基于多模态特征融合的车辆网络波束赋形方法

聂佳莉 崔原豪* 张 迪 张荣辉 穆俊生 景晓军
(北京邮电大学 北京 100876)

摘要：波束赋形技术通过向特定方向发射信号，提高了接收信号的功率。然而，在高速动态的车辆网络场景下，频繁的信道状态更新与波束调整导致系统开销过大；波束与用户位置难以实时对齐，易出现错位现象，影响通信稳定性；复杂路况中的遮挡和信道衰落进一步限制了波束赋形的效果。为了解决上述问题，该文提出了一种基于卷积神经网络和注意力机制模型的多模态特征融合波束赋形方法，以实现感知辅助的高可靠通信。模型首先对传感器采集的雷达、激光雷达数据分别定制数据转换和标准化策略，解决数据异构问题。然后使用三维卷积残差块提取多层次高阶多模态特征后，利用注意力机制模型融合特征并预测最佳波束，实现通信性能的优化。实验结果表明，该文所提方法在高速场景下可达到接近90%的平均Top-3波束预测精度，相比单模态方案性能显著提升，验证了其在提升通信性能和可靠性方面的优越性。

关键词：感知辅助通信；多模态特征融合；雷达信号处理；波束赋形；注意力机制

中图分类号：TN957.52 **文献标识码**：A **文章编号**：2095-283X(2025)04-0994-11

DOI: 10.12000/JR24242 **CSTR**: 32380.14.JR24242

引用格式：聂佳莉, 崔原豪, 张迪, 等. 基于多模态特征融合的车辆网络波束赋形方法[J]. 雷达学报(中英文), 2025, 14(4): 994–1004. doi: 10.12000/JR24242.

Reference format: NIE Jiali, CUI Yuanhao, ZHANG Di, *et al.* Vehicle network beamforming method based on multimodal feature fusion[J]. *Journal of Radars*, 2025, 14(4): 994–1004. doi: 10.12000/JR24242.

Vehicle Network Beamforming Method Based on Multimodal Feature Fusion

NIE Jiali CUI Yuanhao* ZHANG Di ZHANG Ronghui
MU Junsheng JING Xiaojun
(*Beijing University of Posts and Telecommunications*, *Beijing* 100876, *China*)

Abstract: Beamforming enhances the received signal power by transmitting signals in specific directions. However, in high-speed and dynamic vehicular network scenarios, frequent channel state updates and beam adjustments impose substantial system overhead. Furthermore, real-time alignment between the beam and user location becomes challenging, leading to potential misalignment that undermines communication stability. Obstructions and channel fading in complex road environments further constrain the effectiveness of beamforming. To address these challenges, this study proposes a multimodal feature fusion beamforming method based on a convolutional neural network and an attention mechanism model to achieve sensor-assisted high-reliability communication. Data heterogeneity is solved by customizing data conversion and standardization strategies for radar and lidar data collected by sensors. Three-dimensional convolutional residual blocks are employed to extract multimodal features, while the cross-attention mechanism integrates

收稿日期：2024-12-04；改回日期：2025-02-19；网络出版：2025-03-17
*通信作者： 崔原豪 cuiyuanhao@bupt.edu.cn *Corresponding Author: CUI Yuanhao, cuiyuanhao@bupt.edu.cn
基金项目：国家重点研发计划(202304D3)，国家自然科学基金(62171049)
Foundation Items: National Key Research and Development Program of China (202304D3), The National Natural Science Foundation of China (62171049)
责任主编：刘凡 Corresponding Editor: LIU Fan

©The Author(s) 2025. This is an open access article under the CC-BY 4.0 License
(https://creativecommons.org/licenses/by/4.0/)

integrate these features for beamforming. Experimental results show that the proposed method achieves an average Top-3 accuracy of nearly 90% in high-speed environments, which is substantially improved compared with the single-modal beamforming scheme.

Key words: Sensor-assisted communication; Multimodal feature fusion; Radar signal processing; Beamforming; Attention mechanism

1　引言

目前，第5代移动通信(5G)已经在各个行业广泛实施和商业部署[1]。业界对于5G-A和第6代移动通信(6G)[2]发展方向展开了积极的探索。通信网络将向更高频段和多天线技术发展[3]。相较于低频段的传统通信系统，毫米波和太赫兹在传播时会经历严重的损耗，从而导致通信可靠性的下降[4]。为有效解决这类问题，波束赋形已经成为了高频段通信中必不可少的技术[5]。

典型的波束训练过程包括发射机通过导频信号对多个方向进行波束扫描，接收机从接收到的导频信号中提取信道状态信息(如到达角、路径损耗等)并反馈给发射机，随后发射机根据反馈信息调整波束实现优化传输。传统波束训练方法主要分为穷举搜索、分层搜索和基于压缩感知的方法[6−8]。其中，穷举搜索在预定义码本上执行穷举搜索，其高计算复杂度和训练开销限制了实际应用；分层搜索通过逐步缩小扫描范围以减少训练开销，但其精度容易受到初始搜索条件的影响；基于压缩感知的波束训练方法则通过利用信道稀疏性显著降低了训练复杂度，但需要额外的信道建模与先验假设支持。在高动态场景下，这些传统方法普遍存在难以兼顾低延迟、低复杂度和高精度的问题。

为了应对这些挑战，感知辅助的波束赋形方案应运而生[9]。通信系统收发机间的视距链路对用户空间信息具有很强的依赖性[10]，可以利用从环境中收集的多模态先验信息提高通信性能[11]。因此，研究人员开始尝试利用感知数据辅助波束赋形。文献[12]提出了一种针对多输入多输出(Multiple Input Multiple Output, MIMO) 通信感知一体化(Integrated Sensing And Communications, ISAC)系统中联合目标传感和多用户非视距通信优化的波束赋形设计。文献[13]开发了一种扩展的卡尔曼滤波方案来跟踪和预测车辆的状态演变，从而提高了波束对准性能。但该方案需要额外获取其他动力学参数，会带来较高的信令开销。文献[14]采用目标检测锁定图像中的用户位置，结合位置信息进行波束赋形，从而能够更好地利用无线传感进行定位并实现可靠的波束对准。虽然感知信息可以以较低的通信开销辅助波束管理，但同时也带来了相对较长的处理延迟。此外，传统的感知辅助波束赋形方案在恶劣天气条件下会出现信号退化。环境对信号的衰减、散射或吸收都会导致波束赋形的精度降低[15]。

