The Hong Kong Polytechnic University

Department of Computing

COMP4913 Capstone Project

Proposal

Project Name: Understanding indoor Localization by Deep Learning

Student Name:	CHEN Fengyuan
Student ID No.:	23096069D
Programme-Stream Code:	62435-FFT
Supervisor:	YANG Ray
Submission Date:	Oct. 24th, 2025

Table of Contents

Background and Problem Statement

Indoor positioning technologies have become pivotal for applications such as smart buildings, asset tracking, unmanned retail, and security. Unlike outdoor environments where GPS is available, indoor spaces are rife with reflections, occlusions, and complex structures, resulting in highly dynamic channels and imposing stricter requirements on positioning accuracy, stability, and real-time performance. Among available technologies, Bluetooth has emerged as an important choice in both industry and academia due to its low cost, low power consumption, and ease of deployment. However, traditional Bluetooth positioning methods are constrained by large fluctuations in signal strength, making it difficult to achieve stable sub-meter or better accuracy in complex scenarios.

In recent years, array techniques focusing on phase and Angle of Arrival (AoA) and deep learning have pushed the performance frontier of indoor positioning. The iArk platform achieves protocol-agnostic phase estimation and spatial construction via "massive antenna arrays + side channels", enabling high-precision AoA and triangulation across protocols, significantly mitigating multipath interference and providing an engineering-feasible hardware and middleware solution (An et al., 2020). At the algorithmic level, Zhao et al. (2024) further introduced Transformers into the positioning task, using A-Subnetworks learn AoA, T-Subnetwork learns triangulation and motion context, and the pre-trained model LocGPT supports cross-scenario transfer, substantially reducing data requirements and deployment costs. These two lines of work, from systems and algorithms respectively, validate the superiority of “spatial spectrum + deep learning” in complex indoor environments.

Building on these advances, this project focuses on the following core question: under multi-gateway Bluetooth AoA acquisition, how can deep learning models effectively represent high-dimensional spatial spectra and perform multi-view fusion to achieve stable, reproducible, and real-time high-precision localization in complex multipath environments? To address this question, the project must systematically tackle four challenges: (1)observation aliasing caused by multipath interference and distance attenuation; (2)representation learning of high-dimensional spatial spectra and temporal context modeling; (3)the impact of multi-gateway spatial geometry on localization observability and robustness; and (4)the trade-off between cross-scenario generalization and engineering real-time performance.

The significance of this project lies in proposing and validating an end-to-end “spatial spectrum–driven” deep learning solution grounded in Bluetooth’s cost and power advantages. It will incorporate iArk’s spatial spectrum construction philosophy and introduce Transformer-based multi-view fusion and context modeling, with the potential to significantly improve accuracy, robustness, and deployability of indoor positioning without requiring expensive ultra-wideband (UWB) or complex hardware modifications.

Objectives and Outcomes

This project aims to achieve the following objectives:

1. Data and mechanism understanding: Systematically analyze the statistical properties of Bluetooth spatial spectra, visualize multipath and its relationship with accuracy, and quantify the effects of gateway count, placement, and directivity on observability.
2. Model design and implementation: Build and compare multiple deep learning localization models; at a controlled scale, draw on the division of labor in A-Subnetworks/T-Subnetwork to form an "AoA first, triangulation next” hierarchical architecture.
3. End-to-end evaluation: Train and test on a given dataset; evaluate localization error (mean/median/90th percentile), inference latency, and throughput.
4. Methodological comparison and recommendations: Compare with traditional geometric methods and CNN/MLP baselines; provide recommended configurations and deployment guidelines for engineering practice.

Expected outcomes include: (1)a reproducible deep learning–based Bluetooth indoor localization prototype (including data processing, training/inference, and visualization toolchain); and (2)a quantitative report and visualizations, offering configuration recommendations and experiential summaries for different scenarios.

