Leveraging Semantics Information to Recognizing Seen and Unseen Locations Using a Mobile Rover

M.Tech ICT(ML)
Dhirubhai Ambani Institute of Information and Communication Technology
Jigar Shekhat-202211004@daiict.ac.in
Supervisor: Prof. Rahul Mishra
Co-Supervisor: Prof. Tapas Kumar Maiti

***Abstract*—The complicated nature of rapidly changing environments is not well captured by traditional static mapping techniques. To solve this problem, we describe a novel initiative that utilizes cutting-edge technology and innovative approaches to provide mobile robots with the ability to acquire information about their surroundings. First, We examine Graph-Structured Sum-product networks (GraphSPNs) are a probabilistic method for structured prediction in circumstances where arbitrary, dynamic graphs represent latent variable dependencies. While many structured prediction methods impose strict limits on the relationships between inferred variables, many real-world Problems can only be described using complicated graph structures of various sizes, frequently tainted with noise when derived from actual data. We show how noisy topological relations found by a robot navigating extensive office areas can be used to support an inference about semantic, conceptual place descriptions. We implement a graph neural network to address these problems. However, because this data contains spatial data on the neighboring node, we cannot get reliable conclusions from it. Based entirely on spatial information, we are unable to estimate location. To work on the natural, significant aspects of place. Our project focuses on integrating YOLOv8 object detection, WiFi measurements, and GPS coordinates to create a complete understanding of indoor spaces. By collecting this knowledge regarding places, Our project can build semantic information as a form of GPS, WiFi, and Objects. The key objective is to enable mobile robots to identify their location, even when GPS signals are weak or unavailable, providing relevant information about the places they navigate. This project exceeds the limitations of static mapping by offering dynamic, real-time insights into constantly changing environments. To achieve robust location identification, we employed Logistic Regression, Naive Bayes, and Random Forest. These models effectively categorize place names and types, offering mobile robots a comprehensive understanding of their surroundings. Our project represents a significant advancement in location identification for mobile robots in indoor environments. Combining state-of-the-art object detection, WiFi measurements, and GPS coordinates with a machine-learning framework equips robots with the knowledge and adaptability needed to excel in dynamic surroundings.**

***Index Terms*—Object Detection, WiFi Measurements, GPS Coordinates**

I. Introduction

The field of robotics continues to develop, and mobile robots must be able to see and understand their surroundings, especially in dynamic interior situations. For a mobile robot to remain aware of its surroundings, it needs a representation of spatial information, a framework that organizes the understanding of the environment. Constantly changing environments provide challenges for traditional mapping and localization approaches. A representation should take advantage of established spatial links in order to integrate knowledge across geographical scales and levels of abstraction. Topological maps are a well-established framework for characterizing the spatial connections between neighbouring locations. They make it easier to establish high-level conceptual data and provide planning algorithms easy access. To integrate the acquired geographical knowledge, resolve ambiguities, and provide predictions, such frameworks often make use of structured prediction algorithms. Unfortunately, a robot exploring a real-world environment often finds complicated and noisy relations, which poses difficult inference problems. As the robot explores its surroundings, the number of nodes and relations in topological maps increases and varies based on the surroundings. However, robots will not be able to obtain reliable results if they simply have spatial knowledge about their surroundings. We describe an innovative approach that uses WiFi measurements, GPS coordinates, and YOLOv8 object identification to overcome that challenge and enable mobile robots to completely understand their surroundings. This initiative has the potential to completely change how mobile robots react and interact with indoor environments. The immense complexity and dynamic of indoor surroundings are unique features. The dynamic nature of indoor places is frequently difficult for traditional static mapping systems to adapt to because they depend on pre-existing maps or infrastructure. In tackling this problem, our research uses innovative strategies to allow mobile robots to navigate well despite the lack of dependable GPS signals. This allows them to recognize their location and comprehend the environments they navigate. Our project's main element is a comprehensive dataset carefully gathered to represent various indoor features. Several types of information are included in this collection, such as GPS coordinates, WiFi signal measurements, place categories, building names, and photos of various inside spaces. We added YOLOv8 object identification, a state-of-the-art deep learning model, to this dataset to enhance its ability to identify things unique to each location. By applying semantic understanding to the dataset, this method helps the system identify and classify locations according to the things in them. Our approach to improving position recognition accuracy and resilience, even in scenarios with poor or inconsistent GPS signals, heavily relies on this semantic understanding. We used machine learning models like Logistic Regression, Naive Bayes, and Random Forest to further improve the system's capabilities. These models act as the system's brain, allowing location names and categories to be accurately classified. Mobile robots are able to adapt to their environment and carry out duties successfully by efficiently classifying these features. This project has extensive potential effects that extend methods in addition to robots. This technology can be extremely useful for mobile robots in various industries, including security, logistics, and healthcare. They are flexible advantages since they can move around and navigate effectively in many environments. Our approach also presents a novel function for recognizing places that have never been observed. When the robot visits a new location for the first time, it gathers information and allocates a special tag. Subsequent visits to the same location result in rapid and accurate recognition. Our study is a major advancement in mobile robot location recognition in dynamic indoor situations. By integrating advanced object identification, WiFi measurements, and GPS coordinates with a robust machine-learning framework, we equip robots with the intelligence and flexibility required to perform well in dynamic situations. The technology has the strength to improve indoor navigation and enable a broad range of applications, changing the capabilities of mobile robots in a variety of industries. As the project progresses, we will examine the particulars of each part and show how the system could impact indoor robots in the future.