感知辅助波束赋形可以提高动态信道下的信噪比，而动态信道通常是由通信用户的可预测运动引起的。若只依靠通信回波信号的角度或距离估计，易受多径效应影响[16]，进而降低波束赋形的精度和可靠性。相比之下，传感器采集的多模态数据可以增强数据多样性，避免对频谱资源的额外需求。同时，全面的感知信息综合有助于提高复杂环境下波束赋形的可靠性。例如，文献[17]利用深度学习提取来自三维场景激光雷达数据的环境信息，根据此完成毫米波系统的波束赋形。文献[18]提出了一种基于目标检测的毫米波通信波束赋形方案，根据摄像机捕获的视觉图像直接预测最优波束。此外，文献[19]利用特定环境下的历史波束管理数据，提出了一种基于深度学习的波束训练方法，以减少训练开销。文献[20,21]分别提出了针对多模态数据进行波束预测的方案，这些方案通过神经网络提取单模态特征，并采用简单的特征向量聚合方法实现不同模态特征的融合。现有的基于感知数据的波束赋形方案存在模态单一[22]、特征融合不充分[23]等问题，多模态感知数据的特征融合在波束赋形中的应用仍有待探索。

为了解决上述问题，本文提出了一种基于深度学习的多模态特征融合波束赋形方法，以应对高动态通信和复杂路况带来的挑战。首先，针对不同模态感知数据的异构问题，定制了数据转换和标准化策略，以确保每种模式能够有效的贡献数据。其次，为了解决波束赋形中的适应性和泛化问题，该方案采用三维卷积残差块提取多层次高阶多模态特征。然后采用注意力机制模型分别在不同特征层次进行多模态特征融合。最后使用多层全连接得到波束赋形结果。本文通过真实场景采集的多模态数据集，验证了所提方法的有效性和可靠性，且与基线方法相比，本文所提方法表现出更好的性能。

2　系统模型

如图1所示，一个毫米波基站配备了三维激光雷达和77 GHz调频连续波雷达，该基站试图与路

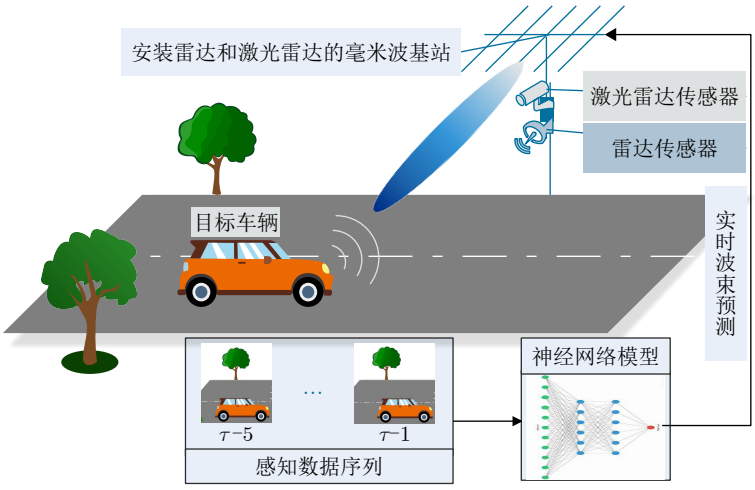

图 1 感知辅助通信场景示意图

Fig. 1 Schematic diagram of sensor-assisted communication scenario

过的车辆通信。在这种情况下，基站可以使用上述传感器获得的感知信息来辅助波束赋形。车辆配备了单个天线以接收基站发送的信号。

2.1 通信模型

系统采用具有N个子载波的正交频分复用波形传输信息。假设基站天线数为M，基站可以基于波束赋形码本$\boldsymbol{G}=\{\boldsymbol{g}_d\}_{d=1}^{D}$来构建波束，其中$\boldsymbol{g}_d\in\mathbb{C}^{M\times 1}$，$D$是码本向量的个数。下行链路中基站和车辆之间的信道被认为是窄带信道，表示为$\boldsymbol{h}\in\mathbb{C}^{M\times 1}$。车辆接收的信号可以表示为

$$y=\boldsymbol{h}^{\mathrm{H}}\boldsymbol{g}_d x+s \tag{1}$$

其中，$\boldsymbol{g}_d$波束赋形向量，发射信号x的平均功率为$E(|x|^2)=P$，s是信号噪声，其服从复高斯分布$N_C(0,\sigma^2)$。当信道状态信息$\boldsymbol{h}$和码本$\boldsymbol{G}$确定后，波束赋形的目标是选择最佳波束索引$\hat{d}$使得接收信号功率最大。该问题可以表示为

$$\hat{d}=\underset{d\in\{1,2,\cdots,D\}}{\arg\max}\left|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{g}_d\right|^2 \tag{2}$$

波束赋形器的设计思路是从波束码本$d\in\{1,2,\cdots,D\}$中预测出最优波束的索引，以获得最大化的波束赋形增益。由于毫米波通信系统的波束较窄，可用空间信道数量巨大。在高动态环境下，频繁的信道估计和波束赋形会导致计算开销显著增加，这使得系统难以满足高可靠性和低时延的业务需求[24]。为降低波束训练和波束追踪带来的信令开销和延迟，可以借助传感数据减少信道估计的频率，并结合机器学习方法，主动预测下一时隙的波束指向，从而精准地与目标车辆进行通信，提升系统的通信效率和响应速度。

2.2 问题定义

基站波束赋形器的初始信道信息可以从毫米波雷达和激光雷达数据中获得。我们使用采集连续5帧感知数据预测下一时刻波束指向，感知数据集$\boldsymbol{B}$被表示为

$$\boldsymbol{B}=\{\boldsymbol{R}_{\tau-5},\cdots,\boldsymbol{R}_{\tau-1},\boldsymbol{L}_{\tau-5},\cdots,\boldsymbol{L}_{\tau-1}\} \tag{3}$$

其中，$\boldsymbol{R}_{\tau-i}$和$\boldsymbol{L}_{\tau-i}$分别表示在第$\tau-i$时刻采集的雷达和激光雷达特征数据。本文利用上述多模态数据预测未来的波束指向，从而最大化接收功率。上述问题可以被表示为

$$\begin{aligned}&\max_{\Theta}P_\tau(\Theta)=\left|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{g}_{\hat{d}_\tau}\right|^2\\&\text{s.t.}\ \ \hat{d}(\tau)=\Phi_\Theta(\boldsymbol{B}(\tau))\end{aligned} \tag{4}$$

其中，$\Phi_\Theta(\cdot)$表示待开发的多模态特征融合模型映射函数，Θ为模型参数集合。

本文尝试设计一个参数为Θ的映射函数Φ_Θ实现感知数据到最佳波束索引的映射。我们采用深度学习来优化从传感数据到最佳通信波束形成向量的映射。具体来说，深度学习模型根据给定感知数据集$\boldsymbol{B}$，返回波束成形码本中提供最大增益的波束的索引。在更一般的意义上，可以使模型返回前K个最可能的波束索引的有序集合。待设计的映射函数可以表示为：

$$\Phi_\Theta:\{\boldsymbol{B}(\tau)\}\rightarrow\left\{\hat{d}(\tau)\right\} \tag{5}$$

其中，$\boldsymbol{B}_\tau$表示帧τ期间的感知数据，$\hat{d}(\tau)$表示帧τ处的最佳波束映射。本文通过深度学习方法设计映射函数Φ_Θ，将感知数据映射到最优波束索引。为达到此目标，定义如下优化问题：

$$\Phi_\Theta=\underset{\Phi_\Theta}{\arg\max}\,\frac{1}{L}\sum_{\tau=1}^{L}I\left\{\Phi_\Theta(\boldsymbol{B}(\tau))=\hat{d}(\tau)\right\} \tag{6}$$

