第六代移动通信通感一体化中用于波束预测的多模态与带内毫米波感知系统级比较

Abidemi Orimogunje*†、Hyunwoo Park†、Igbafe Orikumhi†、Sunwoo Kim†、Dejan Vukobratovic‡
* 非洲物联网卓越中心，卢旺达大学，卢旺达
†电子工程系，汉阳大学，韩国
‡技术科学学院，诺维萨德大学，塞尔维亚。

arXiv:2601.01033v1 [eess.SP] 3 Jan 2026

摘要—**集成感知与通信（通感一体化）可通过实现基于带内感知的波束预测，降低毫米波车联网到基础设施（V2I）链路中的波束训练开销；外感知传感器则可进一步提升预测准确率。本研究构建了一个系统级框架，利用 DeepSense-6G Scenario-33 数据集评估摄像头、激光雷达、雷达、GPS 和带内毫米波功率的性能，包括各传感器单独使用及多模态融合的情况。该框架采用由轻量级卷积神经网络（CNN）和多层感知机（MLP）编码器组成的时延感知神经网络，预测 64 个波束中的波束索引。我们使用 Top-k 准确率以及频谱效率（SE）差距、信噪比（SNR）差距、速率损失和端到端延迟评估性能。结果表明，毫米波功率向量是性能强劲的独立预测器，与外感知传感器融合后仍能保持优异性能：仅使用毫米波以及毫米波+激光雷达/GPS/雷达均达到 98% 的 Top-5准确率，而毫米波+摄像头达到 94% 的 Top-5准确率。该框架为第六代移动通信通感一体化辅助的车联网系统波束预测建立了经过校准的基准。**

索引词—**波束管理、波束预测、通感一体化、多模态感知、第六代移动通信。**

一、引言

未来的第六代移动通信（6G）系统正逐步融合为集成感知与通信（通感一体化）系统，其中，基础设施处的共址传感器持续感知环境，同时维持高速率链路 [1]。在毫米波/太赫兹频段，窄模拟波束对于克服严重的路径损耗并维持较高信噪比（SNR）至关重要 [2]–[4]。然而，对此类波束进行穷举式波束扫描并频繁重新训练会引入时延和控制开销，尤其是在车辆到基础设施（V2I）场景中，系统动态变化迅速，且视距（LoS）/非视距（非LoS）状态频繁切换 [5], [6]。利用外部感知信息（来自摄像头、雷达等）预测波束，可以减少带内训练并加快链路恢复；这一方向已在真实世界的 DeepSense-6G 数据集和新兴的通感一体化原型系统上得到验证 [7]–[9]。

近期研究表明：(i) 使用真实毫米波数据的视觉-位置融合可实现较高的 Top-3 准确率，并显著降低波束训练开销 [10], (ii) 基于位置的模型仅使用商用全球定位系统 (GPS) 时，Top-1 准确率虽一般，但相较于最优波束仍能实现较低的功率损失 (速率损失)，这表明准确率本身并不能真实反映系统性能 [11], (iii)改进模态编码器和池化器式 Transformer 可提升准确率并降低复杂度 [12], 以及 (iv)采用激光雷达和摄像头进行三维定位的几何感知流程可直接引导波束并减少搜索开销 [13],而数字孪生射线追踪的出现，则使波束突发能够实时感知障碍物 [14]。

基于这些发现，我们提出了一个完整、经过验证且可复现的通感一体化辅助多模态波束预测框架，并使用 DeepSense-6G 场景-33 数据集。我们的贡献总结如下：

1）我们比较了单一感知模态（摄像头、激光雷达、雷达、GPS 和带内毫米波功率）、两两融合及多模态融合，不仅考察 Top-k准确率，还评估频谱效率（SE）、信噪比、速率损失和端到端延迟，从而使模型性能评估与实际通信行为相契合。2）我们将64维毫米波接收功率向量视为一种重要的通感一体化感知模态，与摄像头、激光雷达、雷达和 GPS 并列。每种传感器数据均由紧凑型卷积神经网络或多层感知机骨干网络编码，再由基于Transformer的融合模块汇总各嵌入表示，以预测接收波束索引。3）在场景-33上，我们证明毫米波功率向量本身就是一种强有力的基础模态，而融合 GPS 或激光雷达可带来稳定增益，摄像头和雷达则能在特定情况下提升性能。我们量化了 Top-1准确率的微小变化如何转化为速率损失和信噪比差距的显著变化。

