

考试方式： 申请考核 本科直博 硕博连读

身份证号： 33030220020602203X

# 博士生科研计划书

学生姓名： 吴伯涛

报考院系： 计算机学院

报考专业： 计算机科学与技术

研究方向： 软件工程

导师： 张谦

论文题目： 基于视觉语言模型的时空信道预测技术

日期： 2025年12月13日

(注：不区分研究方向和不区分导师的，相应选项可不填。)

## 一、立题依据

## 一、科学意义

随着 6G 移动通信向毫米波/太赫兹高频段及空天地海全场景延伸，无线信道呈现出极高的非平稳性与动态复杂性。在高速移动或复杂散射环境下，信道状态信息 (CSI) 的快速时变导致严重的“老化”效应，极大限制了大规模 MIMO 波束赋形的增益。突破传统的统计建模局限，探索高精度、低开销的信道预测机制，是实现通信感知一体化 (ISAC) 与语义通信愿景的关键科学问题。本项目提出一种基于视觉-语言大模型 (VLM) 的双重引导预测框架，旨在通过跨模态信号表征学习 (Cross-Modal Signal Representation Learning) 重塑物理层信道建模范式。

本项目的核心价值在于打破了传统深度学习方法将射频信号视为单纯数值序列或者单纯图像结构的局限性，创新性地提出了 CSI 视为图像和语言的表征假设。通过将高维 CSI 数据映射为适合 Vision Encoder 处理的伪视觉张量和适合 LLM 处理的信号序列 Token，本项目利用冻结的基础模型强大的跨模态对齐能力与泛化知识，分别从空间与时间两个维度对信道进行双重引导与解耦预测。这不仅能够有效解决长序列预测中的梯度消失与过平滑问题，更能利用大模型的预训练知识增强模型在少样本、低信噪比场景下的鲁棒性。

## 二、国内外现状

无线信道预测技术的发展经历了从统计信号处理到深度表征学习的深刻变革，目前正处于向生成式大模型范式跨越的关键窗口期。纵观国内外研究进展，早期的信道预测主要依赖自回归[1] (AR) 模型、卡尔曼滤波[2]等经典统计方法，这些方法假设信道服从特定的平稳分布，但在复杂多变的移动环境下往往失效。随后，循环神经网络[3] (RNN) 及其变体 LSTM、GRU 以及 NLP 领域的 Transformer[4] 模型、GPT-2[4]模型被广泛应用，它们通过学习 CSI 的时间序列依赖来提升预测精度。然而，这一阶段的方法通常将高维的 MIMO 信道数据视为孤立的数值序列，忽略了 CSI 在天线阵列与频域载波上固有的物理结构，导致模型在面对高维数据时参数冗余严重，且难以捕捉复杂的空间相关性。

为了解决空间特征提取难题，研究逐渐过渡到将 CSI 视觉化的图像建模阶段。鉴于 CSI 矩阵在时频域上呈现出明显的纹理特征，国内外学者开始借鉴计算机视觉 (CV) 的成果，将 CSI 视为二维或三维图像 (Image-based View)，利用卷积神经网络[6] (CNN) 或者复值卷积神经网络[7] (CVCNN) 提取其局部特征。近期，随着 Vision Transformer[8] (ViT) 的兴起，研究[9]进一步将 Swin Transformer 引入该领域，证明了利用自注意力机制处理 CSI 能够比 CNN 更好地捕捉非局部的长距离空间依赖。这一阶段的研究成功挖掘了信道的空间结构信息，但在处理长时间跨度的序列演进方面仍存在局限。

本项目提出直接挖掘射频信号自身的视觉与语言属性的方法，利用预训练的 VLM 进行时空双重特征提取与引导预测的架构设计，是对上述单一模态路径的整合，在国内外范围仍属空白。