其中，L为总样本数，I为指示函数，当$\{\cdot\}$成立时，I=1，否则I=0。通过不断更新参数集Θ，可以确定最佳的映射函数Φ_Θ进行波束赋形。

3　感知数据预处理

本文使用的数据集[25]包含来自4个场景的多模态数据。表1给出了这些场景中样本分布的信息。其中，场景1和场景2的数据是白天收集的，场景3和场景4的数据是在夜间收集的。利用部署在基站上的雷达和激光雷达传感器采集目标车辆的传感数据，并将每$\tau-5$到$\tau-1$时间段内的连续5帧被存储为一个样本集。

基站处采用64波束码本接收机，并将接收功率最强的波束索引作为波束训练的真实标签。波束设计如图2所示，每个波束采用等间距分布在0°到180°之间的211个测量值来衡量。为提高波束赋形精度，对异构的感知数据进行了以下预处理。

3.1　雷达数据

与其他感知信息相比，雷达对运动物体更敏感。在本文的方法中，采集的雷达立方体数据格式如图3所示。原始雷达数据的维度为4×256×250，包含3个维度：天线维度、啁啾维度和采样维度，分别提供目标的角度、速度和距离信息。为了从数据中提取关键特征，本文采用“距离-角度”和“距离-速度”两种预处理映射方法[26]。通过该预处理方法，我们可以将空间和运动信息有效地表示，并为神经网络模型提供输入特征。雷达处理过程步骤如下：

表 1　数据集描述

Tab. 1　Dataset description

场景	采样数	采集时间
场景1	3506	白天
场景2	3235	白天
场景3	3981	夜间
场景4	4431	夜间

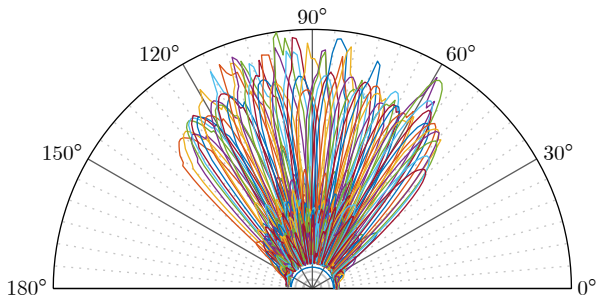

图 2 64波束码本可视化

Fig. 2 Visualization of the 64-beam codebook

“距离-角度”处理方法旨在提取目标的相对距离和角度特征，不同场景的“距离-角度”示意图如图4所示。首先，对数据的距离维度进行快速傅里叶变换(Fast Fourier Transform, FFT)，提取距离信息。随后在速度维度上去均值，消除环境噪声和背景干扰，从而获得更清晰的目标特征。之后在天线维度上进行快速傅里叶变换，从而获得目标角度特征。在速度维度上求和，生成最终的距离-角度矩阵。

“距离-速度”处理方法旨在提取目标的径向速度特征。首先对距离维度进行快速傅里叶变换以提取频域中的距离特征，然后在速度维度上进行快速傅里叶变换，以获得目标的速度特征。最后，计算数据幅值并在天线维度上求和，构建最终的“距离-速度”矩阵，如图5所示，其反映了物体的距离和速度关系。

通过以上两种方法得到的“距离-角度”和“距离-速度”矩阵分别包含了目标在空间中的方位和运动信息。如图6所示，鉴于激光雷达不提供明确的速度信息，我们将雷达的距离-角度图与距离-速度图堆叠起来输入网络，以保留移动汽车的速度信息。相比原始数据，这种预处理方式在降低数据维度复杂性的同时保留了目标车辆相关的重要信息。

3.2　激光雷达数据

激光雷达数据被表示为点云，每个点代表一个扫描的离散空间点。与视觉数据相比，激光雷达数据提供了更直观的空间位置信息和深度细节。此外，激光雷达传感器通常具备高密度空间采样功能，从而能够捕获更为详尽的环境信息。

激光雷达数据包括空间位置信息矩阵$\boldsymbol{P}\in\mathbb{C}^{A\times 3}$和特征向量$\boldsymbol{F}\in\mathbb{C}^{A\times 1}$。$A$是每个时间帧内三维点的个数。本文使用鸟瞰图将原始点云数据转换为类似图像的表示，该方法保持了点云的基本结构

