



# RLoc: Towards Robust Indoor Localization by Quantifying Uncertainty

TIANYU ZHANG, University of Science and Technology of China, China and Hefei Comprehensive National Science Center, China

DONGHENG ZHANG, GUANZHONG WANG, and YADONG LI, University of Science and Technology of China, China

YANG HU, University of Science and Technology of China, China

QIBIN SUN and YAN CHEN\*, University of Science and Technology of China, China

In recent years, decimeter-level accuracy in WiFi indoor localization has become attainable within controlled environments. However, existing methods encounter challenges in maintaining robustness in more complex indoor environments: angle-based methods are compromised by the significant localization errors due to unreliable Angle of Arrival (AoA) estimations, and fingerprint-based methods suffer from performance degradation due to environmental changes. In this paper, we propose RLoc, a learning-based system designed for reliable localization and tracking. The key design principle of RLoc lies in quantifying the uncertainty level arises in the AoA estimation task and then exploiting the uncertainty to enhance the reliability of localization and tracking. To this end, RLoc first manually extracts the underutilized beamwidth feature via signal processing techniques. Then, it integrates the uncertainty quantification into neural network design through Kullback-Leibler (KL) divergence loss and ensemble techniques. Finally, these quantified uncertainties guide RLoc to optimally leverage the diversity of Access Points (APs) and the temporal continuous information of AoAs. Our experiments, evaluating on two datasets gathered from commercial off-the-shelf WiFi devices, demonstrate that RLoc surpasses state-of-the-art approaches by an average of 36.27% in in-domain scenarios and 20.40% in cross-domain scenarios.

CCS Concepts: • Human-centered computing → *Ubiquitous and mobile computing systems and tools*.

Additional Key Words and Phrases: Indoor Localization; WiFi; Uncertainty Learning; AoA;

## ACM Reference Format:

Tianyu Zhang, Dongheng Zhang, Guanzhong Wang, Yadong Li, Yang Hu, Qibin sun, and Yan Chen. 2023. RLoc: Towards Robust Indoor Localization by Quantifying Uncertainty. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 200 (December 2023), 28 pages. <https://doi.org/10.1145/3631437>

## 1 INTRODUCTION

Numerous real-world applications [56], such as indoor navigation in large spaces and device detection in smart homes, require decimeter-level localization that is robust to furniture, human bodies, and other occlusions in the

\*This is the corresponding author.

---

Authors' addresses: Tianyu Zhang, tianyuzhang@mail.ustc.edu.cn, School of Cyber Science and Technology, University of Science and Technology of China, Hefei, China, Institute of Artificial Intelligence and Hefei Comprehensive National Science Center, Hefei, China; Dongheng Zhang, dongheng@ustc.edu.cn; Guanzhong Wang, wangguanz@mail.ustc.edu.cn; Yadong Li, yadongli@mail.ustc.edu.cn, School of Cyber Science and Technology, University of Science and Technology of China, Hefei, China; Yang Hu, eeyhu@ustc.edu.cn, School of Information Science and Technology, University of Science and Technology of China, Hefei, China; Qibin sun, qibinsun@ustc.edu.cn; Yan Chen, eecyan@ustc.edu.cn, School of Cyber Science and Technology, University of Science and Technology of China, Hefei, China.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2474-9567/2023/12-ART200 \$15.00

<https://doi.org/10.1145/3631437>

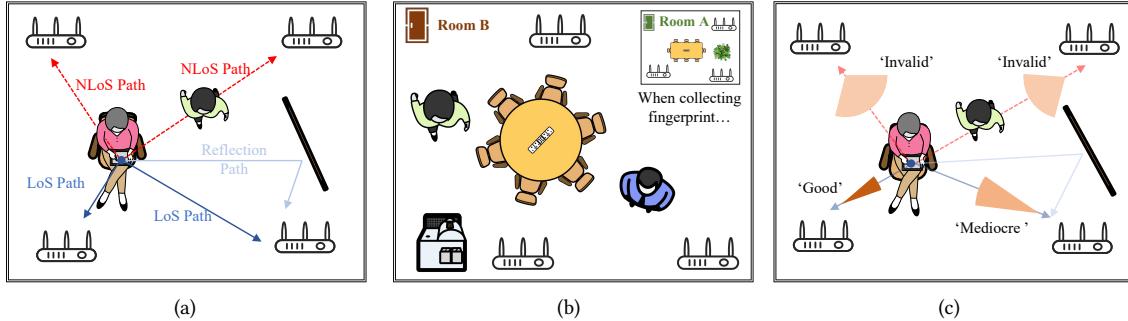


Fig. 1. An overall illustration of the existing WiFi localization systems and RLoc. (a) Angle-based localization is compromised when the direct path is blocked; (b) Changing environments can lead to localization errors as the previously stored fingerprints may not correspond to current conditions; (c) RLoc quantifies the uncertainty level in angle estimations and leverages these uncertainties to enhance the reliability of localization and tracking.

environment. Given this demand, WiFi has emerged as the most promising candidate and attracted significant attention due to its ubiquitousness [31]. If we can utilize the existing WiFi infrastructure in homes, universities, office buildings, and shopping malls for location-based services, it is conceivable that indoor positioning will become as widespread and indispensable as the Global Positioning System (GPS).

