Reinforcement Learning for RFID Localization

Weitung Chen[1], Tara Boroushaki[1], Isaac Perper[1,2], Fadel Adib[1,2]
[1] *Massachusetts Institute of Technology,* [2] *Cartesian Systems*

***Abstract*—We present RL2, a robotic system for efficient and accurate localization of UHF RFID tags. In contrast to past robotic RFID localization systems, which have mostly focused on location accuracy, RL2 learns how to jointly optimize the *accuracy and speed* of localization. To do so, it introduces a reinforcement-learning-based (RL) trajectory optimization network that learns the next best trajectory for a robot-mounted reader antenna. Our algorithm encodes the aperture length and location confidence (using a synthetic-aperture-radar formulation) from multiple RFID tags into the state observations and uses them to learn the optimal trajectory. We built an end-to-end prototype of RL2 with an antenna moving on a ceiling-mounted 2D robotic track. We evaluated RL2 and demonstrated that with the median 3D localization accuracy of 0.55m, it locates multiple RFID tags 2.13x faster compared to a baseline strategy. Our results show the potential for RL-based RFID localization to enhance the efficiency of RFID inventory processes in areas spanning manufacturing, retail, and logistics.**

***Index Terms*—Reinforcement Learning, RFID Localization, Robotics, Autonomous Localization, RF sensing**

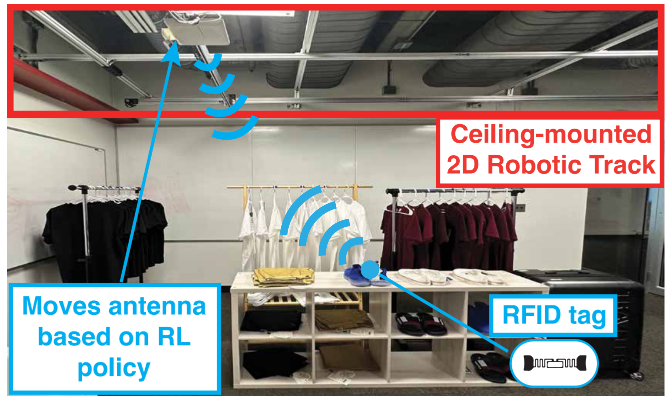


Fig. 1: **RL2 - SAR-based robotic RFID localization system using RL.** RL2 moves a ceiling-mounted antenna autonomously on a 2D track system. The antenna can move in x, y, and diagonal directions to perform Synthetic Aperture Radar (SAR) and localize RFID tags in the environment. It uses Reinforcement Learning to learn a strategy to localize multiple RFID tags in the environment more efficiently.

I. Introduction

RFID localization is an important problem with numerous applications across industries spanning retail, manufacturing, and warehousing. Many businesses have adopted Ultra High Frequency (UHF) RFID tags to manage their product inventory and assets. State-of-the-art research in RFID localization has demonstrated the potential of accurate (decimeter-level) localization in practice indoor environments, including those with dense multipath reflections from furniture and building interiors [3]–[6], [8], [10], [11], [14], [15]. One of the most promising approaches for RFID localization is to leverage robotic systems. In these systems, the RFID reader antennas are mounted on a robot. As the robot moves, it emulates a large number of antennas and uses the collected measurements to accurately localize the UHF RFID tags in the environments via sophisticated array processing methods, such as the Synthetic Aperture Radar (SAR) [14], [15].

Unfortunately, the majority of past proposals for robotic RFID localization systems have focused on the accuracy of localizing individual tags, paying little attention to localization efficiency, especially in scenarios where hundreds or thousands of tags are present in the environment. For example, systems such as PinIt [15] and MobiTagbot [14] localize UHF RFID tags by moving the robot along a line and collecting a large number of measurements to estimate the tags' locations. However, this approach is prone to inefficiencies and localization errors. To see why, consider a robot that opportunistically collects RFID measurements as it traverses a predefined linear trajectory. Because of the stochastic nature of the reader protocol and multipath interference, different tags may have significantly different array aperture sizes, leading to some tags being localized with significantly lower accuracy than others (e.g., multi-meter). One potential solution is to repeat the entire scan along the same path until all tags have been localized within a target confidence interval. However, such an approach can lead to significant inefficiency as the robot repeatedly traverses the same large trajectories. Ideally, we would like to design a robot that can optimally scan the environment to both localize all tags and minimize the trajectory length, enabling a more efficient localization.