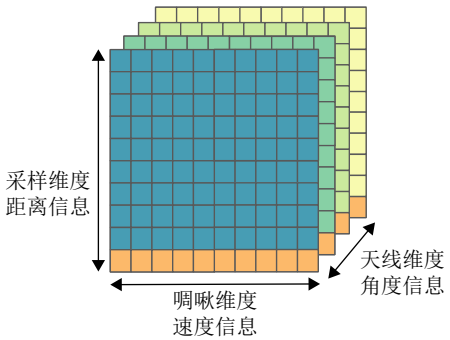

图 3 雷达立方体

Fig. 3 Radar cube

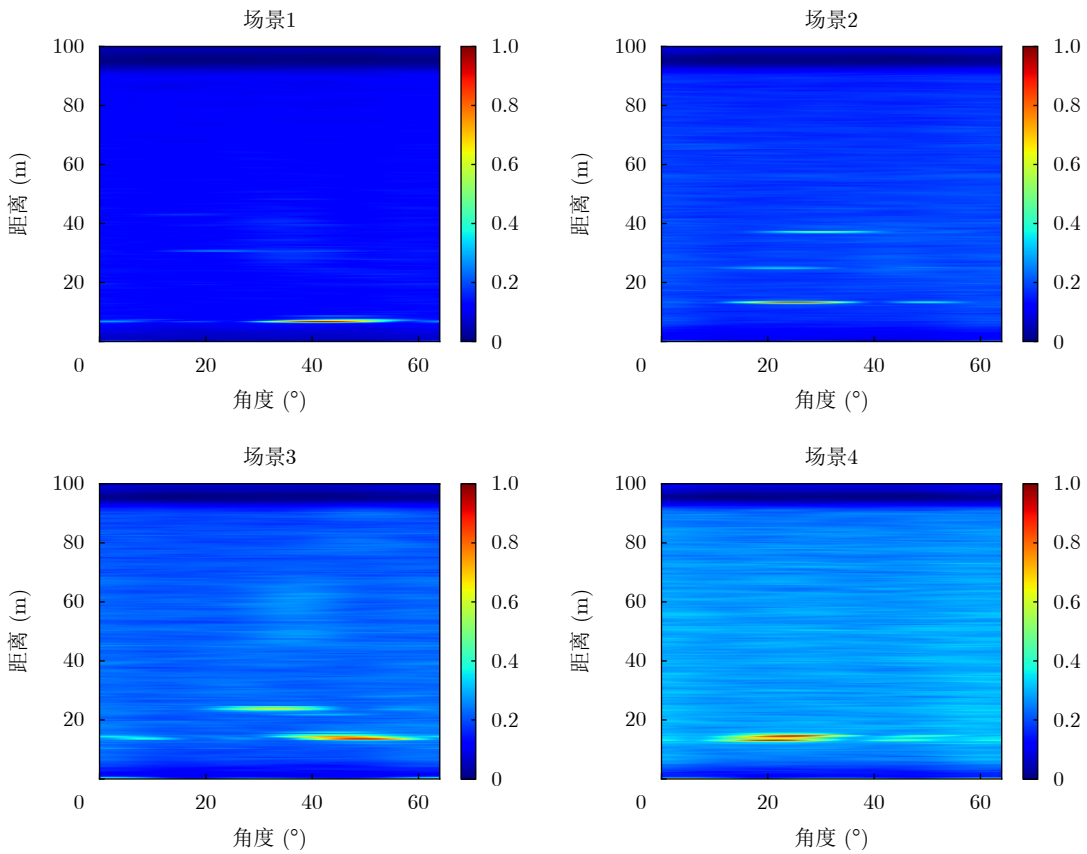

图 4 雷达信号距离-角度特征

Fig. 4 Radar signal range-angle information

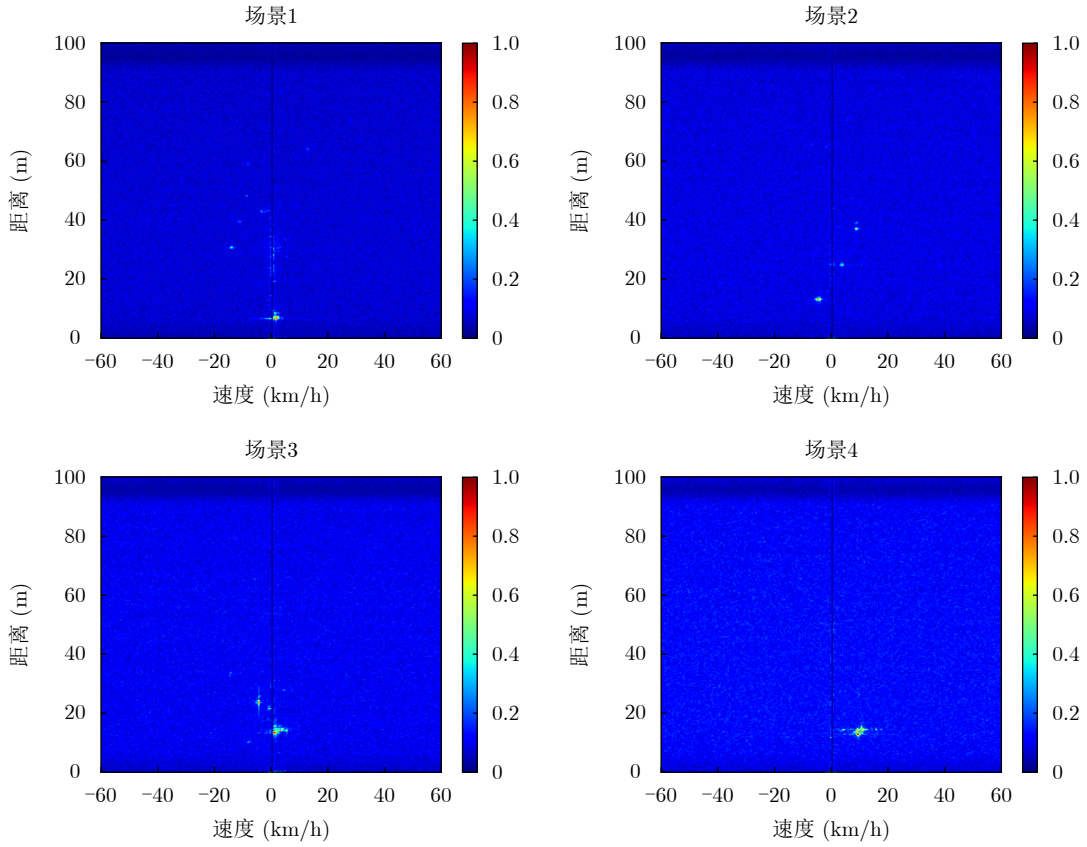

图 5 雷达信号距离-速度特征

Fig. 5 Radar signal range-velocity information

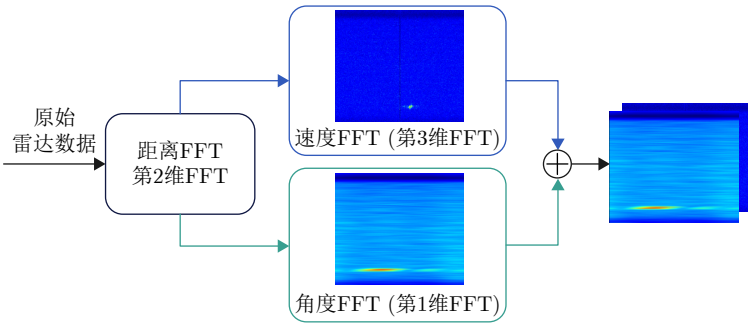

图 6 雷达距离-角度和距离-速度矩阵拼接示意图

Fig. 6 Diagram of radar range-angle and range-Doppler matrix splicing

和深度信息，同时具有较低的计算复杂度。具体来说，投影到横纵坐标分别为$(-X,\ X)$米和$(-Y,\ Y)$米的平面上并离散成二维网格，遍历每个网格单元，确定单元内的最大高度和最大强度值，这些值作为该网格单元的高度和强度编码。计算每个网格内的点的数量来得到点密度。然后将三维点云的高度、强度和密度分别映射到彩色图像的红、绿、蓝通道，形成鸟瞰图[27]。

本文采用点云滤波对激光雷达数据进行处理，从而确保注意力集中在基站和车辆之间的视距路径上。我们首先通过对场景中所有帧的点云数据进行移动平均，生成一个背景模型，用于表示静态物体(如地面、建筑物、树木等)的空间分布。然后，将每个点云帧与该背景模型进行比较，从每帧点云数据中减去背景信息，仅保留车辆周围的动态目标区域。滤除静态对象前后的数据点如图7所示。滤波后的数据更加聚焦于动态对象，有效去除了静态结构和物体。这一过程确保了后续波束赋形算法能够专注于与波束方向调整相关的动态分量。

雷达和激光雷达数据在预处理阶段分别经过去噪、数据对齐等步骤，以消除不同模态间的噪声和时间偏移。随后，通过归一化处理，将数据映射到一个标准化的矩阵表示中，其维度为$C\times5\times256\times256$。其中$C$表示通道数，雷达和激光雷达数据的通道数分别为1和2，5对应于连续5个时间帧，256×256是数据大小。上述预处理步骤有效地解决了多模态数据融合中可能出现的异构性问题。

