	256
隐藏层 5+ReLU 函数	128
输出层	32

具体来说,该全连接网络包含 1 层输入层、5 层隐藏层和 1 层输出层,各层的每一个神经元均与其前一层的所有神经元相连,其中输入层负责将输入数据传递至隐藏层,而隐藏层负责通过引入 ReLU 激活函数对数据进行非线性特征变换和提取,最后输出层负责将特征信息转换为最终结果。

将向量 $S' \in \mathbb{R}^{20 \times 1}$ 输入波束索引推理网络后, 输出最优波束索引 \hat{i} , 完成基于多模态环境语义的毫米波波束预测任务, 进而确定与该索引 \hat{i} 唯一的码本中指示毫米波波束方向的波束成形向量 f_i , 从而实现毫米波信号的定向发射或接收。

3 实验分析

所提轻量级端到端图像去雾网络 Lite-DHNet 由 9 层深度可分离卷积层、1 层注意力层和 3 层全连接层组成, 采用文献[20]中的雾霾和清晰图像共存的数据集进行训练, 训练过程基于均方误差损失, 并采用 Adam 优化器, 其中学习率为 0.000 1, 批大小为 8, 训练轮次为 100。所提掩码图像特征提取网络使用文献[19]在真实世界中采集的 DeepSense 6G 数据集, 基于交叉熵损失, 采用 Adam 优化器, 通过 ReduceLROnPlateau 学习率调度器初始学习率为 0.01、最小学习率为 0.001、批大小为 32 的参数进行轮次为 100 的训练。

本文选择 DeepSense 6G 数据集中的场景 1 和场景 5 对所提方法进行评估和分析。数据集中采集数据的实验平台由两个硬件单元组成, 第一单元的功能是作为接收器, 并配备一台 RGB 摄像头、一个 16 单元 60 GHz 的毫米波均匀线性阵列。第二单元作为发射器, 配备了 GPS 传感器和 60 GHz 的全向天线。其中, 场景 1 为白天环境下的双向车道场景, 场景 5 为另一个地理位置的夜间双向车道场景。两个场景分别覆盖了不同光照条件和不同地理位置的典型差异, 能够充分体现本文方法在差异化条件下的预测性能, 从而实现对模型泛化能力的全面评估。场景 1 中的总样本数量为 1 160, 场景 5 中的总样本数量为 2 083, 实验中数据集按照 70%、20% 和 10% 的比例切分为训练集、验证集和测试集。同时, 为了验证所提方法的性能, 本文对 4 种已有的毫米波波束预测基线方法在本文所提的背景中进行复现: 文献[9]提出的位置辅助的毫米波波束预测方法, 其仅使用位置数据作为 MLP 网络的输入进行波束预测; 文献[10]和文献[11]提出的视觉和视觉-位置辅助的毫米波波束预测方法, 其联合图像或位置数据, 将其输入经过修改的 ResNet 中进行波束预测。此外, 还设置了整体架构与本文方法相同, 但目标识别网络由 SSDLite 替换为采用 YOLOv7^[21]的毫米波波束预测方法。

图 3 展示了不同波束预测方法的模型大小以及在场景 1 和场景 5 中对应的波束预测准确率。其中, 波束预测准确率采用 Top- n 表示, 其代表真实最优波束索引被包含在模型输出的前 n 个候选波束索引集合中的概率。首先, 使用 YOLOv7 作为目标检测网络的波束预测方法拥有最大的模型大小, 超过 140 MB, 同时也取得了较高的 Top-1 毫米波波束预测准确率, 在场景 1 中为 79.5%, 在场景 5 中为 81.3%。而后, 视觉-位置和视觉辅助的波束预测方法的模型大小次之, 分别约为 89.93 MB 和 42.7 MB, 同时视觉辅助的波束预测方法取得了相对较好的 Top-1 毫米波波束预测准确率, 在场景 1 中为 75.2%, 在场景 5 中为 79.9%。值得注意的是, 相比于使用 YOLOv7 作为目标检测网络的波束预测方法, 本文方法的模型大小约为 14.93 MB, 其中语义提取和波束推理部分的参数量所占内存分别约为 13.64 MB 和 1.28 MB。本文方法在取得相似 Top-1 毫米波波束预测准确率的同时, 降低了约 89.39% 的模型大小。相比于视觉辅助的波束预测方法, 本文方法仅以约 34.96% 的模型大小, 在场景 1 和场景 5 中分别提升了 4.3 和 1.4 个百分点的 Top-1 毫米波波束预测准确率。最后, 位置辅助的波束预测方法拥有最小的模型大小和最低的 Top-1 毫米波波束预测准确率。