In this paper, we explore the possibility of adapting recent advancements in (Reinforcement Learning (RL)) to enable efficient and accurate localization of RFID tags. We design a trajectory optimization algorithm that leverages the Deep RL framework. Our algorithm aims to optimize the antenna scanning path by utilizing feedback from RFID localization of multiple tags within the environment. Our aim is to develop a robotic system capable of optimally selecting a SAR trajectory to collect RF measurements and minimize the total scanning distance required to localize all the target tags in the environment. The challenge of achieving this goal lies in integrating multi-tag RFID localization RF information into the RL framework. This involves encoding the tag localization and confidence information for multiple tags and investigating how this information can be used to guide an RL agent to learn the optimal trajectory for efficiently localizing all the tags.

To address the challenge, we developed RL2, the first SAR-based robotic RFID localization system that incorporates RL to optimize the SAR scanning trajectories, significantly boosting the efficiency of locating multiple tags in the environment. RL2 operates as a robotic RFID localization system with an antenna moving on a ceiling-mounted 2D track as shown in Fig. 1. The core of this system is a RL framework that utilizes multi-tag RFID localization feedback and identifies the next best SAR trajectory for localizing multiple tags. The paper presents the following contributions:

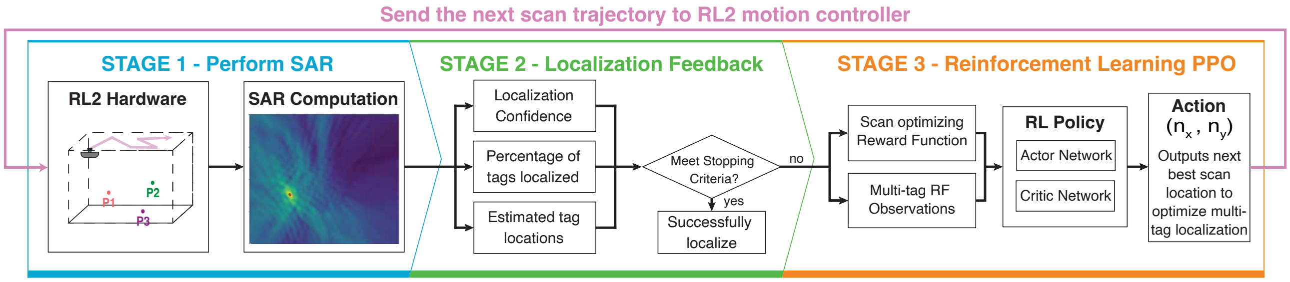


Fig. 2: **RL2 System Overview.** RL2 includes three stages: (1) Control robotic track system to move the antenna and perform SAR calculations for multiple tags (2) Compute the localization confidence, estimated locations, and the percentage of tags converged (3) Feed the multi-tag RF information into RL module and output the next best scanning location.

- It presents the first system to incorporate RL for optimizing the trajectory of robotic scanning for RFID SAR localization on multiple RFID tags in the environment.
- It introduces an RL-based SAR trajectory optimization algorithm. This algorithm encodes information such as the localization confidence of each of the tags, percentage of tags localized, and the antenna scanning distance to learn a policy with Proximal Policy Optimization (PPO) [13]. This combination of RL framework and RF information enables a more efficient localization of multiple tags by optimizing the SAR scanning trajectory.
- It presents an end-to-end prototype implementation and evaluation of RL2. The system is implemented using a moving antenna on a ceiling-mounted 2D track, achieving end-to-end autonomous SAR-based RFID localization. The system has been evaluated in over 100 real-world experiments, each involving 50 to 64 tags, and compared against three other scanning strategies. The evaluation demonstrates that RL2 localizes the RFID tags 2.13x faster compared to the baseline strategy, with a median 3D localization error of 0.55m.

II. Related Work

Early work in RFID localization relied on methods such as measuring received signal strength (RSS) [7], [12], angle of arrival (AoA) [2], [9] the received signal phase [1]. While these approaches demonstrated good accuracy in large open spaces, the struggled to maintain fine-grained (decimeter-level) accuracy in practical multipath-dense indoor environments. To address these challenges, later research adopted more sophisticated techniques such as SAR to achieve high spatial resolution and accurate localization by emulating a large number of antennas. Various systems achieve SAR RFID localization by mounting antennas on a robot to read RFID tags along a trajectory using Roombas [14], robotic arms [11], drones [6], or track systems [8].