4　多模态特征融合波束赋形

4.1　模型结构

本文构建了一个多模态特征融合架构，如图8所示。模型首先将雷达和激光雷达数据按时间帧排列，得到2个数据立方体。然后，将时间帧维度作为深度维度，使用三维卷积残差块[28]提取雷达和激光雷达点云数据的特征。在每两次特征提取后使用注意力机制模型[29]融合不同模态的特征。最后，数据经过三层全连接后获得64波束索引。特征提取和多模态融合的具体细节将在后文详细介绍。

4.2　特征提取

激光雷达和雷达数据经过第3节的预处理后进行统一编码。数据经过7×7的卷积后再进行3×3最大池化层压缩特征。最终预处理输出的两个模态编码特征的大小为64×5×64×64。

用于特征提取的三维卷积残差块结构如图9所示。该残差块利用卷积神经网络对输入数据进行多层次特征提取。具体而言，预处理好的数据通过第1个3×3×3卷积层，卷积结果经过批量标准化层后，再经ReLu激活函数处理。经过激活函数的结果通过第2个3×3×3卷积层后再次经过批标准化层。批量标准化和激活函数的使用引入非线性，提升了训练效率和模型稳定性。将上述结果与原始输入数据进行元素级加法运算(残差连接)，然后再经过一层激活层。残差连接有效缓解了深层网络中的梯度消失问题，确保了信息在网络中的高效传递。两次上述过程构成一个三维卷积残差块。模型多次使用三维卷积残差块构建深层神经网络，从而实现对异构感知数据的多模态特征提取。

4.3　多模态特征融合

不同模态的数据对于波束赋形具有不同的重要性和相互依赖性，我们采用注意力机制模块完成特征融合。其中，输入特征通过3个线性变换矩阵映射到查询、键和值向量，分别表示为$\boldsymbol{Q}$, $\boldsymbol{K}$和$\boldsymbol{V}$。

$$\boldsymbol{Q}=\boldsymbol{X}\boldsymbol{W}_Q,\ \ \boldsymbol{K}=\boldsymbol{X}\boldsymbol{W}_K,\ \ \boldsymbol{V}=\boldsymbol{X}\boldsymbol{W}_V \tag{7}$$

其中，$\boldsymbol{X}$为输入特征，$\boldsymbol{W}_Q$, $\boldsymbol{W}_K$和$\boldsymbol{W}_V$为可学习的权重矩阵。在本文模型中，$\boldsymbol{Q}$, $\boldsymbol{K}$和$\boldsymbol{V}$分别涉及不同模态的信息，从而体现特征的交叉关系。计算$\boldsymbol{Q}$和$\boldsymbol{K}$之间的点积后应用缩放机制得到注意力权重，其计算公式为