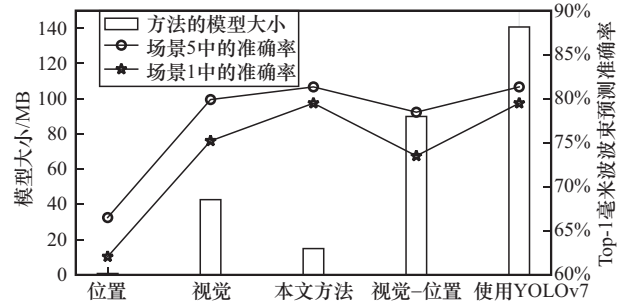


图3 不同波束预测方法的模型大小和在不同场景中的 Top-1 毫米波波束预测准确率

为了评估模型的实时性能, 本文在搭载 Intel Core i7-12700H 处理器 (4.70 GHz CPU, 32 GB RAM) 的硬件平台上, 基于深度学习框架 PyTorch 2.0, 对不同模型在波束预测阶段的推理时间进行了对比。图 4 展示了不同波束预测方法处理不同数量的样本时在对数尺度上的推理时间, 本文方法进行单次波束推理的时间不超过 0.01 s, 而相比于视觉或视觉-位置辅助的波束预测方法, 本文方法将推理时间缩短了至少一个数量级。

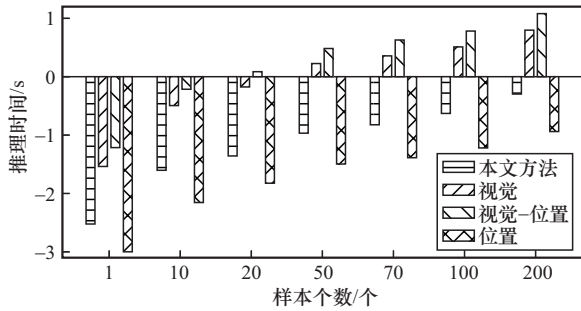


图4 不同波束预测方法处理不同数量的样本时在对数尺度上的推理时间

以上实验结果表明，相比于通过 ResNet 等深度卷积神经网络直接处理原始图像，从而完成波束预测的视觉或视觉-位置辅助方法，本文方法通过将语义提取和波束推理过程解耦，显著降低了波束预测过程中的模型开销和推理时长，同时有效提高了波束预测的性能，体现了该方法在工业无线通信场景中的应用潜力。

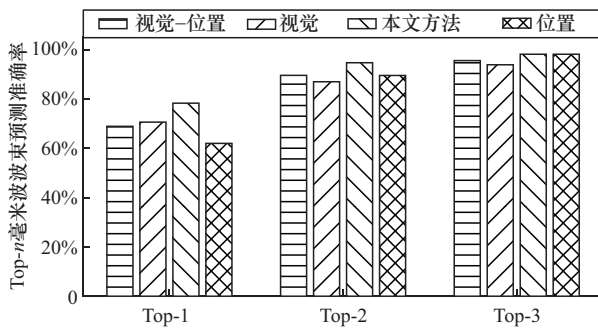
图5展示了不同波束预测方法在雾霾场景1和雾霾场景5中的Top-n毫米波波束预测准确率，场景中的雾霾密度由式(7)随机生成。其中，所提方法在雾霾场景1和雾霾场景5中的Top-3波束预测准确率分别为98.3%和99.5%。相比于视觉或视觉-位置辅助的波束预测方法在雾霾场景1和雾霾场景5中的Top-1、Top-2和Top-3波束预测表现，本文方法均具备更优的预测性能。相较于前者，本文方法在雾霾场景1和雾霾场景5中分别实现了约7.7和1.4个百分点的Top-1波束预测准确率提升；相较于后者，本文方法在雾霾场景1和雾霾场景5中分别实现了9.4和2.9个百分点的Top-1波束预测准确率提升。这是由于在雾霾条件下，随着图像质量的降低，视觉辅助的波束预测方法的性能会降低。需要注意的是，尽管位置辅助的波束预测