However, these past proposals for RFID localization mainly focused on enhancing localization precision but not its speed. Recognizing this limitation, recent proposals have investigated new approaches to improve RFID localization efficiency via path optimization strategies. For example, RF-AR [3] introduces a path optimization algorithm based on Dilution of Precision (DoP) and guides users to search for tags more efficiently. Similarly, RFusion [5] employs a robotic arm and uses an RF-visual RL algorithm to iteratively find the optimal position for tag localization. However, both RF-AR and RFusion are designed mainly for optimizing the search trajectory for a *single RFID tag*. In the presence of multiple tags, the entire process must be repeated, resulting in significantly longer trajectories and latencies (as we demonstrate empirically in our results). RL2 is inspired by these recent advances and takes a step forward to incorporate RL in trajectory optimization of multiple tags, delivering significant efficiency gains.

III. Efficient SAR-based RFID Localization with RL

In this section, we first describe how RL2 uses SAR to perfrom 3D RFID localization with a robotic track system. We then describe how RL2 leverages RL to optimize the trajectory of the robotic track system for simultaneous efficient localization of multiple tags.

This system is implemented on a ceiling-mounted 2D robotic track system with an RFID antenna installed on the movement platform, as shown in Fig. 1. This robotic track system can move the antenna in any direction across the 2D plane on the ceiling. RL2 operates in three stages, as shown in Fig. 2, namely: 1) The robotic track system moves the antenna while collecting RF measurements and calculates SAR. 2) Based on the SAR calculation, RL2 derives the estimated location and its localization confidence for each tag. 3) The encoded observation is fed to the RL model to choose the best next scanning trajectories. This process is repeated until the stopping criteria is met. We describe these steps in more details in the rest of this section.

A. SAR-based RFID Localization

To locate RFIDs through SAR, the robotic track system moves an antenna as it collects measurements, emulating an antenna array. Formally, we define the list of collected RF measurement Λ_s as follows:

$$\Lambda = [..., (epc_n, rssi_n, \phi_n, f_n, x_n, y_n, t_n)] \tag{1}$$

where epc_n is the RFID tag electronic product code (EPC) identifier, $rssi_n$ is the Received Signal Strength Indicator (RSSI), ϕ_n is the measured phase value, f_n is the frequency of the received signal, (x_n, y_n) is the coordinate at which the measurement is collected from, and t_n is the timestamp

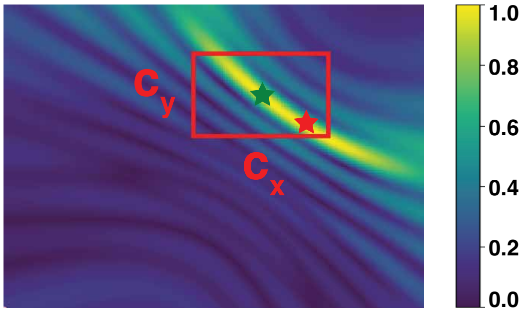


Fig. 3: **SAR Computation and Confidence Interval** This heatmap shows the SAR output of a tag localization result. We use the peak estimated power to obtain the location estimate of the tag. The green star shows the estimated location while the red star shows the ground truth location of the tag. The confidence interval of the location estimate is shown as the red rectangle, defined by c_x and c_y.

information. The power emanated by tag with epc_n from each point can be estimated through standard SAR formulation as follows [3]:

$$P_{epc_n}(x,y,z) = \left\| \frac{1}{N} \sum_{i=1}^{N} h_i e^{j\frac{2\pi d_i(x,y,z)}{\lambda}} \right\| \tag{2}$$

where N is the total number of measurements from the tag with epc_n, h_i is the channel estimation from the i^{th} measurement from tag with epc_n, d_i is the round trip distance from (x,y,z) to the i^{th} measurement location, and λ is the wavelength of the received signal. The tag with epc_n can be located at the peak value of SAR result, $\hat{p}_{epc_n}$:

$$\hat{p}_{epc_n} = \texttt{argmax}_{(x,y,z)}(P_{epc_n}(x,y,z)) \tag{3}$$

B. Localization Confidence

After obtaining a location estimate, $\hat{p}_{epc_n}$, we aim to understand the confidence level of this localization result. We define the confidence interval of the localization result as a region encompassing all (x,y,z) locations where the estimated power $P_{epc_n}(x,y,z)$ is within 0.75 dB of the peak estimated power. The confidence interval includes three numbers: (c_x, c_y, c_z), and they describe the maximum span of the three axes to encompass all locations within 0.75 dB of the identified peak. Larger confidence interval means larger variance in the localization results, indicating a lower confidence in localization results. Figure 3 shows an example of the SAR result with the confidence interval c_x and c_y.