## 三、主要参考文献

- [1]Kareem E. Baddour and Norman C. Beaulieu, “Autoregressive modeling for fading channel simulation,” IEEE Trans. Wirel. Commun., vol. 4, no.4, pp. 1650 – 1662, 2005.
- [2]Hwanjin Kim, Sucheol Kim, Hyeongtaek Lee, Chulhee Jang, Yongyun Choi, and Junil Choi, “Massive MIMO channel prediction: Kalman filtering vs. machine learning,” IEEE Trans. Commun., vol. 69, no. 1, pp. 518 – 528, 2021.
- [3]Lemayian Joel Poncha and Jihad M. Hamamreh, “Recurrent neural network-based channel prediction in mmimo for enhanced performance in future wireless communication,” in 2020 International Conference on UK-China Emerging Technologies, UCET 2020, Glasgow, United Kingdom, August 20-21, 2020. 2020, pp. 1 – 4, IEEE.
- [4]Hao Jiang, Mingyao Cui, Derrick Wing Kwan Ng, and Linglong Dai, “Accurate channel prediction based on transformer: Making mobility negligible,” IEEE J. Sel. Areas Commun., vol. 40, no. 9, pp. 2717 – 2732, 2022.
- [5]Boxun Liu, Xuanyu Liu, Shijian Gao, Xiang Cheng, and Liuqing Yang, “LLM4CP: adapting large language models for channel prediction,” J.Commun. Inf. Networks, vol. 9, no. 2, pp. 113 – 125, 2024.
- [6]Jie Wang, Ying Ding, Shujie Bian, Yang Peng, Miao Liu, and Guan Gui, “UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems,” IEEE Access, vol. 7, pp. 96105 – 96112, 2019.
- [7]Chi Wu, Xinping Yi, Yiming Zhu, Wenjin Wang, Li You, and Xiqi Gao, “Channel prediction in high-mobility massive MIMO: from spatio-temporal autoregression to deep learning,” IEEE J. Sel. Areas Commun., vol. 39, no. 7, pp. 1915 – 1930, 2021.
- [8]Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” in 9th International

Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. 2021, OpenReview.net.

[9]Jiaming Cheng, Wei Chen, Jialong Xu, Yiran Guo, Lun Li, and Bo Ai, “Swin transformer-based CSI feedback for massive MIMO,” in 23rd IEEE International Conference on Communication Technology, ICCT 2023, Wuxi, China, October 20-22, 2023. 2023, pp. 809 – 814, IEEE.

## 二、研究内容

### 一、研究目标

本项目旨在针对通信系统中高频段、高移动性场景下信道状态信息（CSI）极易老化且难以精准建模的痛点，构建一套基于双重引导视觉-语言大模型（Dual-Guided VLM）的时空信道预测框架。突破传统通信模型仅依赖统计规律或单一模态特征的局限，验证利用冻结的 VLM 模型解决专用物理层问题的可行性。最终目标是实现对大规模 MIMO 复杂信道在时、频、空多尺度上的高精度预测，显著降低导频开销，提升波束赋形的鲁棒性。

### 二、研究内容

研究面向 VLM 的 CSI 跨模态表征与映射机制。这包括探索如何将复数域的高维 CSI 张量重构为适配 Vision Encoder 输入的伪视觉体征，以此保留波束的空间纹理特征；同时研究将 CSI 序列离散化为适配 LLM 输入的 Token 信号，以此捕捉信道的时序依赖。

研究构建基于冻结预训练的 VLM 的空间引导预测模型架构。重点研究在保持 VLM 预训练参数冻结的前提下，如何利用轻量级的 Q-Former 架构，将信道的先验知识和统计特征转化为提示词（Prompt）来激发 Vision Encoder 对信道空间结构的特征提取能力。

研究设计物理约束下的 Coherence Embedding 策略。拼接在 Q-Former 的输出后，一起引导大模型输出符合物理规律的预测结果，并在公开的信道数据集上进行验证。