II. Background

We give a thorough overview of the project's background in this section. Initially, we provide a thorough explanation of the characteristics of GraphSPN and GCN, the process of obtaining inferences, and the techniques used for learning the network's parameters and structure. The spatial knowledge model that we utilize to facilitate environmental awareness for mobile robots is called Deep Affordance Spatial Hierarchy (DASH), and it is available in the COLD dataset. We discuss topological maps in general and the DASH topological map generation process[14][17]. After that, we describe that This Project focuses on the challenges of indoor localization, the significance of YOLOv8 object detection, WiFi frequency analysis, and GPS signal utilization. The challenges of signal attenuation and multipath effects in indoor environments are highlighted, which impact traditional GPS-based methods. YOLOv8's real-time object detection capabilities and WiFi frequency analysis's ability to capture spatial signal distributions are explained. While GPS signals are less reliable indoors, their potential for contributing to indoor context is acknowledged. The concept of multimodal data fusion, combining visual, wireless, and satellite-based data, is introduced to enhance indoor localization accuracy[1].

A. Localization

One essential prerequisite for mobile robots functioning in interior environments is precise localization. Finding a robot's location in relation to its environment is what it involves. Traditional GPS-based localization is sometimes insufficient in dynamic interior environments due to the vulnerability of GPS signals to multipath interference and signal blockages, which can cause the signals to become nonexistent or inaccurate. Other techniques, such as WiFi-based location, have become more popular to overcome this difficulty. WiFi positioning uses an estimate of a device's location by using the signals from wireless access points. Robots can now traverse within buildings without GPS signals because of the effectiveness of this technology's indoor localization solutions.

B. Object Detection

For mobile robots, object recognition is essential to contextual comprehension. It involves identifying and recognizing things in the robot's surroundings. The cutting-edge deep learning-based object identification system YOLO (You Only Look Once) is excellent at recognizing objects in real time. YOLO's integration into the project gives the system the capacity to recognize and classify things that are exclusive to particular interior spaces. By recognizing things typical of particular locations, this semantic knowledge improves the system's capacity to recognize and navigate interior surroundings.

C. Spatial Signal Distributions of WiFi

Indoor spaces are frequently linked to WiFi signals, which provide useful information for spatial signal distributions and localization. Spatial signal distributions may be obtained by examining neighbouring WiFi access points' frequencies and signal strengths. A robot's position may be determined by observing the unique WiFi signal patterns that differ across places. In dynamic contexts where conventional mapping might not be sufficient, this technique has advantages. In order to improve the robot's contextual awareness, the project makes use of WiFi frequency analysis, which may supply the system with essential data for location recognition.

D. GPS Signal

Although GPS signals are a dependable method for location in open spaces, they pose difficulties when used indoors. GPS signals have limited penetration capabilities, which causes poor or nonexistent signals within structures. Incorrect location estimations can also result from signal obstructions and multipaths. This study acknowledges that GPS signals have limited use in interior environments and uses other techniques, such as WiFi measurements and object recognition, to improve position identification. Rather than being abandoned, GPS signals are used in conjunction with other technologies to increase the overall accuracy of localization in dynamic interior contexts.

III. Literature Survey

- Research by Guvenc and Chong provides a comprehensive survey of Time of Arrival (TOA) based wireless localization and non-line-of-sight (NLOS) mitigation techniques. Understanding WiFi signal strengths and distributions is crucial for our project's localization algorithm.[4]
- Autonomous indoor mobile robots must have WiFi localization and navigation. Examining the latest techniques

in WiFi-based localization offers insightful information that helps us improve the strategy of our project.[2]

- Satellites facilitate accurate positioning and navigation across outdoor environments. This integration not only enhances the accuracy of the overall system but also ensures a continuous and reliable localization experience as mobile robots traverse between diverse environments.[3]
- ORB-SLAM, a monocular Simultaneous Localization and Mapping (SLAM) system. Understanding visual SLAM techniques enhances your project's capability to integrate object detection from images into the localization process. [5]
- This survey focuses on cooperative localization in wireless sensor networks, providing insights into the fusion of multiple sensor modalities. Understanding how different sensors collaborate enhances our project's adaptability to diverse environmental conditions. [7]
- This study explores challenges and opportunities in localization using Radio-Frequency Identification (RFID). Investigating challenges enhances our project's ability to address potential hurdles in dynamic indoor environments.[16]
- The paper discusses Radio Frequency (RF) fingerprinting for indoor localization, emphasizing Received Signal Strength (RSS)-based techniques. Investigating the strengths and challenges of fingerprinting is relevant for our project's WiFi-based localization.[15]

IV. Techniques used in this Work

A. Sum-Product Network (SPN)

SPN, proposed by Poon and Domingos, is a new class of probabilistic graphical models with built-in properties that allow tractable inference, a major advantage over traditional graphical models such as Bayesian networks. The idea is built upon Darwiche's work on network polynomials and differentials in arithmetic circuit representation of the polynomial [5]. Here, we provide the definition of SPN and several of its key properties[8].

Definition 1: **Sum-Product Network** (SPN) Let $X = \{X_1, \cdots X_n\}$ be a set of variables. A Sum-Product Network(SPN) defined over X is a rooted directed acyclic graph. The leaves are indicators. The internal nodes are sum nodes and product nodes. Each edge (i, j) from sum node i has a nonnegative weight w_{ij}. The weight of a sum node is $\sum_{j \in Ch(i)}, w_{ij} v_j$, where $Ch(i)$ is s the children of i. The value of a product node is the product of the values of its children. The value of an SPN is the value of its root.

We use S to denote an SPN as a function of the indicator variables (i.e. the leaves). Let x be an instantiation of the indicator variables, a full state. Let e be an evidence (partial instantiation). For a given node i, we use S_i to denote the sub-SPN rooted at i. Also, we use x_p^a to mean $[X_p = a]$ is true, and use $x_p^- a$ to mean the negation, for simplicity. We define the following properties of SPN.

Figure 1: SPN implementing a naive Bayes mixture model (three components, two variables).

Fig. 1: A simple SPN for a naive Bayes mixture model P(X1, X2), with three components over two binary variables.