方法相比本文方法在Top-1波束预测中的性能表现欠佳，但其仍然获得了较高的Top-3波束预测准确率，在雾霾场景1中为98.3%，在雾霾场景5中为91.4%。

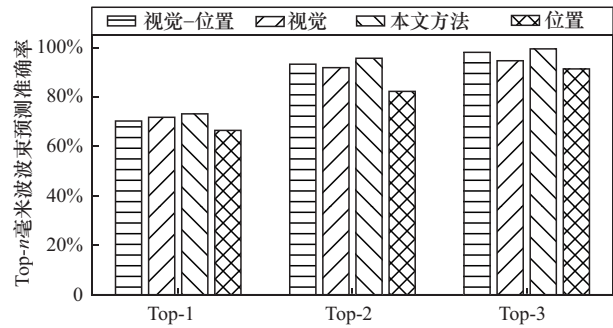
相较于视觉或视觉-位置辅助的波束预测方法，本文方法的波束预测性能提升表明，首先使用 Lite-DHNet 进行图像去雾，而后将语义提取和波束预测过程解耦，在大幅降低模型开销的同时，实现了雾霾环境下波束预测精度的有效提升，即首先提取用于描述移动用户和BS间相对方位的相对位置环境语义（用户掩码图像和用户边界框坐标），以及用于描述移动用户在全球的绝对方位的绝对位置环境语义（用户位置向量），完成语义提取过程，而后将多模态环境语义通过构造的神经网络输出波束索引，完成波束预测过程。

为了进一步验证雾霾环境下所提轻量级端到端图像去雾网络 Lite-DHNet 对毫米波波束预测的影响，本文采用文献[22]和文献[20]提出的图像去雾网络 DehazeNet 和 AOD-Net 进行清晰图像重建，形成多模态语义辅助的毫米波波束预测基线方法。

表3展现了不同去雾方法在式(7)所生成的雾霾图像的峰值信噪比（peak signal-to-noise ratio, PSNR）^[23]和结构相似性（structural similarity index, SSIM）^[24]的对比。从表3中可以看出，相较于原始雾霾图像，本文提出的 Lite-DHNet 在雾霾场景1和雾霾场景5中，PSNR 分别提升了12.02 dB和13.41 dB，SSIM 则分别提升了0.083 7和0.102 7。该方法在两类雾霾场景下均实现了更高的PSNR与SSIM值，这表明其复原图像的质量更接近清晰参考图像，有效提升了图像的视觉与像素级质量，而高质量的去雾图像能够为后续的目标检测和波束预测下游任务提供可靠的数据支撑。



(a) 雾霾场景1中的预测准确率



(b) 雾霾场景5中的预测准确率

图5 不同波束预测方法在雾霾场景1和雾霾场景5中的Top-n毫米波波束预测准确率

表3 不同去雾方法的PSNR和SSIM对比

方法	雾霾场景1		雾霾场景5	
	PSNR/dB	SSIM	PSNR/dB	SSIM
雾霾图像	14.05	0.843 6	11.63	0.640 9
DehazeNet	25.79	0.926 9	24.71	0.740 1
AOD-Net	24.75	0.925 4	23.80	0.730 5
Lite-DHNet	26.07	0.927 3	25.04	0.743 6

图6展示了工业雾霾环境下使用不同去雾网络进行清晰图像重建后波束预测方法的Top-1波束预测准确率。其中, 直接对雾霾图像进行相对位置环境语义的提取, 进而进行波束预测的方法, 在雾霾场景1和雾霾场景5中仅取得了57.8%和67.3%的Top-1波束预测准确率。而使用Lite-DHNet完成图像去雾后, 本文方法在雾霾场景1和雾霾场景5中取得了20.6和5.9个百分点的Top-1波束预测准确率提升, 分别达到78.4%和73.2%。同时, 相比于使用AOD-Net、DehazeNet以及不引入CBAM的Lite-DHNet(Lite-DHNet w/o CBAM)进行图像去雾后在雾霾场景1和雾霾场景5中的Top-1波束预测准确率, 使用Lite-DHNet作为图像去雾网络的波束预测方法均取得了不同程度的性能提升。此外, 根据式(7), 可以通过调整参数 β 的取值来改变生成雾霾的密度。本文选取0.4、1.0和1.6这3个不同的 β 值, 分别对应轻度、中度和重度雾霾情况。基于此, 图7进一步展示了在雾霾场景1(日间)与雾霾场景5(夜间)中, 本文方法在不同雾霾程度下的Top-1波束预测准确率。由于夜间环境能见度较低, 雾霾场景5的整体准确率低于日间的雾霾场景1。实验结果表明, 在轻度和中度雾霾情况下, 本文方法的Top-1波束预测准确率仅出现小幅下降。即便在重度雾霾情况下, 尽管准确率下降幅度较中度雾霾有所增加, 但仍能保持在较高水平, 这说明了本文方法在不同雾霾程度下均具有良好的鲁棒性。