总而言之，本研究推动了通感一体化辅助波束预测的发展，不再局限于仅比较准确率，而是在真实的 60 GHz车联网测量数据上，将波束选择质量与信噪比/频谱效率性能以及感知到波束成形的时延结合起来。该框架为未来第六代移动通信通感一体化的部署提供了实用指导和可复现的基线。

本文其余部分的组织结构如下。第二节详细介绍系统模型。第三节描述所提出的多模态架构、融合策略和训练流程。第四节给出实验结果，第五节总结全文。

通讯作者：remero@hanyang.ac.kr。共同通讯作者：oigbafe2@hanyang.ac.kr

II. 系统模型

我们考虑一个包含两个单元的车联网一体化感知与通信（V2I ISAC）系统 [15]。**单元1**（基础设施）是一个工作在 60 GHz 的固定接收机，配备包含 N_r=16 个单元的相控阵以及大小为 B=64 的接收码本。共址传感器包括 RGB 摄像头、三维激光雷达、调频连续波雷达（FMCW）和 GPS-RTK接收机，用于观测场景。**单元2**（移动端）是一个配备单个有源天线单元的发射机。所有传感器均与时间戳 t处的毫米波扫描在时间上保持同步。

A. 通信模型

在时刻 t，发射机发射一个复符号 $x_t \in \mathbb{C}$。单元1处的阵列观测为

$$\mathbf{y}_t \;=\; \mathbf{h}_t\, x_t \;+\; \mathbf{n}_t, \tag{1}$$

其中 $\mathbf{h}_t \in \mathbb{C}^{N_r}$ 是接收端观测到的下行信道，$\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ 是热噪声。应用第 i 个接收波束后，得到标量合并后输出

$$z_{t,i} \;=\; \mathbf{w}_i^H \mathbf{y}_t, \tag{2}$$

相应的接收功率测量为

$$p_{t,i} \;=\; |z_{t,i}|^2, \qquad i = 1, \ldots, 64. \tag{3}$$

汇总所有波束的测量后，得到功率向量$\mathbf{p}_t=[p_{t,1}, \ldots, p_{t,64}]^\top \in \mathbb{R}^{64}_{\geq 0}$，我们将其同时视为通信测量和感知特征。时刻 t 的最优波束索引为

$$b_t^\star \;=\; \arg\max_{i \in \{1,\ldots,64\}} p_{t,i}. \tag{4}$$

在噪声方差为 σ^2 时，候选波束 i 对应的瞬时波束成形后信噪比和速率为

$$\mathrm{SNR}_{t,i} \;=\; \frac{p_{t,i}}{\sigma^2}, \tag{5}$$

$$R_{t,i} \;=\; \log_2\!\big(1 + \mathrm{SNR}_{t,i}\big), \tag{6}$$

预言机频谱效率为 $R_{t,\ b_t^\star}$。在全文中，我们采用多模态波束预测研究中使用的标准基于码本的波束选择抽象。对于预测的索引 $\hat{}\ b_t$，可达速率为

$$R_{t,\hat{b}_t} = \log_2\!\left(1 + \frac{p_{t,\hat{b}_t}}{\sigma^2}\right). \tag{7}$$

我们定义了能够将模型质量与系统影响相对应的性能差距：

$$\mathrm{SNR_gap}_t\ [\mathrm{dB}] = 10 \log_{10}\!\left(\frac{\mathrm{SNR}_{t,b_t^\star}}{\mathrm{SNR}_{t,\hat{b}_t}}\right) \tag{8}$$

$$\mathrm{Rate_loss}_t = R_{t,b_t^\star} - R_{t,\hat{b}_t}, \tag{9}$$

$$\mathrm{SE_opt} = \mathbb{E}_t[R_{t,b_t^\star}], \tag{10}$$

$$\mathrm{SE_pred} = \mathbb{E}_t[R_{t,\hat{b}_t}] \tag{11}$$

此外，还报告建议用于实际评估的平均功率损失指标。预言机频谱效率始终是上界，即最优频谱效率 $\geq$预测频谱效率。

B. 感知模型

在每个同步时间戳处，单元1提供融合感知集

$$\mathcal{S}_t \;=\; \{\, \mathbf{I}_t,\ L_t,\ \mathcal{R}_t,\ \mathbf{g}_t \},\tag{12}$$

其中$\mathbf{I}_t$，L_t，$\mathcal{R}_t$，$\mathbf{g}_t$，分别表示摄像头图像、激光雷达点云、雷达特征和基于 GPS 的特征。各个元素概述如下。

1) RGB 相机

RGB 相机与毫米波阵列共址安装，我们使用以 30 fps 采样的 RGB 流 $\mathbf{I}_t \in \mathbb{R}^{540\times960\times3}$。它提供基站视场内的语义和几何线索。