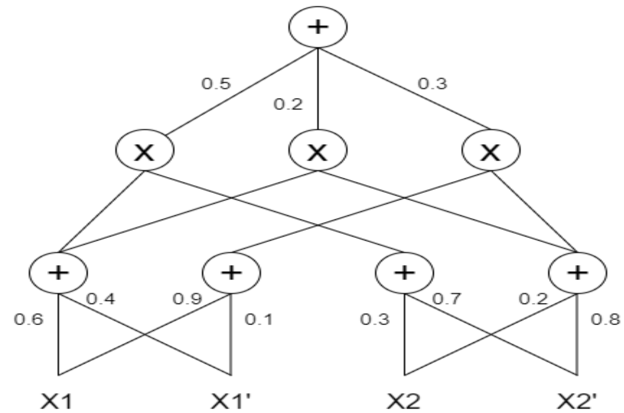

Fig. 1: Example of SPNs.

The bottom layer consists of indicators for X1 and X2. Weighted sum nodes, with weights attached to inputs, are marked with +, while product nodes are marked with ×. we will assume (without loss of generality) that sums and products are arranged in alternating layers, i.e., all children of a sum are products or leaves, and vice-versa.

Example 4.1: Considering the SPN in Figure 1, $S(x_1, x_2, \overline{x}_1, \overline{x}_2) = 0.5(0.6x_1 + 0.4\overline{x}_1)(0.3x_2 + 0.7\overline{x}_2) + 0.2(0.6x_1 + 0.4\overline{x}_1)(0.2x_2 + 0.8\overline{x}_2) + 0.3(0.9x_1 + 0.1\overline{x}_1)(0.2x_2 + 0.8\overline{x}_2)$ The network polynomial $S(x) = (0.5 \times 0.6 \times 0.3 + 0.2 \times 0.6 \times 0.2 + 0.3 \times 0.9 \times 0.2)x_1 x_2 + \cdots$ If a complete state x is $X_1 = 1$, $X_2 = 0$, then $S(x) = S(1, 0, 0, 1)$ If The evidence e is $X_1 = 1$, then $S(e) = S(1, 1, 0, 1)$. Finally, $S(*) = S(1, 1, 1, 1)$.

B. Topological Graphs

GraphSPNs are applicable to arbitrary graphs. However, here our dataset is specifically a topological graph built by a mobile robot exploring a large-scale environment. The primary purpose of our topological graph is to support the behaviour of the robot. As a result, nodes in the graph represent places the robot can visit and the edges represent navigability. The graph nodes are associated with latent variables representing semantics and the edges can be seen as spatial relations forming a global semantic map. Local evidence about the semantics of a place might be available and we assume that such evidence is inherently uncertain and noisy. Additional nodes in the graph are created to represent exploration frontiers, possible places the robot has not yet visited, but can navigate to. We call such nodes placeholders and assume that the robot has not yet obtained any evidence about their semantics. The topological graph is assembled incrementally based on a dynamically expanding 2D occupancy map. The 2D map is built from laser range data captured by the robot using a grid mapping approach based on Rao-Blackwellized particle filters. Placeholders are added at neighbouring, reachable, but unexplored locations and connected to existing places. Then, once the robot performs an exploration action, a placeholder is converted into a place and local evidence captured by the robot about the semantic place category is anchored to the graph node [18]. An example of such a semantic-topological map is shown in Fig. 2

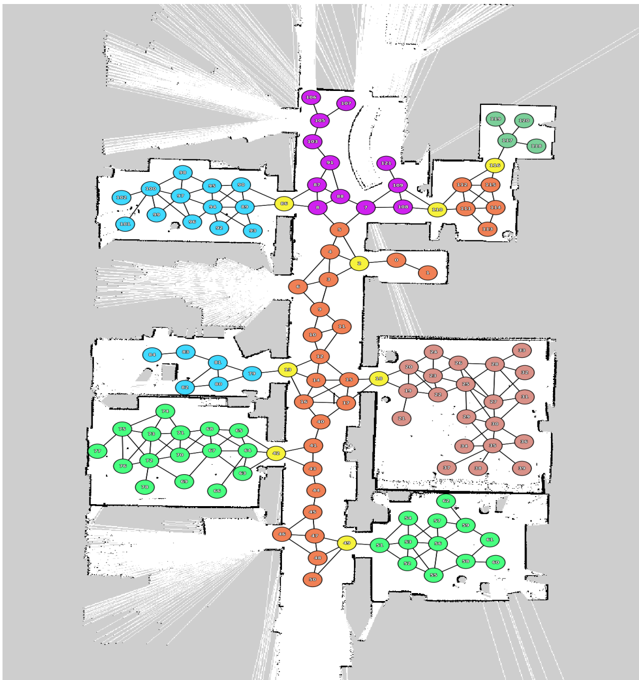

Fig. 2: Topological graph (Stockholm-Floor4)

C. GraphSPNs

GraphSPNs learn a template model over arbitrary graph-structured data, with local evidence Xi and latent variables $Y_i = \{Y_{i1}, \cdots, Y_{iM}\}$ for each graph node or edge i, with dependencies between the latent variables expressed in terms of the graph structure. Then, an instance of GraphSPN distribution $P(X_1, Y_1, \cdots, X_N, Y_N)$ is assembled for a specific graph to perform inference. To this end, we define a set S of template sub-graphs, and associate each template sub-graph S $\in \mathbb{S}$ with a separate template SPN modelling the distribution over variables X_i and Y_i corresponding to the nodes and edges of the template sub-graph. The structure and parameters of each template SPN can be learned directly from data obtained by decomposing training graphs into sub-graphs corresponding to $\mathbb{S}$.