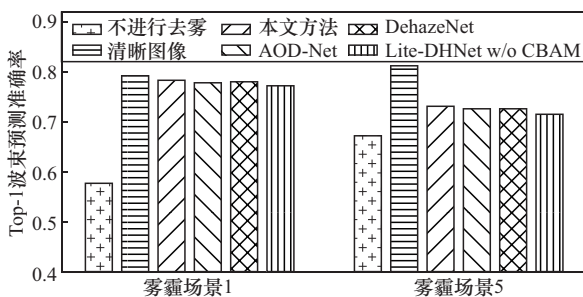


图6 工业雾霾环境下使用不同去雾网络进行清晰图像重建后波束预测方法的Top-1波束预测准确率

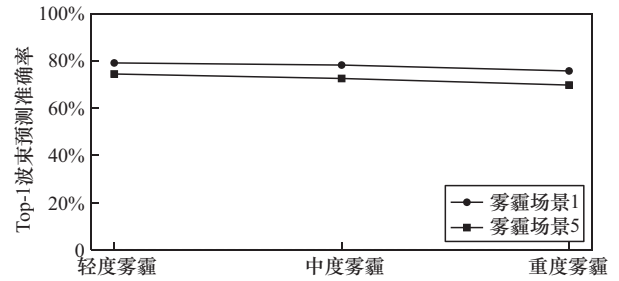


图7 不同雾霾程度下本文方法的Top-1波束预测准确率

以上实验结果表明, 所提轻量级端到端图像去雾网络Lite-DHNet在清晰图像重建过程中, 以及在工业雾霾环境下多模态语义辅助的毫米波波束预测过程中的有效性, 即对于BS端摄像头拍摄的雾霾图像, 首先通过Lite-DHNet进行清晰图像的端到端重建, 而后基于去雾图像进行相对位置环境语义的提取, 进一步完成波束预测的这一流程可以有效降低工业雾霾环境对波束预测性能的影响, 满足工业复杂环境下的毫米波通信需求。

4 结束语

为了在工业雾霾环境下以较小的模型开销实现对毫米波波束的有效预测, 本文考虑雾霾环境对波束预测性能的影响, 针对BS和移动UE间的毫米波通信过程, 提出了工业雾霾环境下多模态语义辅助的毫米波波束预测方法。首先通过构造的Lite-DHNet对BS端摄像头拍摄的雾霾图像进行清晰图像的端到端重建, 而后从去雾图像数据和用户定位器获得的位置数据中提取相对和绝对位置环境语义, 最后通过图像特征提取网络和波束索引推理网络进一步完成波束预测。实验结果表明, 本文方法以较低的模型开销实现了较高的预测精度, 同时有效减少了工业雾霾环境对波束预测的影响, 能够有效满足实际工业复杂环境中的毫米波通信需求。未来研究将探索多用户场景下波束预测与资源调度的协同优化机制, 通过构建多目标与用户资源的动态关联调度方案, 在保障多用户波束预测实时性的同时, 进一步提升资源利用效率。

参考文献:

[1] Xue Q, Ji C W, Ma S D, et al. A survey of beam management for mmWave and THz communications towards 6G[J]. IEEE Communications Surveys & Tutorials, 2024, 26(3): 1520-1559.
 [2] Khan B S, Jangsher S, Ahmed A, et al. URLLC and eMBB in 5G industrial IoT: a survey[J]. IEEE Open Journal of the Communications Soci-