To declare that a tag localization process has converged, RL2 need to satisfy two criteria: 1) The confidence interval size should be smaller than threshold[1]

$$c_x < \sigma_x \;\;\&\;\; c_y < \sigma_y \;\;\&\;\; c_z < \sigma_z$$

2) The changes in confidence intervals should be below threshold[2]

$$\Delta c_x < \epsilon \;\;\&\;\; \Delta c_y < \epsilon \;\;\&\;\; \Delta c_z < \epsilon$$

C. RL-based Trajectory Optimization

RL2 uses RL to develop a trajectory optimization algorithm for efficiently localizing a large number of RFID tags. The objective of the algorithm is to find the next best location for the antenna to move to improve the quality of localization results for many tags . The RL problem's state observations, action space, and reward function can be formulated as follows:

1) State Observation: We encode the observation as

$$S = \{(x^i_{est},\ y^i_{est},\ c^i_x,\ c^i_y)^N_{i=1},\ l_r,\ x_a,\ y_a\} \tag{4}$$

where (x^i_{est}, y^i_{est}) represents the current estimated location of tag i^{th}, (c^i_x, c^i_y) indicates the confidence interval for tag i^{th} [3], l_r represent the percentage of the tags that have been localized, and (x_a, y_a) is current antenna position. Consequently, our state space has a size of $4N+3$.

2) Action Space: Each action can be represented as $a = (x, y)$ where x and y are the next coordinates that the robotic track should move the antenna to.

3) Reward Function: The reward function is dependent on the total antenna scanning distance, confidence intervals, and the localization convergence percentage. Formally, the reward function is defined as follow [4]:

$$R = \alpha \cdot s_{total} + \beta \cdot (C_t - C_0) + \gamma \cdot L_t + B_p + R_c \tag{5}$$

where s_{total} is the total scanning distance, C_t is the average L2-norm confidence interval of all K tags after the last scan in the environment, i.e. $\frac{1}{K}\sum_{i=1}^{K}\sqrt{(c^i_x)^2 + (c^i_y)^2 + (c^i_z)^2}$, C_0 is the initial confidence interval and can be defined as $\sqrt{(x_{max}-x_{min})^2 + (y_{max}-y_{min})^2 + (z_{max}-z_{min})^2}$, L_t is the percentage of tags that meet the convergence criteria, B_p is the scanning step length penalty, and R_c is the task completion reward. B_p is a term that penalizes very short or very long scans. The intuition is that we want to penalize the agent for moving to a location that is too close to the previous location, as it does not provide new information. We also want to penalize the agent for moving to a location that is way too far from the previous location than it needs. B_p is defined as $B_p = \sum_s b^s_p$ where

$$b^s_p = \begin{cases} G\cdot(1-l^s_r) & \Delta a_s < 0.20 \\ G\cdot(1-l^s_r) & \Delta a_s > \min(x_{max}-x_{min}, y_{max}-y_{min}) \\ 0 & o.w. \end{cases} \tag{6}$$

where $\Delta a_s = ||a_s - a_{s-1}||$ is the length of the antenna motion at step s, and l^s_r is the percentage of tags that have been localized until step s [5].

The agent is also awarded a task completion reward of $R_c = 50$ if it localizes η[6] percent of all the tags, and is penalized $R_c = -50$ if it doesn't localize and reaches the maximum allowed scanning distance [7].

D. RL Architecture and Training Details

We trained the deep RL network in simulation based on the state observation, action, and reward function defined above.

[1] In our implementation $\sigma_x = \sigma_y = 0.3$ and $\sigma_z = 0.6$

[2] RL2 selects ϵ = 0.03

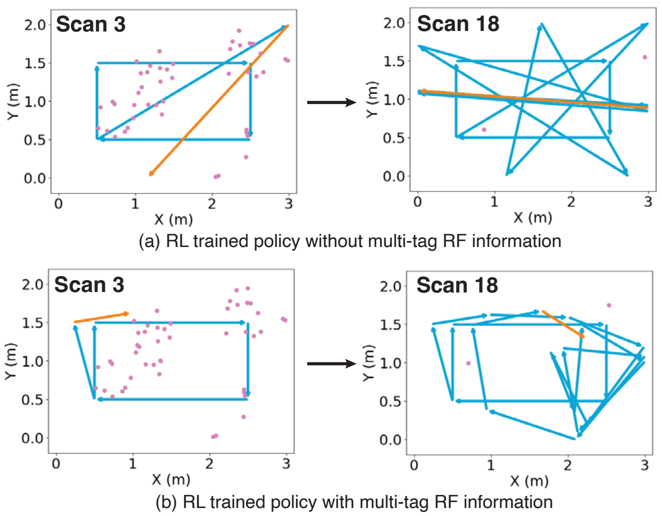
[3] If the total number of tags is greater than N, we select the closest N tags to the current antenna location to iteratively optimize a group of more than N tags