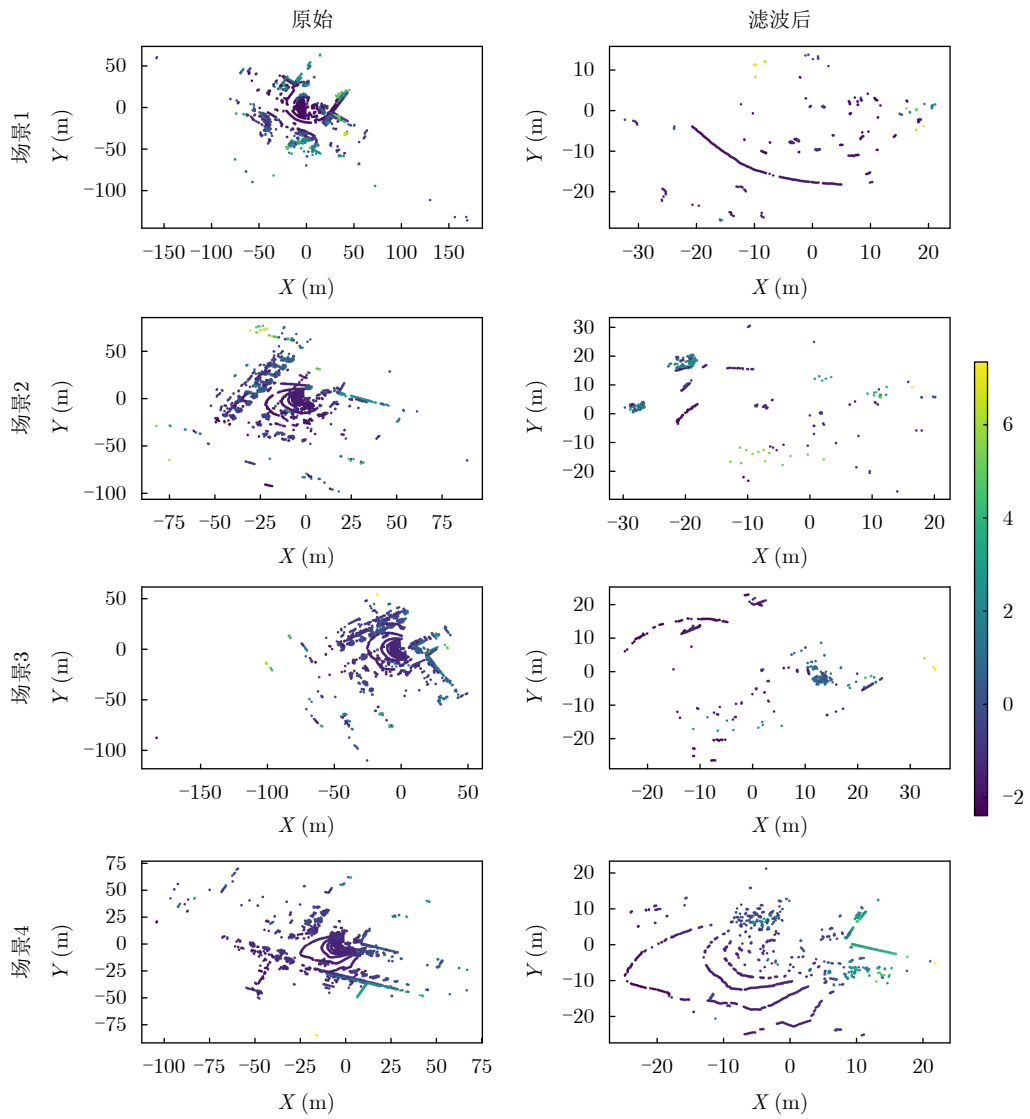

图 7 点云滤波可视化

Fig. 7 Visualization of point cloud filtering

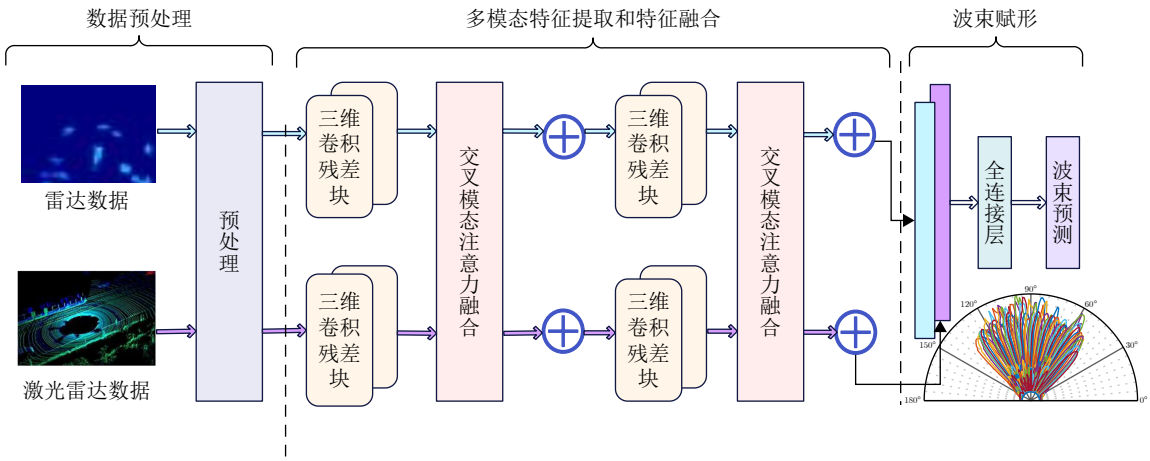

图 8 多模态融合模型架构

Fig. 8 Multimodal fusion model architecture

$$\mathbf{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{8}$$

其中，$(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}})/\sqrt{d_k}$表示查询和键的点积结果除以

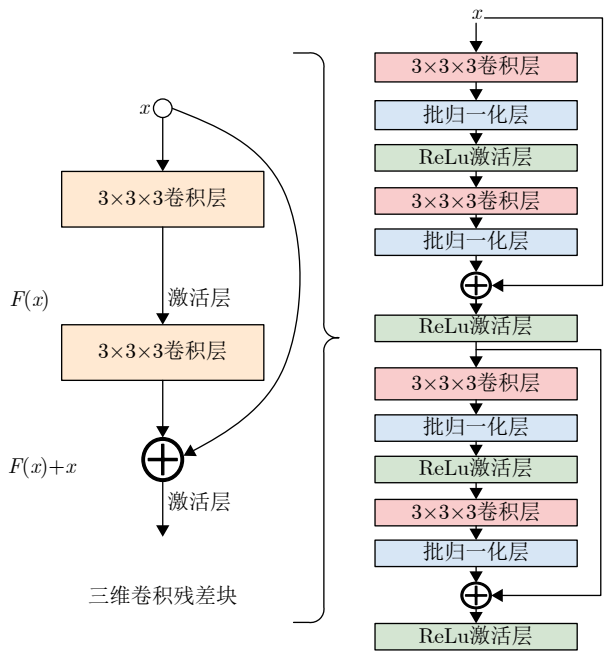

图 9 用于特征提取的三维卷积残差块

Fig. 9 3D convolution residuals block for feature extraction

键的维度的平方根，这样可以防止点积值过大。Softmax函数将点积结果转换为概率分布，得到每个值向量的加权权重。然后，根据权重对$\boldsymbol{V}$进行加权求和得到特征表示。上述结果拼接后进行线性变换，得到多头注意力的输出。最后利用前馈网络对多头注意力的输出进行非线性变换得到输出特征。

图10展示了注意力机制模型在任务中的具体应用细节。两个模态的特征分别作为查询($\boldsymbol{Q}$)、键($\boldsymbol{K}$)和值($\boldsymbol{V}$)输入到基于交叉注意力机制的特征融合模块中，输出和输入大小相同的4个特征图。这一融合机制在多模态特征提取器中被多次应用(图8)。鉴于注意力机制的开销较大[30]，本方案使用自适应池化将特征图下采样到较低分辨率后输入给注意力机制模块，然后利用双线性插值完成分辨率增强，从而有效降低了计算开销。

将经过最后一次特征融合的4个模态数据进行拼接，形成一个维度为17×512的矩阵。随后，在第一维度对该矩阵进行求和操作，得到一个维度为1×1024的向量。这个向量被传递到输出维度分别是256, 128和64的全连接层。这3个层通过映射和非线性变换，生成波束赋形的结果。

5　实验设置与性能评价

5.1　实验设置

实验中使用的FMCW雷达配备了1个发射天线和4个接收天线，采用全数字接收方式。雷达的频率为77 GHz，带宽为4 GHz，最大探测距离为

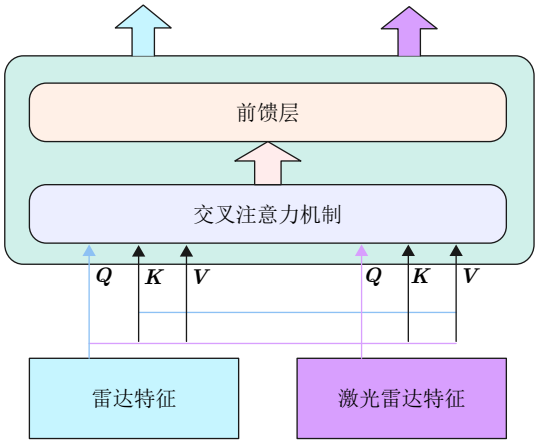

图 10 基于注意力机制的特征融合模块

Fig. 10 Feature fusion module based on attention mechanism

100 m。雷达输出为三维复数I/Q测量数据，尺寸为4(接收天线数量)×256(每啁啾采样数)×250(每帧啁啾数)。此外，激光雷达的测量范围为100 m，最大电机旋转频率为20 Hz，输出为360°的点云采样数据。

本文进行实验评估所提的模型和方法。训练集和验证集数据被随机分成80%用于训练，20%用于验证。模型批量大小设置为8，学习速率设定为从10^{-4}开始，在训练过程中使用余弦衰减调整学习率。模型采用预测波束索引与全局真实波束索引之间交叉熵作为损失函数。模型采用的主要评估指标是前1(Top-1)和前3(Top-3)准确度。前k(Top-k)准确度被定义为其中最佳真实波束索引在前k个预测波束索引内的样本的百分比。

5.2　性能评价

三维卷积残差块的使用次数对模型的收敛性能的影响如图11所示。可看出，模型在前5轮快速收敛，第10轮后趋于稳定。使用2个卷积残差块时，训练和验证损失值较高，且训练过程中波动较大，表现出一定的过拟合现象。而使用4个卷积残差块时，训练和验证损失显著降低，且损失曲线更为平滑，表明模型具有更好的稳定性和泛化能力。本文最终采用4个卷积残差块进行特征提取。

在图12中，以两个波束为例展示了目标波束和预测波束的归一化功率分布。图中实线和虚线分别表示波束7和波束56的真实值和预测值。可以看到，在各个波束索引上，预测的功率值与真实功率值基本趋于一致。这表明模型在整体功率分布预测上具有较高的准确性。模型预测的最大归一化功率对应的波束索引与真实值的波束索引一致，这进一步说明模型在波束选择上做出了正确的判断。

图13展示了所提多模态特征融合方案与文献[20]