2）激光雷达

旋转式三维激光雷达（探测距离为 100 m，最高 20 Hz）返回一个点云 $L_t=\{\mathbf{q}_\ell\}_{\ell=1}^{N_L} \subset \mathbb{R}^3$（$N_L$ 表示激光雷达扫描中的点数），以近似均匀的方位角-仰角间隔对环境进行采样。激光雷达提供结构和移动障碍物具有度量精度的距离信息，通过揭示可能的遮挡物和影响多径传播的反射表面，为毫米波功率向量提供补充。

3）调频连续波雷达

每个雷达帧由复数样本组成，其维度为 (A×S×C)，分别对应接收天线数量 A、每个啁啾的样本数 S 以及每帧的啁啾数C。经过标准预处理后，我们通过二维快速傅里叶变换构建幅度距离-速度（$\mathbf{H}_{\mathrm{RV}}$）和距离-角度（$\mathbf{H}_{\mathrm{RA}}$）表示，并将 $\mathcal{R}_t$={$\mathbf{H}_{\mathrm{RA}}$,$\mathbf{H}_{\mathrm{RV}}$} 作为雷达特征。雷达提供对光照具有鲁棒性的多普勒和角度信息，能够适应各种天气条件下的工作，并为动态目标提供摄像头-激光雷达信息的补充。

4) GPS-RTK

GPS 接收机提供一个大地测量特征向量 $\mathbf{g}_t$ 。在位置辅助波束管理中， [11],由 GPS 得出的相对几何关系与视觉和雷达线索融合后，可减少波束训练开销；不过，原始 GPS 可能存在非高斯误差，从而影响直接波束分类。

5) 毫米波接收功率向量

64 维毫米波接收功率向量（$\mathbf{p}_t$）作为显式感知信道纳入其中。它捕获了 64 波束码本上的瞬时角度接收功率分布，因此对于学习型通感一体化系统中的波束对准具有丰富的信息。将 $\mathbf{p}_t$ 用作输入特征，符合近期多模态建模的做法，即融合外感知与带内无线电特征 [16]。

C. 问题表述

令 $\mathcal{B} = \{1, \ldots, 64\}$ 表示接收波束码本，$\mathcal{S}_t$表示时刻 t的同步通感一体化观测。理想波束索引根据无线电测量定义为 (4)

目标是训练一个预测器 $\mathcal{M}_\theta$，将多模态观测映射为波束后验 $\boldsymbol{\pi}_t \;=\; \mathcal{M}_\theta(\mathcal{S}_t;\mathbf{p}_t) \;\in\; \mathbb{R}^{|\mathcal{B}|}$，并输出

$$\hat{b}_t \;=\; \arg\max_{i \in \mathcal{B}} \;\pi_{t,i}. \tag{13}$$

除 top-k 分类外，该问题还以系统为中心：我们寻求能够 (i) 最大化频谱效率的预测，

$$\mathrm{SE}_{\mathrm{pred}} \;=\; \mathbb{E}\Big[\log_2\big(1+\tfrac{p_{t,\hat{b}_t}}{\sigma^2}\big)\Big], \tag{14}$$

并且 (ii) 将与预言机的信噪比差距降至最低，

$$\mathrm{SNR_gap} \;=\; \mathbb{E}\Big[10\log_{10}\Big(\tfrac{p_{t,b_t^\star}}{p_{t,\hat{b}_t}}\Big)\Big], \tag{15}$$

同时保持较高的 Top-K 准确率和较低的感知与推理延迟。形式化地，我们学习 θ ，以优化多目标准则

$$\min_{\theta}\ \mathbb{E}\Big[\ \underbrace{\mathrm{CE}(b_t^\star,\boldsymbol{\pi}_t)}_{\text{beam classification}}\ +\lambda_{\mathrm{gap}}\cdot\mathrm{SNR_gap}\ +\ \lambda_\tau\cdot\tau_t\Big], \tag{16}$$

其中 CE 表示交叉熵， τ_t 汇总传感器和网络推理延迟，而 (λ_{gap}， λ_τ) 用于权衡准确率、速率/信噪比最优性和延迟。我们评估 $\mathcal{M}_\theta$ 在单一模态及其融合情况下的表现，以量化现实硬件损伤下的准确率、频谱效率损失、信噪比差距和端到端延迟。