[4] We set $\alpha = -0.9$, $\beta = -5.0$, $\gamma = 5.0$

[5] In out implementation, we set $G = -20$

[6] We set η to 98

[7] The maximum allowed scanning distance is set to 50m in RL training



(a) RL trained policy without multi-tag RF information

(b) RL trained policy with multi-tag RF information

Fig. 4: **Example of RL2 SAR Trajectory Optimization.** We compare RL2's scanning trajectory of (a) RL-trained policy without incorporating multi-tag RF information and (b) RL policy of RL2, which includes multi-tag RF information at scan 3 and 18. Blue arrows represent past scanning trajectories, while the orange arrow indicates the latest antenna scan at that step. Pink dots denote the locations of RFID tags that have not yet been localized.

The RL architecture mainly consists of an actor critic policy that outputs the next best scanning trajectory. The actor and critic network incorporate Multilayer Perceptron (MLP) layers, and they both have 2 hidden layers with 128 neurons each. The output layer has 2 neurons for the x and y coordinates. We apply the state-of-the-art deep RL algorithm - PPO [13] to learn the optimal policy and minimize scanning trajectory length needed to localize all the tags. PPO is a policy gradient method that incorporates a clipping function to achieve a more stable process of policy optimization. It calculates the reward value based on interaction with the environment and computes an advantage estimate using the current value function. With these values, the PPO algorithm updates the network parameters to improve the policy. Consequently, the trained policy is then used to suggest the next best location for the antenna to move to.

Figure 4 illustrates an example of utilizing RL for SAR localization trajectory optimization in simulation. To show the advantage of our observation encoding, we compare two different path optimization strategies using an RL model trained with (a) RL agent with states encoded as $S = \{l_r,\ x_a,\ y_a\}$ instead of Eq.4 and with completion reward of $R_c = 50$ instead of what was described in Eq.5, and (b) RL agent with states encoded with multi-tag RF information as formulated as described in Eq.4 and Eq.5 in the previous section. The total scanning trajectories (blue and orange arrows) and the RFID tags yet to be localized (pink dots) are shown for the 3rd and the 18th scan. The tags' locations are set to have the same distribution in both scenarios. Note that we define an initial trajectory τ_{init} to collect the first set of measurements[8]. After this initial scan, the RL agent takes over and suggests the next best location for the antenna to move to.

[8]We define τ_{init} to be a rectangular scan trajectory with setpoints $[(0.5, 0.5)$, $(0.5, 1.5)$, $(2.5, 1.5)$, $(2.5, 0.5)$, $(0.5, 0.5)]$

Figure 4a reveals that the RL agent has learned to scan the environment in a star-shaped pattern to maximize aperture at different angles. As a result, the system manages to localize 48 out of 50 tags with a scanning length of 52.42m. In contrast, Figure 4b displays the trajectories when using RL2's path optimization algorithm, indicating that the inclusion of multi-tag RF information enables the algorithm to scan more efficiently in clusters potentially containing RFID tags. After 18 scans, the system can also localize 48 out of 50 tags, but it traveled only 19.64m of scanning distance. This demonstrates the effectiveness of integrating multi-tag RF information into the RL training and RL2's ability to localize multiple RFID tags more efficiently.

Sim-to-Real: As the RL agent learns the policy in a simulation environment, we need to consider the generalization of the policy network when it is implemented on the real hardware. RL2 apply domain randomization technique to close the gap between the simulation and the real-world environment. We trained our RL agent with randomized system parameters such as the noise range of the measured phase value and the tag readability parameters. This approach helps generalize the RL network to have similar performance on the real-world system.