Given a set of trained template SPNs, and a specific the graph to be modelled, an instance of GraphSPN is assembled as illustrated in Fig. 2. First, the graph is decomposed multiple times, each time differently, into sub-graphs using sub-graph templates S in descending order of the template size (i.e. more complex templates have priority). The subgraphs should not overlap in each decomposition and the corresponding template SPNs should cover all variables X_i, Y_i in the model. This condition guarantees the completeness and decomposability resulting in a valid instance GraphSPN. For each decomposition and each sub-graph, we instantiate the corresponding template SPN resulting in multiple SPNs sharing weights and structure. The instantiations for a single graph decomposition is combined with a product node and the product nodes for all decompositions become children of a root sum node realizing the complete mixture model.

In order to incorporate the latent variables Y_{ij}, we include an intermediate layer of product nodes into the template SPNs. Each such product node combines arbitrary distribution of ${D_i}_j^k(X_i)$ with an indicator $\lambda_{Y_{ij}} = c_j^k$ for a specific value c_j^k of Y_{ij}. The template SPN is built on top of the product nodes and can be learned from data and the distributions ${D_i}_j^k(X_i)$ can be arbitrary, potentially also realized with an SPN with a data-driven structure[19].

In our experiments, we assumed only one latent variable (semantic place category) Y_i per graph node i, with $Val(Y_i) = \{c^1, \cdots, c^L\}$, and we defined $D_i^k(X_i)$ for a single hypothetical binary observation x_i, which we assumed to be observed:

$$D_i^k(X_i) = \begin{cases} \alpha_i^k & \text{if } X_i = x_i \\ 1 - \alpha_i^k & \text{if } X_i = \overline{x}_i \end{cases}$$

Such simplification allows us to thoroughly evaluate GraphSPNs for the problem of learning topological semantic maps by directly simulating hypothetical evidence about the semantic category of varying uncertainty and under various noise conditions. Furthermore, it allows us to compare GraphSPNs with Markov Random Fields using the same α_i^k as the value of local potentials, i.e. $\phi_i(Y_i = c^k) = \alpha_i^k$. The proposed approach naturally extends to the case where a more complex distribution is used to model semantic place categories based on robot observations.

D. GraphNNs

Graph Neural Networks (GNNs) are a type of neural network that are designed to process graph-structured data. They are used to perform various tasks such as node classification, graph classification, and link prediction. GNNs have become popular due to their ability to handle complex relationships and dependencies in graph data.

GNNs work by using a message-passing algorithm to update the hidden states of nodes in a graph based on the hidden states of their neighbors. The algorithm involves passing messages between nodes and aggregating the messages to update the hidden states. The process is repeated for multiple iterations until the hidden states converge to a stable state. GNNs can be designed with different types of message-passing functions such as convolutional or attention-based functions.

There are different architectures of GNNs such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), GraphSAGE, and Graph Isomorphism Networks (GINs). Each of these architectures has its own strengths and weaknesses, and the choice of architecture depends on the specific task and the characteristics of the graph data.

GCNs are one of the most popular GNN architectures and are designed to perform convolution operations on graphs[14].

E. GCNs

Graph convolutional networks (GCNs) (Kipf & Welling, 2017) generalize convolutional neural networks (CNNs) (LeCun et al., 1995) to graph-structured data. To learn the graph representations, the “graph convolution” operation applies the same linear transformation to all the neighbours of a node followed by a nonlinear activation function. GCNs have become one of the most popular GNN architectures due to their simplicity, effectiveness, and scalability.

GCNs work by propagating information from a node's neighbours to update its own representation. This is achieved through a convolution operation in the graph domain, where the filter is applied to the node and its neighbours. The output of the convolution operation is a new feature representation of the node, which incorporates information from its local neighbourhood.

The mathematical formula for a single layer of GCN can be written as follows:

$$H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$

where $H^{(l)}$ is the hidden state matrix of the nodes in layer l, $\hat{A}$ is the adjacency matrix of the graph with self-loops, $\hat{D}$ is a diagonal node degree matrix of $\hat{A}$, $W^{(l)}$ is the weight matrix of layer l, and σ is the activation function.

The $\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}$ term is known as the normalized adjacency matrix, which ensures that the weight of the connections between nodes is scaled based on their degree. The output of the convolution operation is then passed through the activation function to produce the updated hidden state matrix $H^{(l+1)}$. Multiple layers of GCN can be stacked to learn increasingly complex representations of the nodes in the graph. The output of the final layer can be used for various downstream tasks such as node classification and graph classification[17].

F. YOLOv8

YOLOv8 was released in January 2023 by Ultralytics, the company that developed YOLOv5. YOLOv8 provided five scaled versions: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large) and YOLOv8x (extra large). YOLOv8 supports multiple vision tasks such as object detection, segmentation, pose estimation, tracking, and classification [10]. YOLOv8 is the latest iteration in the YOLO family of detection models, which are known for their capabilities for joint detection and segmentation. Similar to YOLOv5, the architecture consists of a backbone, head, and neck. It boasts a new architecture, improved convolutional layers (backbone), and a more advanced detection head, making it a top choice for real-time object detection. YOLOv8 also offers support for the latest computer vision algorithms such as instance segmentation, enabling multiple object detection in an image or videos. The model uses the Darknet-53 backbone network, which is faster and more precise than the previous YOLOv7 [12] network. YOLOv8 predicts bounding boxes through an anchor-free. [1]

G. WiFi Frequency Analysis

WiFi frequency analysis involves the collection and analysis of WiFi signal strength measurements within indoor environments. WiFi networks emit radio waves at specific frequencies, and the strength of these signals can be measured in decibels relative to 1 milliwatt (dBm). By analyzing the distribution of WiFi signal strengths across an indoor space, it's possible to gain insights into the spatial characteristics of the environment, including signal attenuation, interference, and the presence of obstacles.