III. 提出的方法

我们设计了一种多模态波束预测网络，用于在 Scenario-33 V2I 场景中，根据同步通感一体化观测 $\mathcal{S}_t$ 估计接收波束索引 b_t 。该设计遵循三项原则：(i) 将带内通信信号视为额外的感知流，通过嵌入 64维毫米波功率向量来实现；(ii) 使用轻量级 CNN/MLP编码器提取各模态特征，并将所有传感器映射到共享嵌入维度；(iii) 使用基于 Transformer 的融合模块融合这些嵌入，该模块可灵活处理任意可用模态子集。该网络以波束分类器的形式进行训练，并通过准确率和通信中心指标进行评估。

A. 数据准备与同步

场景-33为每个索引提供一个同步样本，其中包括 RGB相机帧、 64×1 毫米波接收功率向量、激光雷达点云、雷达帧和GPS测量值。我们首先将原始日志转换为紧凑的索引数据集。对于每个样本索引，预处理流程：

- 将毫米波测量值转换为64维接收功率特征向量；
- 将激光雷达测量值提取为XYZ点云特征，并以紧凑表示形式存储；
- 将雷达测量值解析为二维雷达强度图，以便后续处理；
- 将每个样本与其对应的相机观测相关联；并且
- 将 GPS 测量值转换为逐样本特征向量，其中包含纬度、经度、速度和质量指标。

所有模态均按索引对齐，并以毫米波序列为参考；对齐后的样本按连续的 70/15/15 比例划分为训练集/验证集/测试集。

B. 模态专用编码器

每个传感器流都由专用神经编码器进行嵌入处理，输出一个 d维特征向量（在我们的实现中，嵌入维度统一为 d=256 ）。这有助于保持各模态之间的对称性，并简化融合。

1) 毫米波功率向量

64维毫米波接收功率向量由一个两层多层感知机进行编码。具体而言，编码器先通过第一个线性层，将 $\mathbb{R}^{64}$ 映射到 $\mathbb{R}^{2d}$ ，再经过 ReLU 激活函数；随后通过第二个线性层，将 $\mathbb{R}^{2d}$ 映射到 $\mathbb{R}^{d}$ ，并再次经过 ReLU 激活函数。得到的嵌入表示 $\mathbf{f}^{(\mathbf{p}_t)}{}_t \in \mathbb{R}^d$捕捉了接收码本上的角度功率分布，并作为通信感知特征。

2) 摄像头

对于 RGB相机流，我们使用卷积骨干网络将图像张量 $\mathbf{I}_t \in \mathbb{R}^{3\times H\times W}$ 映射为一个d维嵌入表示。我们使用一个在 ImageNet 上预训练的 ResNet-18 模型，并将其分类头替换为投影至 $\mathbb{R}^d$的线性投影。此外，还使用了一个带 ReLU激活函数的轻量级卷积神经网络。在这两种情况下，摄像头编码器都会输出 $\mathbf{f}^{(\mathbf{I}_t)}{}_t \in \mathbb{R}^d$，以概括基站周围的语义与几何上下文。

3) 激光雷达

首先，将激光雷达点云转换为预定义视场内固定（ H_{bev}， W_{bev}）网格上的鸟瞰视角（BEV）占用图像。随后，将生成的单通道鸟瞰视角地图输入编码器，该编码器与卷积神经网络骨干的结构相同，但输入通道数为 1。此编码器依次应用带 ReLU 激活函数和池化的堆叠卷积、自适应平均池化以及线性投影，得到 $\mathbf{f}^{(L_t)}{}_t \in \mathbb{R}^d$。这种表示以紧凑的栅格形式编码道路布局、静态结构和潜在的遮挡几何结构。

4) 雷达

雷达帧以实值二维图的形式加载（经过幅值处理、降维、归一化和尺寸调整后）。随后将这些图输入卷积神经网络编码器，该编码器由三个带 ReLU 激活函数的卷积层和中间池化层组成。之后进行自适应平均池化，并投影至 $\mathbb{R}^d$。雷达编码器生成 $\mathbf{f}^{(\mathcal{R}_t)}{}_t \in \mathbb{R}^d$，以捕捉距离/角度结构和运动线索，补充摄像头和激光雷达的信息。

5) GPS-RTK

这些特征表示为一个定长向量（例如纬度和经度），并经过归一化。该向量输入到一个多层感知机编码器，其结构与毫米波功率编码器相同，即由两层 ReLU 多层感知机构成，中间层宽度为 $2d$，输出维度为 d。GPS分支生成 $\mathbf{f}^{(\mathbf{g}_t)}{}_t \in \mathbb{R}^d$,，提供粗略的几何信息（相对位置和运动信息），有助于消除波束歧义，尤其是在其他模态受损时。