Project Methodology

3.1 Data Analysis and Preprocessing

The project will use the spatial spectrum as the core feature carrier. We will conduct exploratory analysis on raw data (distribution, missing values, anomalies), adopt robust normalization and outlier suppression strategies, and construct temporal sample segments via sliding windows to enhance robustness against transient interference. To characterize multipath stability, we will compute statistics such as spectral peaks, sidelobe ratio, and spectral entropy, and establish their correlations with localization error.

3.2 Model Design

- Model choices: Explore CNNs (for spatial spectral feature extraction), RNN/LSTM/Transformer (for temporal modeling), and MLPs.
- Multi-gateway information fusion: Design multi-input branches or fusion layers to fully leverage spatial information from different gateways.

3.3 Training and Experimental Design

This stage will include a proper split of training/validation/test sets to prevent sample leakage, with cross-validation to assess stability and generalization. We will set reasonable ranges for core hyperparameters such as learning rate, batch size, and network depth, and tune them iteratively for improved convergence speed and accuracy. During training, we will continuously monitor the loss and key evaluation metrics, employing early stopping to terminate training when performance ceases to improve, thus suppressing overfitting and conserving compute resources.

3.4 Results Analysis

We will systematically organize the localization outputs, perform statistical analyses of positioning accuracy and error distributions, standardize evaluation protocols, and compile aggregate and stratified accuracy metrics. Visualizations will illustrate concentration and dispersion characteristics of errors, clearly depicting model performance. Under identical data and settings, we will conduct parallel evaluations of traditional methods (e.g., triangulation) and deep learning methods, ensuring comparable training and testing conditions for direct comparison and interpretation of differences, culminating in a comprehensive summary of both approaches.

3.5 Literature Review

In the early and middle stages of the project, we will conduct a continuous review of related work, focusing on two main lines: (1) iArk by An et al. (2020), which uses massive arrays and side channels to achieve protocol-agnostic phase estimation and generate stable spatial spectra, adopting an end-to-end "AoA first, triangulation next" pipeline to achieve high-precision localization under multi-protocol conditions; and (2) Zhao et al. (2024), who standardize spatial spectra into tokens for Transformer-based modeling, with the encoder learning AoA and the decoder performing triangulation, combined with semi-supervised losses and pretraining (LocGPT) to enhance cross-scenario transfer and data efficiency. These works provide direct references for spatial spectrum representation, model partitioning, and evaluation baselines in this project.

Project Schedule

日期	任务内容	里程碑/输出
9.1-10.31	● Literature review ● related work survey	Completion of a proposal
11.1-12.8	● Dataset analysis ● Preprocessing dataset analysis	Data exploration report and preprocessing scripts
12.9 - 1.9(midterm submission) - 1.27	● Network architecture selection ● Preliminary modeling (Complete the mid-term report and presentation video according to progress)	Implementation of a basic deep learning localization model (And midterm report)
1.28-2.28	● Model training and tuning ● Experimental design	Obtain preliminary experimental results
3.1-3.31	● Results analysis ● Model comparison ● Visualization	Localization result visualizations and analysis report
4.1-4.10	● Summary and final report writing	Complete project report

Resource Estimation

Hardware:

- Laptop with RTX 4060 Laptop GPU and i9-14900HX CPU
- System memory: 32 GB

Software:

- Operating system: Windows 11
- Python 3.7+
- Deep learning framework: PyTorch
- Data analysis toolkits: numpy, pandas, matplotlib, scikit-learn, etc.
- IDE: VSCode, etc.

References

An, Z., Lin, Q., Li, P., & Yang, L. (2020). General-purpose deep tracking platform across protocols for the internet of things. *Association for Computing Machinery*, 94–106. https://doi.org/10.1145/3386901.3389029

Zhao, X., Wang, G., An, Z., Pan, Q., & Yang, L. (2024). Understanding localization by a tailored GPT. *Association for Computing Machinery*. https://doi.org/10.1145/3643832.3661869