方案的性能对比。从图中可以看出，随着迭代轮数的增加，所提方案平均波束赋形精度逐渐提升。在模型精度趋于稳定时，所有场景的平均Top-1精度达到60%以上，平均Top-3精度达到90%左右。与文献[20]的方案相比，本文模型的收敛速度稍有逊色，在训练初期性能提升较慢。在迭代10轮后，所提模型的精度超过了对比模型，验证了所提方案在多模态特征融合的应用中具有更高的性能，能够有效提升波束赋形精度。

模型在不同场景下的Top-3精度变化的情况如图14所示。从图中可以看出，场景2显示出最高的波束赋形精度，这可能由于场景2采集于光线较好的白天，数据中提供了更多有益的信息。同时，随着模型训练，各个场景的波束赋形精度最终都稳定在90%左右。表明模型能够有效地学习并适应不同场景下的波束特性，具有较强的场景泛化性。

图15为分别使用雷达、激光雷达以及多模态融合的波束赋形结果。从图中可以明显看出，多模态融合的Top-1, Top-2和Top-3波束赋形精度分别达到了62.8%, 81.1%和89.6%，远高于单一模态下的对应精度。雷达和激光雷达3种单模态的Top-1精度分别为44.6%和42.1%，而多模态融合后的Top-1精度提高了约20%。模型性能的提升表明多模态融合

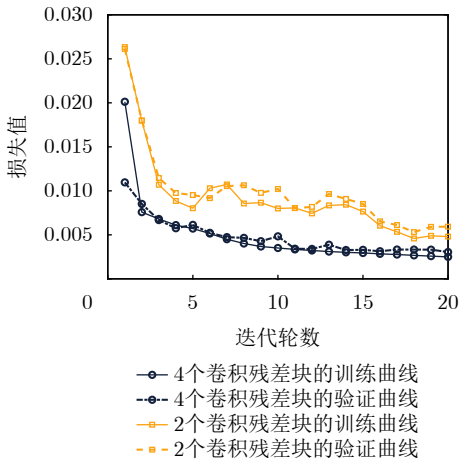

图 11 不同特征提取次数对应的损失曲线

Fig. 11 Loss curves corresponding to different feature extraction times

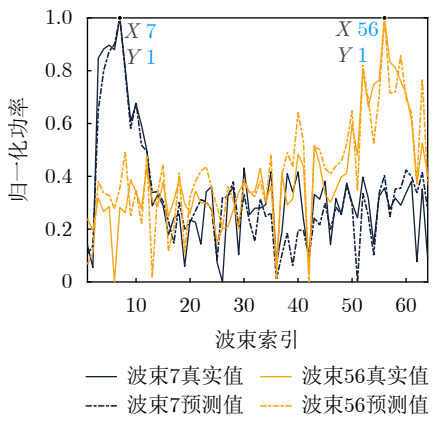

图 12 波束赋形结果

Fig. 12 Beam prediction results

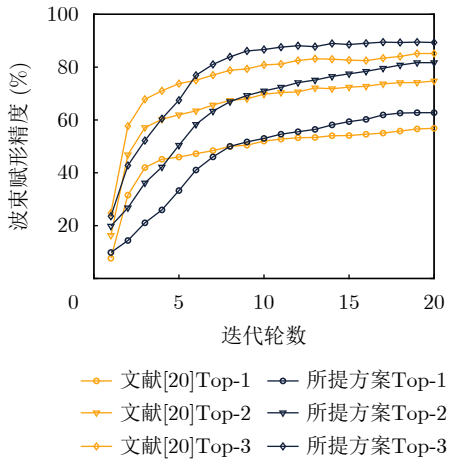

图 13 不同多模态特征融合波束赋形方案性能对比

Fig. 13 Performance comparison of different multimodal feature fusion beamforming schemes

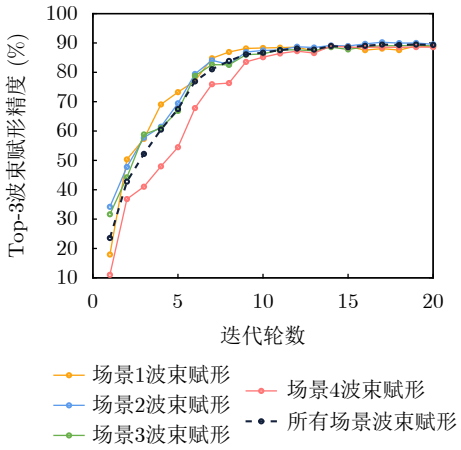

图 14 不同场景的波束赋形精度

Fig. 14 Beamforming accuracy for different scenarios

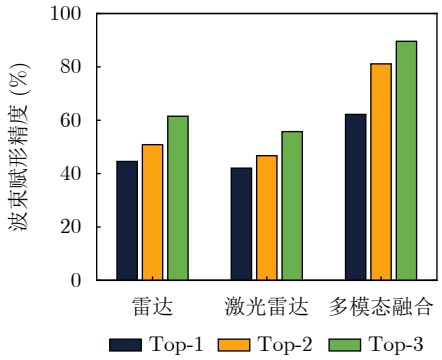

图 15 不同模态的波束赋形精度

Fig. 15 Beamforming accuracy for different modals

能够更全面地捕捉环境信息，从而增强波束赋形的准确性。多模态融合方案不仅能补充单一模态的不足，还能综合多种模态的优点，极大地提升了波束赋形性能。这些结果验证了多模态融合在波束赋形任务中的必要性和有效性。

6　结语

本文提出了一种基于多模态特征融合通信辅助感知进行车辆网络波束赋形的解决方案。通过有效的数据预处理技术过滤掉干扰信息后，使用三维卷积残差块提取多模态的数据特征、采用注意力机制模型融合不同模态的特征。提出的方法利用传感器收集的道路环境信息，在多个场景下实现了较高的波束赋形精度。与传统的波束赋形方法相比，多模态感知辅助波束赋形可以提供更准确的波束赋形。

利益冲突 所有作者均声明不存在利益冲突

Conflict of Interests The authors declare that there is no conflict of interests

参 考 文 献

[1] 王明哲. 5G移动通信发展趋势及关键技术研究[J]. 智慧中国, 2022(2): 68–69.
WANG Mingzhe. Research on the development trend and key technologies of 5G mobile communication[J]. *Wisdom China*, 2022(2): 68–69.

[2] CHEN Wanshi, LIN Xingqin, LEE J, *et al.* 5G-advanced toward 6G: Past, present, and future[J]. *IEEE Journal on Selected Areas in Communications*, 2023, 41(6): 1592–1619. doi: 10.1109/JSAC.2023.3274037.

[3] ZHANG Zhengquan, XIAO Yue, MA Zheng, *et al.* 6G wireless networks: Vision, requirements, architecture, and key technologies[J]. *IEEE Vehicular Technology Magazine*, 2019, 14(3): 28–41. doi: 10.1109/MVT.2019.2921208.

[4] LIU Fan, ZHENG Le, CUI Yuanhao, *et al.* Seventy years of radar and communications: The road from separation to integration[J]. *IEEE Signal Processing Magazine*, 2023, 40(5): 106–121. doi: 10.1109/MSP.2023.3272881.