C. 基于 *Transformer* 的多模态融合

令 $\mathcal{M}_t$ 表示时间 t可用的模态集合，并令 $\mathcal{F}_t = \{\mathbf{f}_t^{(m)} : m \in \mathcal{M}_t\}$ 表示其嵌入表示。我们使用基于令牌的

Transformer 编码器对这些模态嵌入进行融合。首先，模态投影器将嵌入集合映射为令牌序列。随后，在序列前添加一个可学习令牌，并加入一组固定的、经过学习的位置嵌入，以编码令牌顺序。由此得到一个张量，其中的令牌对应摄像头、激光雷达、雷达、GPS 和/或毫米波嵌入，具体取决于当前存在的模态。接着，我们在该令牌序列上应用一个多层 Transformer 编码器，并使用多头自注意力。由于令牌列表仅由批次中存在的模态构成，因此同一架构能够无缝处理任意传感器子集（例如，仅毫米波、摄像头+GPS，或完整融合）。

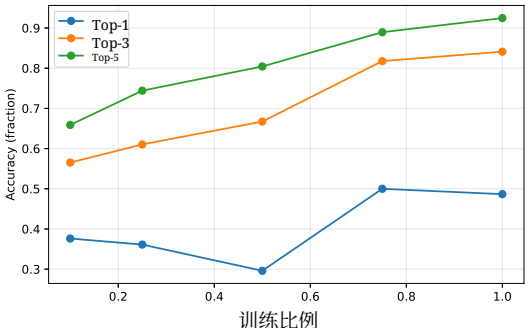


图 1：主要模态及其融合的学习曲线（Top-1/3/5 准确率与数据集的关系）。

D. 延迟感知评估

尽管延迟并未明确纳入训练损失，但实现会在评估期间同时跟踪传感器延迟和推理延迟。对于每个批次，我们测量模型的前向传播时间，并在所有样本上进行汇总，以获得平均推理延迟。随后，将其与用户指定的每种模态传感器延迟相结合，从而估算给定传感器组合的端到端延迟。除 Top-k准确率外，我们还计算预测波束和预言机波束下的频谱效率、以 dB 为单位的信噪比差距，以及预测波束功率与最优波束功率之间的增益比。该评估协议将多模态感知质量与所考虑的车联网一体化感知与通信场景中的通信性能及时延约束直接关联起来。

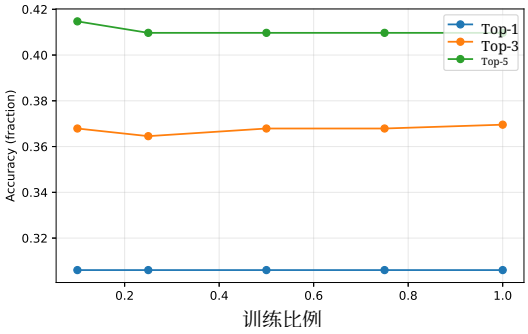


图 2：激光雷达模态的学习曲线。

IV. 结果与讨论

A. 评估指标

我们在 DeepSense-第六代移动通信场景-33 [15], 上使用接收波束码本 $\mathcal{B}$，对每种传感器组合分别进行训练和评估。除 Top-k 准确率外，我们还报告：(i) 与预言机波束之间的平均信噪比间隔（以 dB 为单位的信噪比差距）；(ii) 预测波束与预言机波束的频谱效率（$\mathrm{SE}_{\mathrm{pred}}$ 和 $\mathrm{SE}_{\mathrm{opt}}$，单位为比特/秒/赫兹）；(iii) 平均速率损失（$\mathbb{E}[\Delta R]$）；以及 (iv) 端到端延迟。因此，我们针对每种传感器组合分别呈现所有评估指标，并对不同组合进行比较。

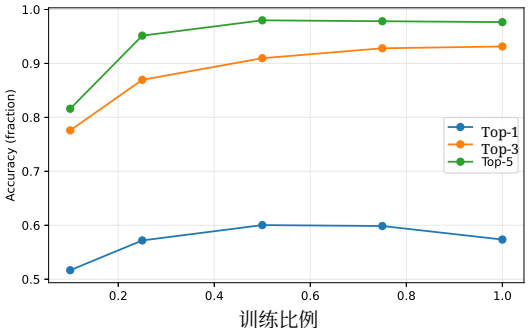


图 3：毫米波功率模态的学习曲线。

B. 总体准确率与学习行为

图 1 展示了主要感知模态融合的训练动态。图 2–6 中各感知模态的曲线凸显了互补行为：随着模型逐渐掌握基站-用户设备几何关系，GPS 的表现持续提升；激光雷达受益于鸟瞰图预处理，早期快速取得增益；摄像头和雷达的进展则更为渐进，但在与几何信息和毫米波融合后，能够带来额外提升。