Understanding WiFi signal distribution is crucial for indoor localization projects. WiFi signals interact with walls, furniture, and other objects, leading to variations in signal strength across different locations. By mapping these variations, it's possible to create location fingerprints that help identify and differentiate between various indoor spaces based on their unique WiFi signal patterns.

WiFi frequency analysis provides a valuable layer of information that contributes to the overall indoor localization process. By combining WiFi signal strengths with other data sources like YOLOv8 object detection and GPS signals, the project aims to create a comprehensive understanding of indoor spaces. This integrated approach enhances the accuracy of indoor positioning and enables the determination of the best-matching room based on a combination of modalities.

H. GPS Signal Utilization

Global Positioning System (GPS) signals, originally designed for outdoor navigation, can still provide valuable context for indoor localization projects, despite their limitations within enclosed spaces. GPS signals are transmitted by a network of satellites orbiting the Earth and contain information about satellite positions and timestamps. Although GPS signals weaken and experience multipath effects indoors due to signal attenuation caused by walls and structures, they can still contribute to indoor localization in certain scenarios.

In indoor environments, where traditional GPS signals might be weak or entirely absent, utilizing available GPS signals or data derived from weak signals can provide additional contextual information. This context can be integrated with other localization methods, such as WiFi frequency analysis and object detection, to enhance the accuracy of indoor positioning.

Methods of Utilization:

1) **Coarse Location Estimation:** Weak GPS signals might still be received indoors, albeit with reduced accuracy. These signals can provide a rough estimate of the user's location within a building or floor.
2) **Contextual Information:** While weak GPS signals may not provide precise indoor positioning, they can offer additional context. For example, they can help validate or corroborate indoor location estimates derived from other sources.
3) **Hybrid Localization:** Hybrid localization methods combine data from multiple sources to enhance accuracy. GPS signals can be integrated with WiFi frequency analysis, object detection, and other techniques to create a more comprehensive localization solution. These hybrid methods aim to mitigate the limitations of individual techniques and leverage the strengths of each data source.

Challenges:

1) **Weak Signals:** Indoor environments typically receive weaker GPS signals due to signal blockage from struc-

tures. These signals might not be as accurate as outdoor signals.

2) **Multipath Effects:** Multipath interference occurs when GPS signals bounce off walls and structures, leading to signal distortions and inaccuracies. This effect can complicate the use of GPS data for indoor positioning.
3) **Coarse Accuracy:** The accuracy of GPS signals indoors is generally lower than in open outdoor spaces. However, even coarse location estimates can be valuable for certain applications.
4) **Integrity Monitoring:** GPS signal integrity monitoring is essential, as weak or inaccurate signals can lead to erroneous location estimates. Implementing measures to assess the quality of received signals is important.

V. Datasets

A. *COLD Dataset*

Cognitive rObot Localization Database (COLD), a large database of localization, sensory information and local environment representations (place scans), as well as topological maps, completely annotated with semantic place categories. There are 100 sequences in total, collected by a mobile robot in three buildings at three different locations: Stockholm, SE (40 seqs), Freiburg, DE (26 seqs), and Saarbrucken, DE (34 seqs).

The **COLD-TopoMaps Dataset** consists of topological maps collected as the robot explores the environment. Each topological map is an undirected graph (also called ”topological graph”), where vertices are places that the robot could access, and edges indicate navigability. Each place is annotated by a semantic place category, such as corridor, kitchen, or doorway.
nodes.dat A CSV file where each row is the data for a certain node on the topological map. The column values (and types) are below:

Each place has 8 views (as shown in the figure below). For every edge connected to the node through a view, there are three entries of data in the row for this node:
If the neighbor node ID is -1 for a view number, then this view has no connected edge. If we think of a node as a disk, then the view numbers correspond to regions as annotated in the following illustration[11].

B. *New Dataset*

The dataset includes a comprehensive collection of photos illustrating various indoor spaces and rooms. A broad representation of the interior areas under examination is ensured by the different angles and views from which these photographs are captured. Essentially, they function as visual representations of the settings in which our mobile robots work, capturing the constantly changing arrangements, light, and item presence that define dynamic interior spaces.

Apart from the images, the collection also includes WiFi signal strength measurements taken from well-placed wireless access points in the target areas. The complex geographical distribution of WiFi signals is captured by this data, which is an invaluable asset for position determination. We obtain a thorough grasp of the signal strength variations by gathering WiFi measurements from various locations within each interior environment, which is an essential component in our system's capability to precisely locate a robot. **Pywifi** is the module we used. For dealing with wireless interfaces, it provides a cross-platform Python package. With the help of this module, we can get a list of WiFi networks that are available, and measure them, and save the results in a dataset.
In order to provide our system with ground truth and spatial reference, GPS coordinates are also captured at certain spots within the interior areas. In order to provide smooth transitions when robots navigate between indoor and outdoor regions, these coordinates serve as a crucial connection between outdoor and inside localization. In situations when GPS signals are weak or unstable, this integration is extremely important since it enables continuous navigation and precise location determination.

Furthermore, the dataset has been modified with object identification outcomes produced by the advanced deep learning algorithm YOLOv8. By correlating identified items to particular interior places, our results provide a layer of semantic knowledge to the dataset. By identifying things typical of certain locations, object detection improves accuracy and adaptability and provides semantic comprehension that helps the system identify and cope with interior environments.