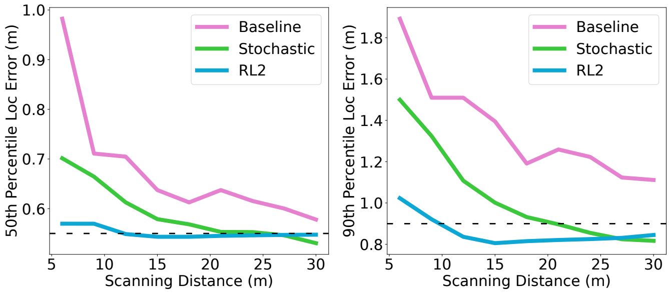
IV. Implementation

Physical Prototype: Our setup comprises of a 2D ceiling-mounted track with a circularly polarized antenna, a motor controller, a Thingmagic M6e UHF RFID reader, and a local server for computations. The track system is constructed with V-slot linear rails and can move the antenna to any location on the X/Y plane, including in x, y, and diagonal directions. It is powered by NEMA 23 stepper motors and controlled by a Protoneer Raspberry Pi CNC board running GRBL software. We interfaced the RFID reader through the Mercury API, which operates within the ISM band. All the RF measurement data are processed on a local server, a Mac Mini 2018 with a 3.0 GHz 6-core Intel Core i5 processor, running Ubuntu 20.04.

Simulator for RL Training: We developed a simulator for the RL agent to perform SAR-based RFID localization and the PPO algorithm in Python. The simulator incorporates random phase noises, multipath effects, and a tag read parameter to model the tag readability based on the distance between the reader and the tag. The RL network was trained in the simulator more than 750,000 interactions. The training process took place on a machine equipped with 8 GTX 1080 Ti GPUs, an 8-core Intel Xeon CPU @ 2.10 GHz, and 256GB of RAM.

V. Evaluation

We evaluated RL2 in a multipath rich indoor environment designed to mimic a storage room setting. Fig 1 shows the evaluation environment for RL2. We selected clothes and shoes as the items we wanted to localize and attached standard UHF RFID tags to them. The shelves, racks, and sofa in the area were randomly placed within the target localization area of RL2, and the clothes and shoes were likewise randomly positioned on these furnitures in the area. Note that some of the tags



(a) 50th Percentile

(b) 90th Percentile

Fig. 5: **Localization Error vs Scanning Distance.** The plots evaluate the efficiency and accuracy of RL2 and other baseline strategies. We present the (a) 50th and (b) 90th percentile localization error at different scanning distances. The dashed lines mark the localization error benchmark value at (a) 0.55m and (b) 0.9m. These benchmarks help compare how each strategy performs in terms of scanning distance needed to achieve these errors.

were in non-line-of-sight situations due to obstructions like shelves or other items. We evaluated RL2 with a baseline and stochastic strategy to assess its performance on minimizing the total scanning distance for multi-tag RFID localization.

- *Baseline*: This baseline strategy exhaustively scans the environment and starts with a rectangular scan at the border of the environment and continue to spiral inwards, reducing the rectangular scan by 0.1m length each round.[9]
- *Stochastic*: We developed this stochastic policy to generate the next scanning location for the antenna to go to based on a random distribution (uniform distribution within the bounds of the scanning area).
- *RL2*: Reinforcement learning based trajectory optimization algorithm explained in Section §III.

Metrics: We evaluated RL2 performance through three main metrics: *1) Localization error* is the error between the target RFID tag ground truth location and the location estimate computed by RL2. *2) Scanning Distance* is the distance traveled by the antenna from the beginning until the RFID tagged item is confidently located. *3) L2-Norm Confidence Interval* refers to the L2 norm of the confidence interval for the RFID tag's location estimate, as defined in Section §III-B. Formally, it is represented by $||(c_x, c_y, c_z)||$.

Ground Truth: We collected the target RFID tags' EPC ids and measured their ground truth locations relative to the origin of the environment using a laser distance meter.

VI. Results

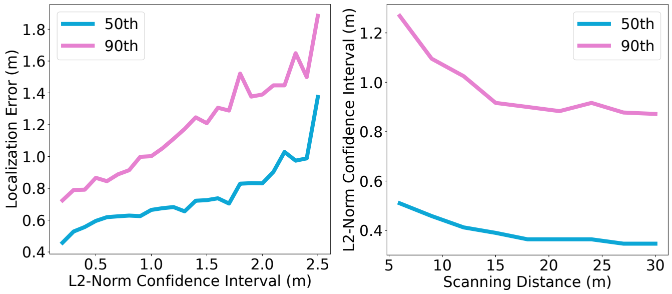
We conducted a total of 100 real-world experimental trials based on two different setups: 60 trials with 51 RFID tags and 40 trials with 64 RFID tags. The goal was to evaluate the performance of RL2's RL-based trajectory optimization algorithm, focusing on the efficiency and accuracy of RFID localization. The localization results were collected at each timestep of the SAR scanning process. After each scan in an experiment, we calculated the localization error and the L2-norm confidence interval for each tag from the SAR computation, and recorded RL2's current cumulative SAR scan length. Based on these results, we compiled a collection of (*localization error, scanning distance, L2-norm confidence interval*) data points. These data points were gathered across all tags at each scan of all trials, allowing us to understand fully the performance of RL2 and compare with other baseline strategies. The results are presented in Figures 5 and 6, with the findings discussed below.