图 7 中的交叉组合比较证实了
这表明毫米波功率是表现强劲的单一模态，Top-3/Top-5准确率约为 93%/98%。如表 I 所示，添加 GPS、雷达或激光雷达后，Top-1/3/5准确率始终保持稳定，且复杂度适中。

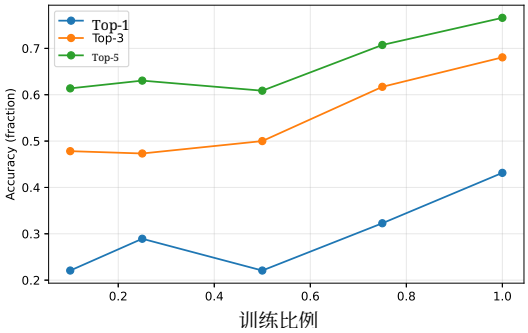


图 4：摄像头模态的学习曲线。

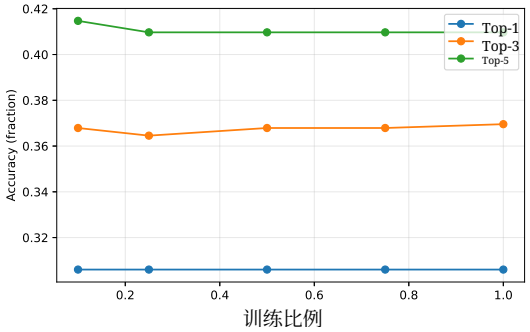


图 5：GPS模态的学习曲线。

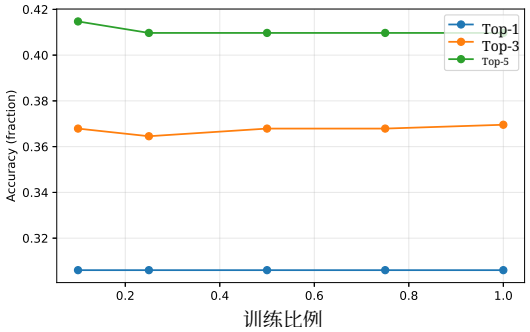


图 6：雷达模态的学习曲线。

C. 频谱效率与信噪比差距

准确率差异并不总是线性地转化为通信质量。图 8 对比了 $\mathrm{SE}^{\mathrm{pred}}$ 与预言机频谱效率$^{\mathrm{opt}}$。仅毫米波基线在场景-33上已实现较小的频谱效率差距；融合 GPS（稳健的方位/探测距离线索）和激光雷达（遮挡/反射体几何结构）可进一步缩小这一差距，使$\mathbb{E}[\Delta R]$ 相对 Top-1准确率的变化呈现出不成比例的下降。图 9 报告了第 II 节所定义的平均信噪比间隔。仅毫米波配置的平均间隔小于 1 dB，而毫米波+GPS/激光雷达配置则进一步降低了该间隔。摄像头和雷达在低能见度或高动态片段中发挥作用；与毫米波/GPS/激光雷达组合融合时，观察到的净收益最大。

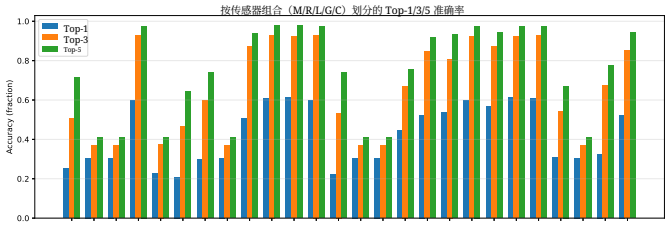


图 7：按传感器组合划分的 Top-$1/3/5$ 准确率。摄像头、GPS、激光雷达、毫米波和雷达感知模态分别以 C、G、L、M 和 R 表示。

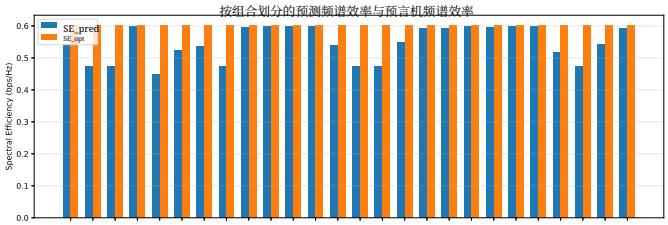


图 8：各组合的 $\mathrm{SE}^{\mathrm{pred}}$ 与 $\mathrm{SE}^{\mathrm{opt}}$ 对比。几何辅助融合（毫米波+GPS/激光雷达）使 SE 差距的缩小幅度显著超过 Top-1 的变化。