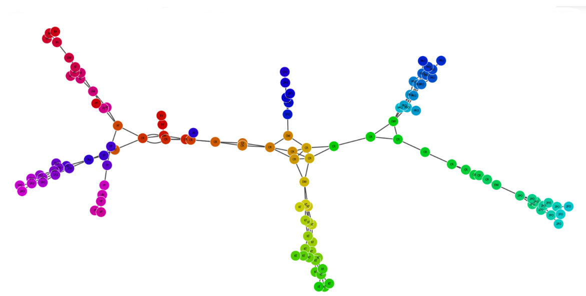

Fig. 3: **Single Floor Graph**

Node_ID	Placeholder	X(m)	Y(m)	label	Views
int	boolean	float	float	string	8 views

Fig. 4: **Data Attributes**

neighbor node ID	affordance	view number
int	float	int

Fig. 5: **3 Entries for Each view**

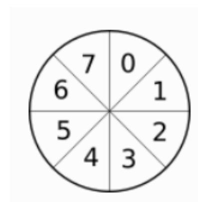

Fig. 6: **8 views**

Columns	DType
space	object
latitude	float64
longitude	float64
w01	float64
w02	float64
w03	float64
w04	float64
folderPath	object
spaceType	object
building	object

Fig. 7: Schema of Dataset

VI. Working Steps

A. Previous Method Working Steps

The project described above involves node classification in the first step of semantic mapping. Here are the general steps that we followed for previous method:

1) First, the COLD (Cognitive rObot Localization Database) dataset was used to launch the project. Working with this dataset proved challenging. Topological maps are acquired and saved in the COLD-TopoMaps Dataset while the robot explores its surroundings. Each map is an undirected graph, with vertices serving as potential access places for the robot and edges denoting navigability. Each place is annotated with a semantic place category.
2) We attempted to create GraphSPN while working with the COLD dataset, but as a result, we encountered several library dependency issues, which we later identified as deprecated versions of the same.
3) In the following phase, we choose to implement GNN. Graph Neural Networks (GNNs) are built to work with graph data, representing a network of connected nodes and edges.
4) An adjacency matrix and a feature matrix are used to represent a graph as the input data for the GCN node classification algorithm. While the feature matrix comprises feature vectors for each node in the graph, the adjacency matrix encodes the connections between the nodes in the network. subsequently turned the dataset into a feature matrix and an adjacency matrix for our model's input.
5) The first step in GCN is to perform a graph convolution operation. This operation involves multiplying the feature matrix with the adjacency matrix to capture the relationship between nodes in the graph.
6) After performing the graph convolution operation, an activation function is applied to introduce non-linearity to the model. The Rectified Linear Unit (ReLU) function is the most commonly used activation function.
7) The output of the final layer of GCN is then fed into a softmax function to obtain a probability distribution over the different classes of nodes.
8) The GCN model is trained using labeled data to minimize the cross-entropy loss between the predicted probability distribution and the true class labels.

B. Main Method Working Steps

Here are the general steps that we followed to create this project:

1) **Data Collection and Preprocessing:** We collected Images, Record GPS Signals Readings, and Measure the WiFi Signal Strengths of different Rooms. We annotated images with bounding boxes around objects of interest using the annotation tool Roboflow. Each bounding box should correspond to an object, and the annotation file should contain object labels and coordinates. We calculate the elevation(height) of the location from its cordinates of GPS and stored them into dataset. After that, We divide our dataset into training, validation, and testing sets.
2) **Model Training:** We used the preprocessed dataset and annotations to train YOLOv8 using the training script provided by the framework. After that, We monitored training progress, analyzed loss curves, and made adjustments to the learning rate and batch size as needed. We evaluate the trained model on the validation set and fine-tune the model parameters based on validation performance. Signal strength maps to each room using the recorded WiFi signal strength measurements and we use the recorded GPS signals to obtain coarse location estimates for each room. In this step, it can predict 'spaceType' from objects of locations. and GPS coordinates, WiFi readings, and space type help collectively predict actual places.
   a) **Logistic Regression for Indoor Localization** Logistic Regression is a statistical technique primarily employed for binary classification tasks[9]. But, it can be extended to handle multi-class classification tasks through a technique known as "One-vs-All" or "One-vs-Rest." This approach involves training multiple binary classifiers, each focused on distinguishing one class from the rest. In this project, We use this technique to predict the likelihood of each room's presence based on WiFi and GPS data.
      - **Training Multiple Classifiers(One-vs-All):** For each room in your dataset, train a separate Logistic Regression classifier. Each classifier learns to distinguish the signals and features specific to its assigned room from signals in all other rooms [6]. For each room, create a binary target variable 1 for the specific room and 0 for all other rooms combined and train separate Logistic Regression classifiers, each focusing on distinguishing a particular room.

b) **Naive Bayes Classification:** The probabilistic classifier Naive Bayes is straightforward and effective. It is used in our project to classify location kinds according to input parameters such as GPS locations, WiFi signal intensities, and object detection outcomes. Despite being simpler than other classifiers, Naive Bayes works effectively in multi-class classification applications. Its probabilistic methodology adds variation to the ensemble of classifiers and increases the project's adaptability by offering a simple and understandable way to categorize locations.

c) **Random Forest:** Another crucial classifier for our project is Random Forest. It is well-known for its ensemble learning methodology, which combines many decision trees to provide predictions as a group. Random Forest is a location identification system that can handle a broad range of input information, such as GPS coordinates, WiFi signal intensities, and object detection results, with high accuracy. Random Forest's ensemble approach lessens overfitting and increases the resilience of the system.

3) **Testing:** We evaluate Localization Accuracy in this phase. In this, We assess the accuracy of your indoor localization system by comparing the predicted room labels with ground truth. and also evaluate for an unseen location like if the location does not exist in the dataset then it can predict location type and neighborhood.
4) **Web App:** We also created a Web Application for this project to collect data automatically. This web Application stores images of GPS coordinates, calculate elevation, and takes readings of WiFi by its own. This web application also has a train and predict feature. So that users can easily use this thing without any difficulties.

VII. Results

In previous experiment, we compared the performance of two different models for node classification: a graph convolutional network (GCN) and an artificial neural network (ANN). We used the COLD-topo map dataset with 22 classes and trained both models to classify nodes into their respective classes.