- ety, 2022, 3: 1134-1163.
- [3] 张泽鹏, 李翠然, 吴昊, 等. 基于列车先验信息的 RIS-MIMO 系统鲁棒波束成形算法[J]. 通信学报, 2024, 45(5): 29-43.
Zhang Z P, Li C R, Wu H, et al. Robust beamforming algorithm for RIS-MIMO system based on train prior information[J]. Journal on Communications, 2024, 45(5): 29-43.
- [4] Xiao Z Q, Chen S Q, Zeng Y. Simultaneous multi-beam sweeping for mmWave massive MIMO integrated sensing and communication[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 8141-8152.
- [5] Kim J, Choi Y J, Noh G, et al. On the feasibility of remote driving applications over mmWave 5G vehicular communications: implementation and demonstration[J]. IEEE Transactions on Vehicular Technology, 2023, 72(2): 2009-2023.
- [6] Zhang Y, Alrabeiah M, Alkhateeb A. Reinforcement learning of beam codebooks in millimeter wave and terahertz MIMO systems[J]. IEEE Transactions on Communications, 2022, 70(2): 904-919.
- [7] Song J, Choi J, Love D J. Common codebook millimeter wave beam design: designing beams for both sounding and communication with uniform planar arrays[J]. IEEE Transactions on Communications, 2017, 65(4): 1859-1872.
- [8] 李双志, 杨睿棋, 郭新, 等. 低导频开销 RIS 辅助毫米波 MIMO 系统参数化信道估计方案[J]. 通信学报, 2024, 45(9): 82-91.
Li S Z, Yang R Q, Guo X, et al. Low pilot overhead parametric channel estimation scheme for RIS-assisted mmWave MIMO systems[J]. Journal on Communications, 2024, 45(9): 82-91.
- [9] Morais J, Bchboodi A, Pezeshki H, et al. Position-aided beam prediction in the real world: how useful GPS locations actually are?[C]//Proceedings of the ICC 2023-IEEE International Conference on Communications. Piscataway: IEEE Press, 2023: 1824-1829.
- [10] Alrabeiah M, Hredzak A, Alkhateeb A. Millimeter wave base stations with cameras: vision-aided beam and blockage prediction[C]//Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). Piscataway: IEEE Press, 2020: 1-5.
- [11] Charan G, Osman T, Hredzak A, et al. Vision-position multi-modal beam prediction using real millimeter wave datasets[C]//Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2022: 2727-2731.
- [12] Jiang S F, Charan G, Alkhateeb A. LiDAR aided future beam prediction in real-world millimeter wave V2I communications[J]. IEEE Wireless Communications Letters, 2023, 12(2): 212-216.
- [13] Demirhan U, Alkhateeb A. Radar aided 6G beam prediction: deep learning algorithms and real-world demonstration[C]//Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2022: 2655-2660.
- [14] Tosi E, Hu P W, Ichkov A, et al. Cross-environment transfer learning for location-aided beam prediction in 5G and beyond millimeter-wave networks[C]//Proceedings of the ICC 2025-IEEE International Conference on Communications. Piscataway: IEEE Press, 2025: 3563-3568.
- [15] Feng Y, Zhao C B, Gao F F, et al. Environment sensing-aided beam prediction with transfer learning for smart factory[J]. IEEE Transactions on Wireless Communications, 2025, 24(1): 676-690.
- [16] Sheng Y C, Huang K, Liang L, et al. Beam prediction based on large language models[J]. IEEE Wireless Communications Letters, 2025, 14(5): 1406-1410.
- [17] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Berlin: Springer, 2018: 3-19.
- [18] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4510-4520.
- [19] Alkhateeb A, Charan G, Osman T, et al. DeepSense 6G: a large-scale real-world multi-modal sensing and communication dataset[J]. IEEE Communications Magazine, 2023, 61(9): 122-128.
- [20] Li B Y, Peng X L, Wang Z Y, et al. AOD-net: all-in-one dehazing network[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 4780-4788.
- [21] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 7464-7475.
- [22] Cai B L, Xu X M, Jia K, et al. DehazeNet: an end-to-end system for single image haze removal[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198.
- [23] Li B Y, Ren W Q, Fu D P, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2019, 28(1): 492-505.
- [24] Liu W, Zhou F, Lu T, et al. Image defogging quality assessment: real-world database and method[J]. IEEE Transactions on Image Processing, 2021, 30: 176-190.

[作者简介]



谢鑫 (1990-), 男, 河南南阳人, 博士, 重庆邮电大学讲师, 主要研究方向为环境辅助的毫米波通信、工业网络实时调度等。



杨迪寒 (2000-), 男, 重庆人, 重庆邮电大学硕士生, 主要研究方向为基于深度学习的毫米波波束预测。



王一凡 (2000-), 男, 陕西汉中, 重庆邮电大学硕士生, 主要研究方向为基于深度学习的毫米波波束预测。



王恒 (1981-), 男, 河北新乐人, 博士, 重庆邮电大学教授, 主要研究方向为工业物联网、无线通信等。