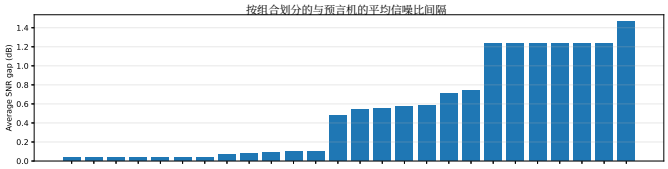


图 9：相对于预言机波束的平均信噪比差距（单位：dB）。在场景-33中，仅使用毫米波时的差距小于 1 dB；GPS/激光雷达可进一步缩小差距；摄像头/雷达则带来情境性增益。

表 I：感知模态（单独使用及与毫米波融合）在场景-33上的性能。

感知模态	**Top-1**	**Top-3**	**Top-5**	$\mathbb{E}[\Delta R]$	**GR**
单一模态					
毫米波	0.60	0.93	0.98	0.0044	0.99
GPS	0.31	0.37	0.41	0.1306	0.77
雷达	0.23	0.38	0.41	0.1530	0.72
激光雷达	0.31	0.37	0.41	0.1306	0.77
摄像头	0.25	0.51	0.72	0.0612	0.89
与毫米波融合					
毫米波+GPS	0.61	0.93	0.98	0.0041	0.99
毫米波+雷达	0.60	0.93	0.98	0.0043	0.99
毫米波+激光雷达	0.61	0.92	0.98	0.0042	0.99
毫米波+摄像头	0.51	0.87	0.94	0.0091	0.98
M+G+L+C+R	0.52	0.85	0.94	0.0102	0.98

缩写： M = 毫米波，G = GPS，L = 激光雷达，C = 摄像头，R = 雷达；$\mathbb{E}[\Delta R]$ = 预期速率损失；GR = 增益比。

D. 延迟-精度权衡

图10将端到端延迟分解为感知和推理两部分。毫米波+GPS配置是延迟效率最高的工作点，只增加极少延迟便能实现接近基线的频谱效率$^{\mathrm{pred}}$。激光雷达和雷达会增加感知时间，但在移动性较强和间歇性遮挡场景下可缩小信噪比差距。摄像头分支以适度的开销提供语义上下文。我们的多模态融合模块（第III节）支持使用部分模态集合，因此必要时可关闭成本较高的分支（如摄像头），同时仍通过毫米波+GPS保留频谱效率$^{\mathrm{pred}}$。

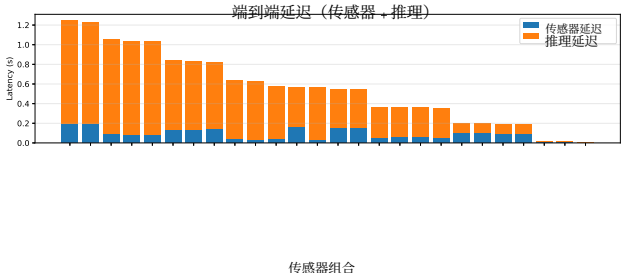


图 10：针对不同传感器配置，将端到端延迟分解为感知和推理组件。

E. 数据效率与消融实验

学习曲线切片（图2–图6）表明，该方法具有很强的数据效率。毫米波和GPS模型迅速趋于饱和，而激光雷达、摄像头和雷达模型随着训练数据增加仍持续改进，并在中等数据量阶段开始趋于平稳。在消融研究（表I）中，我们进一步发现，将纯粹面向分类的目标函数替换为第II节提出的以系统为中心的损失函数后，该损失函数将学习信号与速率和信噪比明确耦合起来，从而持续降低速率损失和信噪比差距，同时保持Top-k 准确率。这一结果使预测器更好地契合了前文定义的系统级目标。

V. 结论

我们针对 DeepSense-6G 场景 33 上的集成感知与通信波束预测进行了以系统为中心的对比分析。我们的延迟感知融合模型将毫米波功率向量作为一种感知特征，与摄像头、激光雷达、雷达和 GPS 特征一同使用，并采用速率感知损失函数进行训练。结果表明，毫米波提供了强有力的锚定信号；GPS 和激光雷达带来稳定的性能提升；摄像头和雷达则能在特定情境下改善性能。值得注意的是，即使 Top-1准确率只有小幅变化，也可能显著缩小频谱效率和信噪比差距。在实际应用中，我们建议默认采用毫米波+GPS，并根据硬件约束选择性地启用其他传感器。总体而言，该框架为可复现的基线方案，兼顾实际限制，并支持跨场景、时序和近场扩展。