### 三、拟解决的关键问题

如何让基于自然图像/语言训练的 VLM 大模型理解无线信道射频信号的特征？

如何在双重引导下平衡 Vision Encoder 的空间特征与 LLM 的时序推理特征，避免信息互斥？

### 三、拟采取的研究方法、技术路线

#### 一、研究方法

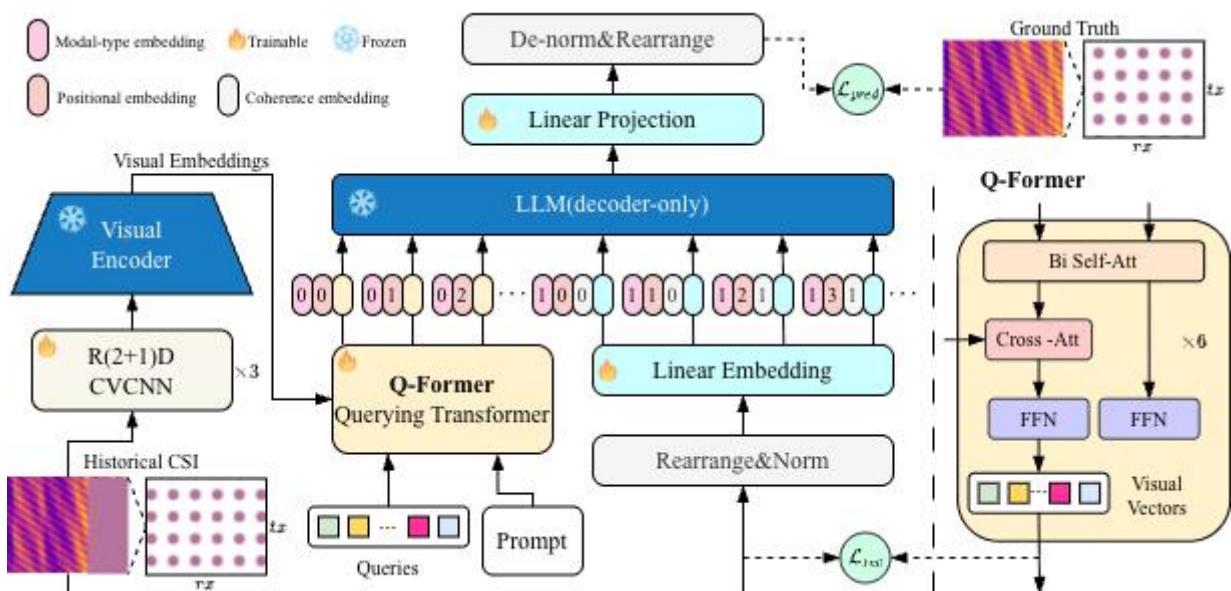
本项目综合采用跨模态表征学习与大模型微调等研究方法。为解决通信领域训练数据不足的问题，本项目直接复用预训练好的 Vision Transformer (Vit) 作为视觉特征提取器，以及预训练好的通用的 Decoder-only LLM，并冻结其绝大部分参数。

采用多视图建模方法。针对无线信道数据的高维特性，摒弃单一的序列化处理。原始 CSI 复数矩阵输入系统后，被分流为两条路径：视觉路径在多个维度上建模信道；序列路径则通过 Rearrange 操作将 CSI 展平，并进行归一化处理，为后续嵌入做准备。在视觉分支，利用 R(2+1)D 复值卷积神经网络对高纬张量进行时空卷积，提取复值 CSI 深层的空间结构特征，以适配预训练 Vision Encoder 的输入；在序列分支，通过线性嵌入层 (Linear Embedding) 将 CSI 序列映射为高维语义向量，并叠加位置编码和模态编码以保留时序特征。

作为核心环节的跨模态融合器 Q-Former 利用初始化的 Learned Queries 输入与冻结的预训练 vision encoder 输出的视觉特征进行多层次交互。同时以包含通信信道全局信息和统计特征的 prompt 为指导，Q-Former 输出的 Token 不仅压缩了冗余的视觉信息，更将其转化为 LLM 能够理解的“视觉提示”(Visual Prompts)。

将 Q-Former 输出的视觉 Token 加上位置编码和模态编码后与序列分支的 CSI Token 进行拼接 (Concatenation) 构成完整的上下文。设计蕴含物理意义的 Coherence Embedding 作为冻结的 LLM 的输入部分，LLM 基于自回归机制推演下一时刻的隐状态 hidden\_states，最后通过线性投影层 (Linear Projection) 与反归一化 (De-norm) 模块，将高维隐变量映射回物理信道空间，输出预测的 CSI 矩阵。

#### 二、技术路线



## 四、预期成果

- 本项目预期将构建一套基于双重引导 VLM 框架的通用时空信道预测理论体系与原型系统。我们将从理论上阐明通用基础模型在物理层信道建模中的迁移机理，验证 CSI 视为图像语言这一表征范式的有效性，为解决通信高移动性场景下的信道老化与波束失效难题提供具有原创性的解决方案。
- 在具体的技术指标与性能表现方面，预期该模型将在公开数据集、QuaDriGa 仿真场景以及实测场景中，相比现有的主流基准模型（如 LLM4CP, Transformer-based predictors），在归一化均方误差 (NMSE) 指标上降低 15%-20%，余弦相似度 SGCS 提升 10%以上。特别是在少样本 (Few-shot) 及未见过的复杂环境下，模型应展现出显著优于传统专用小模型的鲁棒性。
- 在知识产权与学术产出方面，本项目预期在国内外期刊（如 IEEE Journal on Selected Areas in Communications, IEEE Transactions on Wireless Communications）及会议（如 IEEE International Conference on Multimedia and Expo, International Conference on Machine Learning）上发表高水平学术论文 2-3 篇，申请国家发明专利 2-3 项。
- 整理并开源一套包含数据预处理、VLM 适配代码及预训练权重的完整代码库。