After training and testing the models, we found that the GCN achieved an accuracy of **39.5%**, while the ANN achieved an accuracy of **30%**. This suggests that the GCN model was better suited to this particular task of node classification and was able to capture the relationships between nodes in the graph more effectively than the ANN.

One possible reason for the superior performance of the GCN is that it was able to leverage the structure of the graph to improve its predictions. The GCN uses the adjacency matrix of the graph to propagate information between neighboring nodes, allowing it to capture higher-order interactions between nodes. In contrast, the ANN treats each node as an independent data point and does not explicitly model the relationships between nodes. After that in new experiment, YOLOv8 performs well at recognizing things inside photos of indoor space type. The accuracy of 84.62% The accuracy for indoor localization achieved by using different models for WiFi and GPS signals is 80-90%. Currently, for this dataset, Logistic Regression has 76%, Naive Bayes has 85%, and Random Forest has 92%. This level of accuracy indicates the system's ability to forecast the proper room based on WiFi signal strength and GPS context with the help of object detection. The end result combines the results of YOLOv8 Object Recognition with multi-class classification of WiFi and GPS localization. This comprehensive strategy depends on the benefits of both modalities, improving the overall accuracy of indoor localization [13].

VIII. Conclusion

On the COLD dataset, which has 22 distinct labels, we examined the effectiveness of graph convolutional networks (GCNs) in this study to classify nodes. According to our findings, the GCN model classified the nodes into their respective labels with an accuracy of 39.5%.

These findings indicate that GCNs are a potential method for node classification tasks in challenging datasets like COLD, where typical machine learning models have difficulty capturing connections between nodes.

Our work underlines the need for more research to improve the performance of GCNs and solve their weaknesses while also demonstrating the promise of GCNs for node classification tasks in challenging datasets like COLD. The Project effectively integrates YOLOv8's object identification capabilities with Logistic Regression-based WiFi and GPS localization to obtain overall accuracy that improves indoor placement. The findings provide a spotlight on the potential and constraints of multimodal indoor localization. YOLOv8's accuracy of 84.62% demonstrates its effectiveness in recognizing things within indoor room pictures. Accurate object identification improves spatial comprehension and contributes to indoor localization precision. The combined accuracy of 76.92% achieved through WiFi and GPS localization using Logistic Regression highlights their importance in indoor locations. Despite WiFi signal loss and the limitations of indoor GPS signals, these Logistic Regression approaches give useful information.

IX. Future scope

The project has a lot of potential for future development. Here are some possible areas for improvement:

1) **Refining Model and Semantic Understanding**
   - **Fine-Tuning Object Detection:** We have investigated techniques to improve object detection accuracy, especially for objects currently being misidentified, and fine-tune the YOLOv8 model on misclassified objects and update bounding box annotations.
   - **Semantic Object Relationships:** We need to explore methods to establish semantic connections between detected objects and develop algorithms