致谢

本研究由非洲物联网卓越中心（ACEIoT）卢旺达大学、区域奖学金与创新基金（RSIF）以及韩国国家研究基金会共同资助，资助编号为 RS- 2024-00409492。

参考文献

[1] D. Lee、Y. Lee、C. Song、J. Oh、N. T. Vi 和 S. Cho，“集成感知与通信综述：RIS、无人机与多址接入的集成系统”，发表于*2025*信息与通信人工智能国际会议（*ICAIIC*），2025年，第0028–0031页。[2] J. Bang、H. Chung、J. Hong、H. Seo、J. Choi 和 S. Kim，“毫米波通信：硬件与波束管理算法的最新进展和挑战”，*IEEE*通信杂志，第59卷，第8期，第86–92页，2021年。

[3] L. Li、W. Chen、Z. Chen、T. Hu、W. Mei 和 B. Ning，“利用通感一体化辅助波束管理提升太赫兹通信覆盖范围”，*IEEE*无线通信，第31卷，第1期，第34–40页，2024年。[4] H. Ahn、I. Orikumhi、J. Kang、H. Park、H. Jwa、J. Na 和 S. Kim，“基于机器学习的视觉辅助波束选择：面向毫米波多用户MISO系统”，*IEEE*无线通信快报，第11卷，第6期，第1263–1267页，2022年。[5] W. Chen、L. Li、Z. Chen、Y. Liu、B. Ning 和 T. Q. S. Quek，“通感一体化赋能的太赫兹网络波束对准：方案设计与覆盖分析”，*IEEE*车辆通信汇刊，第73卷，第12期，第19 019–19 033页，2024年。[6] H. Park、J. Kang、S. Lee、J. W. Choi 和 S. Kim，“基于深度Q网络的移动毫米波通信波束跟踪”，*IEEE*无线通信汇刊，第22卷，第2期，第961–971页，2023年。[7] K. Tan 和 C. Zhu，“智能车联网多模态感知：面向部署的最新进展综述”，*IEEE*物联网期刊，第12卷，第22期，第46 294–46 315页，2025年。[8] C. Zheng、J. He、C. G. Kang、G. Cai、Z. Yu 和 M. Debbah，“M2BeamLLM：大语言模型赋能的多模态感知毫米波波束预测”，2025年。 [在线]。网址：https://arxiv.org/abs/2506 14532。Y. Guo、W. Qin、Y. Xu、Y. Gu、C. Yin 和 B. Xia，“协作感知辅助车联网通信中的预测波束跟踪与功率分配”，*IEEE*通信学会开放期刊，第5卷，第6048–6063页，2024年。[10] G. Charan、T. Osman、A. Hredzak、N. Thawdar 和 A. Alkhateeb，“基于真实毫米波数据集的视觉-位置多模态波束预测”，发表于*2022 IEEE*无线通信与网络会议（*WCNC*），2022年，第2727–2731页。[11] J. Morais、A. Bchboodi、H. Pezeshki 和 A. Alkhateeb，“真实世界中的位置辅助波束预测：GPS位置究竟有多大用处？”，发表于*ICC 2023——IEEE*国际通信会议，2023年，第1824–1829页。[12] Y. Yeo、J. Kim、J. Kim 和 J. Lee，“利用真实世界数据集的多模态感知辅助波束预测”，通信与网络期刊，第27卷，第5期，第412–419页，2025年。[13] H. Kim、T. Roh 和 B. Shim，“面向6G通信系统的多模态感知辅助波束管理”，发表于*2024 IEEE*第*100*届车辆技术会议（*VTC2024-Fall*），2024年，第1–5页。[14] K. Heimann、J. An、L. Hito、S. Saidi 和 C. Wietfeld，“基于实时射线追踪的毫米波网络障碍物感知主动波束管理”，发表于*2025 IEEE*第*101*届车辆技术会议（*VTC2025-Spring*），2025年，第1–5页。[15] A. Alkhateeb、G. Charan、T. Osman、A. Hredzak、J. Morais、U. Demirhan 和 N. Srinivas，“DeepSense 6G：大规模真实世界多模态感知与通信数据集”，*IEEE*通信杂志，第61卷，第9期，第122–128页，2023年。[16] T. Kikuzuki、M. Boloursaz Mashhadi、Y. Ma 和 R. Tafazolli，“聚焦前导码：利用毫米波信道状态信息进行感知”，*IEEE*通信学会开放期刊，第5卷，第6430–6442页，2024年。