A. Localization Error vs Scanning Distance

We evaluated RL2 by comparing the localization error of RFID tags with the SAR scanning distances to assess accuracy versus efficiency across three scanning strategies. We took the (*localization error, scanning distance*) pair from our data collection, categorizing them into 3m bins along the X-axis based on *scanning distance*. For each bin, we computed the median and 90th percentile errors for each strategy, as shown in Figures 5a and 5b. We show the *scanning distance* from 6m to 30m and focus on the change in localization error within this range of scanning distances. The pink line shows the exhaustive strategy, green for stochastic, and blue for RL2, with dashed lines at 0.55m and 0.9m as error benchmarks. These benchmarks help compare how each strategy performs in terms of scanning distance needed to achieve these errors. From this, we observe the following:

- RL2 requires only 9.77m of scanning distance to reach a 0.55m median localization error, while the stochastic strategy needs 20.83m for the same error, marking a 2.13x efficiency improvement. This shows that RL-based trajectory optimization makes RL2 more effective at localizing RFID tags than the random scanning strategy. One practical constraint is that commercial RFID reader decoding hardware typically suffers from pi phase ambiguity. This ambiguity must be resolved prior to applying the antenna array equations. To deal with this issue, one approach is to perform phase unwrapping, but that requires collecting dense measurements for each RFID (e.g., within $\frac{\lambda}{8}$) along the robot's trajectory. We achieve this by moving the robot at a relatively slow speed (around 16.67mm/sec). Thus, it takes 586 seconds to traverse a 9.77m trajectory. Note that specialized reader hardware that does not suffer from phase ambiguity would be capable of moving at much faster speeds.
- RL2 outperforms both the stochastic and baseline strategies at the 90th percentile. RL2 requires only 11.44m of antenna scanning distance to achieve a 90th percentile localization error of 0.9m, while the stochastic strategy requires 26.29m of scanning distance. This marks a 2.30x improvement in scanning efficiency and underscores the effectiveness of the RL-based trajectory optimization algorithm.
- Among the three strategies, RL2 demonstrates the lowest 50th and 90th percentile localization errors at all scanning distances. The baseline scanning strategy shows the highest localization error and did not reach the reference localization error values set for both the median and 90th percentile. This indicates that the baseline strategy is inefficient at quickly

[9]Doing so emulates what past approaches [5], [14], [15] would need to localize all tags by repeating a rectangular scan but with small variations.



(a) Localization Error vs Conf Interval (b) Conf Interval vs Scan Distance

Fig. 6: **RL2 Microbenchmark.** This microbenchmark shows the relationship between the confidence intervals of the tags, antenna scanning length, as well as the localization accuracy. We present the (a) localization error for various L2-norm confidence interval values and (b) L2-norm confidence interval of the RFID tags at different scanning distances. We plot the 50th percentile of the data in blue and the 90th percentile in pink.

locating multiple tags and does not efficiently localize multiple RFID tags in the environment compared to RL2.

B. RL2 Microbenchmark

We conducted a microbenchmark to evaluate the performance of RL2 and to show the relationship among the tags' confidence intervals, localization accuracy, and antenna scanning distance. In Figure 6a, the 50th (blue) and 90th (pink) percentiles of localization error for each L2-norm confidence interval are shown. We divided the X-axis into 0.1m intervals, assigning data points based on their L2-norm values, and calculated the 50th and 90th percentile errors in each bin, ranging from 0.2m to 2.5m. Similarly, Figure 6b presents the 50th and 90th percentiles for RFID tags at various distances. Here, distances are grouped into 3m intervals from 6m to 30m, with percentile calculations for each bin. Based on these results, we observe the following:

- RL2 demonstrates a trend where the localization error increases as the confidence interval increases. The 50th percentile localization error for RL2 is 0.46m at a confidence interval of 0.2m and 1.37m at a confidence interval of 2.5m. Similarly, the 90th percentile localization error follows the same trend, with 0.72m at a confidence interval of 0.2m and 1.88m at a confidence interval of 2.5m. This indicates that the confidence interval of a location estimate can be a reliable predictor of localization accuracy. RL2 leverages this information to optimize the scanning trajectory with the RL-based algorithm and achieves efficient and accurate localization of multiple RFID tags.