[5] NIE Jiali, CUI Yuanhao, YANG Zhaohui, *et al.* Near-field beam training for extremely large-scale MIMO based on deep learning[J]. *IEEE Transactions on Mobile Computing*, 2025, 24(1): 352–362. doi: 10.1109/TMC.2024.3462960.

[6] WEI Xiuhong, DAI Linglong, ZHAO Yajun, *et al.* Codebook design and beam training for extremely large-scale RIS: Far-field or near-field[J]. *China Communications*, 2022, 19(6): 193–204. doi: 10.23919/JCC.2022.06.015.

[7] NOH S, ZOLTOWSKI M D, and LOVE D J. Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(9): 5689–5701. doi: 10.1109/TWC.2017.2713357.

[8] ABDELREHEEM A, MOHAMED E M, and ESMAIEL H. Location-based millimeter wave multi-level beamforming using compressive sensing[J]. *IEEE Communications Letters*, 2018, 22(1): 185–188. doi: 10.1109/LCOMM.2017.2766629.

[9] CUI Yuanhao, LIU Fan, JING Xiaojun, *et al.* Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges[J]. *IEEE Network*, 2021, 35(5): 158–167. doi: 10.1109/MNET.010.2100152.

[10] LU Shihang, LIU Fan, LI Yunxin, *et al.* Integrated sensing and communications: Recent advances and ten open challenges[J]. *IEEE Internet of Things Journal*, 2024, 11(11): 19094–19120. doi: 10.1109/JIOT.2024.3361173.

[11] CUI Yuanhao, CAO Xiaowen, ZHU Guangxu, *et al.* Edge perception: Intelligent wireless sensing at network edge[J]. *IEEE Communications Magazine*, 2025, 63(3): 166–173. doi: 10.1109/MCOM.001.2300660.

[12] LIU Fan, MASOUROS C, LI Ang, *et al.* MU-MIMO communications with MIMO radar: From co-existence to joint transmission[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(4): 2755–2770. doi: 10.1109/TWC.2018.2803045.

[13] LIU Fan, YUAN Weijie, MASOUROS C, *et al.* Radar-assisted predictive beamforming for vehicular links: Communication served by sensing[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(11): 7704–7719. doi: 10.1109/TWC.2020.3015735.

[14] NIE Jiali, ZHOU Quan, MU Junsheng, *et al.* Vision and radar multimodal aided beam prediction: Facilitating metaverse development[C]. The 2nd Workshop on Integrated Sensing and Communications for Metaverse, Helsinki, Finland, 13–18. doi: 10.1145/3597065.3597449.

[15] LIU Fan, CUI Yuanhao, MASOUROS C, *et al.* Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond[J]. *IEEE Journal on Selected Areas in Communications*, 2022, 40(6): 1728–1767. doi: 10.1109/JSAC.2022.3156632.

[16] VA V, CHOI J, SHIMIZU T, *et al.* Inverse multipath fingerprinting for millimeter wave V2I beam alignment[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(5): 4042–4058. doi: 10.1109/TVT.2017.2787627.

[17] XU Weihua, GAO Feifei, JIN Shi, *et al.* 3D scene-based beam selection for mmWave communications[J]. *IEEE Wireless Communications Letters*, 2020, 9(11): 1850–1854. doi: 10.1109/LWC.2020.3005983.

[18] YING Ziqiang, YANG Haojun, GAO Jia, *et al.* A new vision-aided beam prediction scheme for mmWave wireless

communications[C]. The 2020 IEEE 6th International Conference on Computer and Communications, Chengdu, China, 2020: 232–237. doi: 10.1109/ICCC51575.2020.9344988.

[19] SHEN L H, CHANG Tingwei, FENG K T, *et al.* Design and implementation for deep learning based adjustable beamforming training for millimeter wave communication systems[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(3): 2413–2427. doi: 10.1109/TVT.2021.3058715.

[20] NIE Jiali, CUI Yuanhao, YU Tiankuo, *et al.* An efficient nocturnal scenarios beamforming based on multi-modal enhanced by object detection[C]. 2023 IEEE Globecom Workshops, Kuala Lumpur, Malaysia, 2023: 515–520. doi: 10.1109/GCWkshps58843.2023.10464587.

[21] SHI Binpu, LI Min, ZHAO Mingmin, *et al.* Multimodal deep learning empowered millimeter-wave beam prediction[C]. The 2024 IEEE 99th Vehicular Technology Conference, Singapore, Singapore, 2024: 1–6, doi: 10.1109/VTC2024-Spring62846.2024.10683225.

[22] GU J, SALEHI B, ROY D, *et al.* Multimodality in mmWave MIMO beam selection using deep learning: Datasets and challenges[J]. *IEEE Communications Magazine*, 2022, 60(11): 36–41. doi: 10.1109/MCOM.002.2200028.

[23] CHARAN G, OSMAN T, HREDZAK A, *et al.* Vision-position multi-modal beam prediction using real millimeter wave datasets[C]. 2022 IEEE Wireless Communications and Networking Conference, Austin, TX, USA, 2022: 2727–2731. doi: 10.1109/WCNC51071.2022.9771835.

[24] CUI Yuanhao, NIE Jiali, CAO Xiaowen, *et al.* Sensing-assisted high reliable communication: A transformer-based beamforming approach[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2024, 18(5): 782–795. doi: 10.1109/JSTSP.2024.3405859.

[25] 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. doi: 10.1109/MCOM.006.2200730.

[26] DEMIRHAN U and ALKHATEEB A. Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration[C]. 2022 IEEE Wireless Communications and Networking Conference, Austin, TX, USA, 2022: 2655–2660. doi: 10.1109/WCNC51071.2022.9771564.

[27] ZHOU Bo, XIE Jiapeng, PAN Yan, *et al.* MotionBEV: Attention-aware online LiDAR moving object segmentation with bird's eye view based appearance and motion features[J]. *IEEE Robotics and Automation Letters*, 2023, 8(12): 8074–8081. doi: 10.1109/LRA.2023.3325687.

[28] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 770–778. doi: 10.1109/CVPR.2016.90.

[29] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 2017: 6000–6010.

[30] HAN Dongchen, PAN Xuran, HAN Yizeng, *et al.* FLatten transformer: Vision transformer using focused linear attention[C]. 2023 IEEE/CVF International Conference on Computer Vision, Paris, France, 2023: 5938–5948. doi: 10.1109/ICCV51070.2023.00548.

作 者 简 介

聂佳莉，硕士生，主要研究方向为通信感知一体化和智能无线通信。

崔原豪，博士，硕士生导师，主要研究方向为通信感知一体化。

张　迪，博士生，主要研究方向为无线通信和通信感知一体化。

张荣辉，博士，主要研究方向为通信感知计算一体化。

穆俊生，博士，博士生导师，主要研究方向为通信感知一体化、智能信号处理。

景晓军，博士，博士生导师，主要研究方向为无线通信。

(责任编辑：于青)