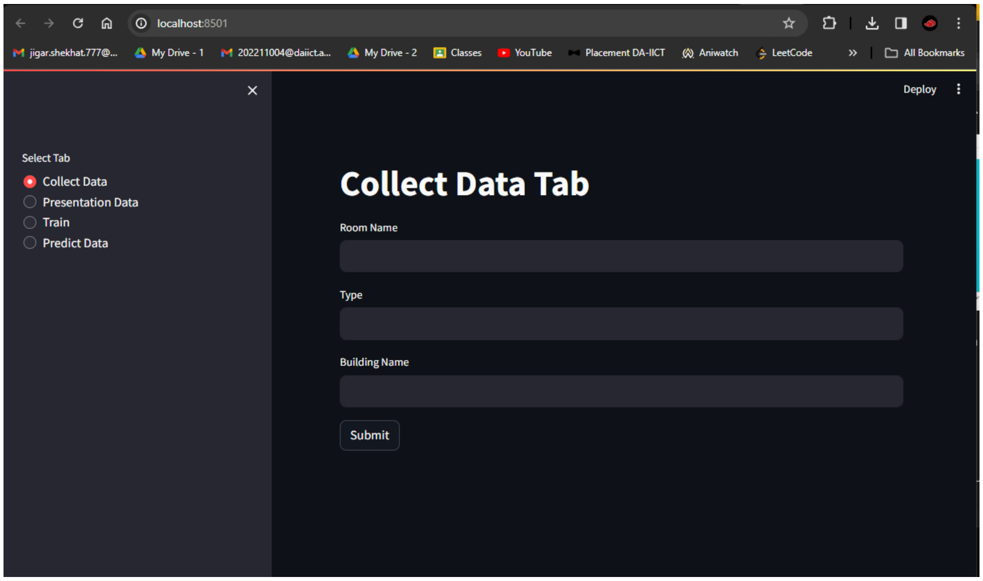

Fig. 8: Dashboard

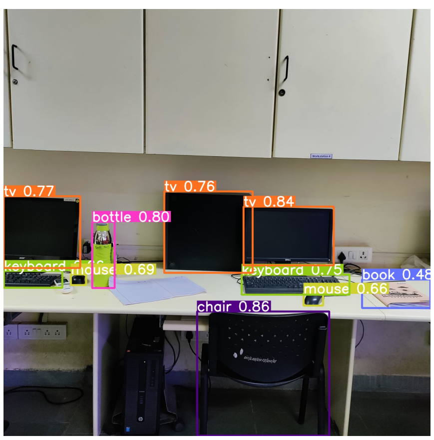

Fig. 9: Example of YOLOv8 Object Detection

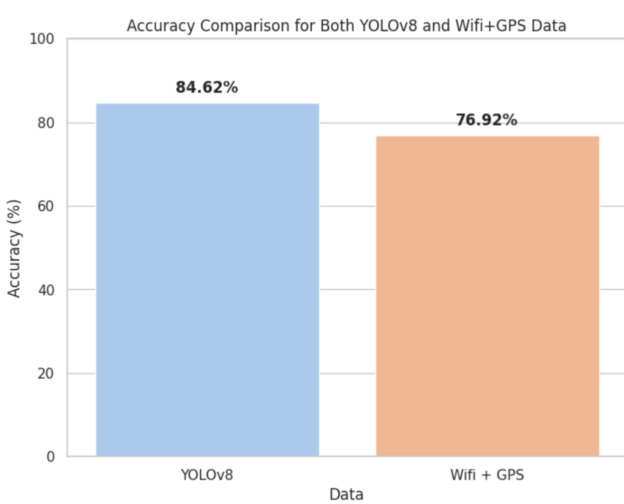

Fig. 10: Accuracy Comparison

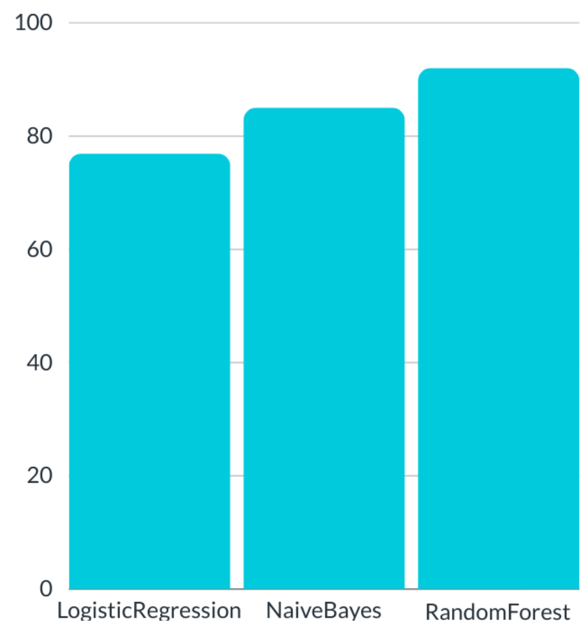

Fig. 11: Different model's Accuracy

to infer relationships, positions, or interactions between objects for a richer spatial understanding.

2) **WiFi Frequency Analysis and Contextual Insights** We need to analyze these frequencies to gain insights into areas with higher or lower WiFi activity. We have to develop techniques to create contextual WiFi signatures for rooms, combine signal strength with frequency count, and use these signatures to enhance the accuracy of WiFi-based localization.
3) **Handling Dynamic and Undefined Objects** We will investigate dynamic object tracking methods to account for objects that change position or appearance.

This could involve employing techniques like object-tracking algorithms and consider ways to track and localize these objects while maintaining overall accuracy.

References

[1] Armstrong Aboah, Bin Wang, Ulas Bagci, and Yaw Adu-Gyamfi. Real-time multi-class helmet violation detection using few-shot data sampling technique and yolov8. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5349–5357, 2023.

[2] Joydeep Biswas and Manuela Veloso. Wifi localization and navigation for autonomous indoor mobile robots. In *2010 IEEE international conference on robotics and automation*, pages 4379–4384. IEEE, 2010.

[3] Paul D Groves. Principles of gnss, inertial, and multisensor integrated navigation systems, [book review]. *IEEE Aerospace and Electronic Systems Magazine*, 30(2):26–27, 2015.

[4] Ismail Guvenc and Chia-Chin Chong. A survey on toa based wireless localization and nlos mitigation techniques. *IEEE Communications Surveys & Tutorials*, 11(3):107–124, 2009.

[5] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.

[6] Todd G Nick and Kathleen M Campbell. Logistic regression. *Topics in biostatistics*, pages 273–301, 2007.

[7] Neal Patwari, Joshua N Ash, Spyros Kyperountas, Alfred O Hero, Randolph L Moses, and Neiyer S Correal. Locating the nodes: cooperative localization in wireless sensor networks. *IEEE Signal processing magazine*, 22(4):54–69, 2005.

[8] Raquel Sanchez-Cauce, Iago París, and Francisco Javier Díez. Sum-product networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3821–3839, 2021.

[9] Sandro Sperandei. Understanding logistic regression analysis. *Biochemia medica*, 24(1):12–18, 2014.

[10] Juan Terven and Diana Cordova-Esparza. A comprehensive review of yolo: From yolov1 to yolov8 and beyond. *arXiv preprint arXiv:2304.00501*, 2023.

[11] Muhammad Muneeb Ullah, Andrzej Pronobis, Barbara Caputo, Jie Luo, and Patric Jensfelt. The cold database. Technical report, Idiap, 2007.

[12] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7464–7475, 2023.

[13] Yi Wu, Edward Y Chang, Kevin Chen-Chuan Chang, and John R Smith. Optimal multimodal fusion for multimedia data analysis. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 572–579, 2004.

[14] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

[15] Kenny Fong Peng Wye, Syed Muhammad Mamduh Syed Zakaria, Latifah Munirah Kamarudin, Ammar Zakaria, Norhawati Binti Ahmad, and Kamarulzaman Kamarudin. Rss-based fingerprinting localization with artificial neural network. In *Journal of Physics: Conference Series*, volume 1755, page 012033. IOP Publishing, 2021.

[16] Lei Xie, Yafeng Yin, Athanasios V Vasilakos, and Sanglu Lu. Managing rfid data: challenges, opportunities and solutions. *IEEE communications surveys & tutorials*, 16(3):1294–1311, 2014.

[17] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):1–23, 2019.

[18] Kaiyu Zheng and Andrzej Pronobis. From pixels to buildings: End-to-end probabilistic deep networks for large-scale semantic mapping. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3511–3518. IEEE, 2019.

[19] Kaiyu Zheng, Andrzej Pronobis, and Rajesh Rao. Learning graph-structured sum-product networks for probabilistic semantic maps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.