- As we perform more scans with RL2, the L2-norm confidence interval for the RFID tags decreases. The 50th percentile of the L2-norm confidence interval for the RFID tags is 0.5m at a scanning distance of 6m and 0.34m at a scanning distance of 30m. The 90th percentile of the L2-norm confidence interval for the RFID tags is 1.27m at a scanning distance of 6m and 0.87m at a scanning distance of 30m. This demonstrates that RL2 can conduct more scans to localize the tags with a lower confidence interval, i.e., higher confidence and lower localization error.

VII. Conclusions

RL2 introduces reinforcement learning into the fine-grained RFID localization problem, demonstrating significant benefits in improving overall localization efficiency. Our reinforcement learning network encodes RF-based metrics, derived from the SAR-based formulation and real-time measurements, to learn optimized robot scanning trajectories. We designed and built a robotic track system that incorporates the learnt RL scanning algorithm. Our evaluations demonstrate the capability of the RL-based algorithm to enhance the efficiency of localizing multiple RFID-tagged items simultaneously. As this research evolves, it would be valuable to evaluate RL2 as a function of different tag densities and multipath environments. It would also be interesting to explore how advancing the localization algorithm with RF-visual sensor fusion (similar to [5]) and with more sophisticated RFID estimation techniques (similar to [10]) can further improve efficiency.

References

[1] Daniel Arnitz, Klaus Witrisal, and Ulrich Muehlmann. Multifrequency continuous-wave radar approach to ranging in passive uhf rfid. *Microwave Theory and Techniques, IEEE Transactions on*, 57:1398 – 1405, 06 2009.

[2] Salah Azzouzi, Markus Cremer, Uwe Dettmar, Rainer Kronberger, and Thomas Knie. New measurement results for the localization of uhf rfid transponders using an angle of arrival (aoa) approach. pages 91 – 97, 05 2011.

[3] Tara Boroushaki, Maisy Lam, Weitung Chen, Laura Dodds, Aline Eid, and Fadel Adib. Exploiting synergies between augmented reality and rfids for item localization and retrieval. In *IEEE RFID 2023*, 2023.

[4] Tara Boroushaki, Maisy Lam, Laura Dodds, Aline Eid, and Fadel Adib. Augmenting augmented reality with non-line-of-sight perception. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*, 2023.

[5] Tara Boroushaki, Isaac Perper, Mergen Nachin, Alberto Rodriguez, and Fadel Adib. Rfusion: Robotic grasping via rf-visual sensing and learning. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 192–205, 2021.

[6] Alice Buffi, Andrea Motroni, Paolo Nepa, Bernardo Tellini, and Riccardo Cioni. A sar-based measurement method for passive-tag positioning with a flying uhf-rfid reader. *IEEE Transactions on Instrumentation and Measurement*, 68:845–853, 2019.

[7] Kirti Chawla, Christopher McFarland, Gabriel Robins, and Connor Shope. Real-time rfid localization using rss. In *2013 International Conference on Localization and GNSS (ICL-GNSS)*, pages 1–6, 2013.

[8] Yue Jiang, Yongtao Ma, Hankai Liu, and Yunlei Zhang. Rf-sml: A sar-based multi-granular and real-time localization method for rfid tags. *Electronics*, 2020.

[9] Rainer Kronberger, Thomas Knie, Roberto Leonardi, Uwe Dettmar, Markus Cremer, and Salah Azzouzi. Uhf rfid localization system based on a phased array antenna. *2011 IEEE International Symposium on Antennas and Propagation (APSURSI)*, pages 525–528, 2011.

[10] Yunfei Ma, Nicholas Stearns Selby, and Fadel M. Adib. Minding the billions: Ultra-wideband localization for deployed rfid tags. *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017.

[11] Robert Miesen, Fabian Kirsch, and Martin Vossiek. Uhf rfid localization based on synthetic apertures. *IEEE Transactions on Automation Science and Engineering*, 10:807–815, 2013.

[12] Lionel Ni, Yunhao Liu, Yiu Lau, and Abhishek Patil. Landmarc: Indoor location sensing using active rfid. volume 10, pages 407– 415, 04 2003.

[13] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. 07 2017.

[14] Longfei Shangguan and Kyle Jamieson. The design and implementation of a mobile rfid tag sorting robot. MobiSys '16, page 31–42, New York, NY, USA, 2016. Association for Computing Machinery.

[15] Jue Wang and Dina Katabi. Dude, where's my card? rfid positioning that works with multipath and non-line of sight. *ACM SIGCOMM Computer Communication Review*, 43:51–62, 08 2013.