Over the past two decades, existing systems have achieved impressive results with median accuracy at the decimeter-level. Most of these systems model and estimate the parameters of the direct path, such as angle-based [12, 26, 39, 52, 62] and distance-based [13, 43, 51, 63] solutions, then use triangulation or trilateration for location services. However, as shown in Fig. 1(a), practical environments pose challenges, as the human body [64] or other obstructions can often block the Line-of-Sight (LoS) path, while changes in surrounding reflectors also impact WiFi signal measurements. When some direct paths are heavily attenuated or obstructed, the model of these systems may no longer be applicable, causing triangulation and trilateration to perform inadequately with significant positioning errors.

To address these issues, prior works [2, 4, 31, 46, 47, 55] use data-driven methods to implicitly model the correspondence between the received signal and location. These systems construct and update the fingerprint database, which consists of unique features of each specific location. While fingerprint-based localization achieves good performance, the approach requires the collection of fingerprints, which can be labor-intensive and subject to over-fitting of specific environmental conditions. This means that the environmental changes, such as moving to a different room, can disrupt the implicit correspondences within the fingerprint database. As illustrated in Fig. 1(b), if we do not adapt the fingerprint database to the new room, the robustness of existing fingerprint-based localization systems is limited.

One straightforward approach to harness the strengths of both fingerprint-based and angle-based approaches is to implicitly model the correlation between the received signal and AoA, and then perform triangulation to obtain the final position of the target device. By integrating the channel-modeling feature of AoA into fingerprints, the system can improve its generalization across different environments, which reduces the cost of collecting fingerprints when adopting to a new environment. However, such an intuitive idea faces a significant challenge: *managing the unreliable AoA estimations caused by environmental changes and complex interferences*. Experiences from deployments have shown learning-based wireless localization methods to be sensitive to environmental changes [28, 58]. This means when deploying a trained model in a new environment, it often results in unreliable performance. Moreover, unreliable AoA estimations can still arise, resulting from complex

and dynamic interferences, as well as the limited spatial resolution [2, 26, 42]. In addition, real-world scenarios often present various occlusions leading to Non-Line-of-Sight (NLoS) conditions between the target device and Access Points (APs). It is unreasonable to expect neural networks to deduce the direct path without incorporating information about other reflecting surfaces. Consequently, uniformly treating all the AoAs from multiple APs would significantly affect the final location accuracy due to these unreliable AoA measurements. This, in turn, would inevitably amplify the impact of errors in triangulation.

In this paper, we introduce RLoc (**R**obust **L**ocalization System), which exploits the potential of deep learning for consolidating the robustness of localization and tracking. As illustrated in Fig.1(c), the key design principle of RLoc lies in quantifying the uncertainty level of those AoA measurements estimated by neural networks and then exploiting the uncertainty to benefit the reliable localization and tracking. Specifically, the uncertainties inevitably arise due to the complex and dynamic interference, as well as the limited spatial resolution of WiFi signal. These factors hinder the neural network to learn an accurate mapping between wireless signals and AoAs. Moreover, environmental changes can cause the test distribution to diverge from the training samples, leading to increased uncertainty and unreliable prediction accuracy [22]. To address these challenges, RLoc incorporates uncertainty quantification in its network design, guiding us in determining when the learning-based angle estimations are reliable. As a bonus, RLoc can adaptively prioritize the APs based on the quantified uncertainties to mitigate the effect of unreliable angle estimates and achieve robust localization. In addition, when a segment of angle estimates has poor uncertainty, we can minimize its impact by leveraging the continuity in the time series to speculate the angle tracks in the corresponding time. More specifically, RLoc recognizes that the beamwidth in the spatial spectrum is not only related to angle estimation but also contains information about uncertainty. Hence, RLoc first utilizes the appropriate manual feature extraction using the Array Signal Processing (ASP) techniques [26] to incorporate wireless localization knowledge in input representation. Then, the Kullback-Leibler (KL) divergence loss [19, 27, 32] and the ensemble technique [15, 27] are employed to evaluate the uncertainty. Finally, RLoc exploits the estimated uncertainty as prior knowledge to guide the localization and tracking process, which detects unreliable AoA measurements and leverages the diversity of APs and the temporal continuous information of AoAs to achieve reliable localization and tracking results.

In our implementation of RLoc, we utilize two datasets collected from the commercial off-the-shelf WiFi devices. We conduct in-domain evaluations in eight unique scenario sets and cross-domain evaluations across all twenty possible combinations of training and testing scenarios. Importantly, the robustness of our system is demonstrated by implementing a five-fold cross-validation technique across all conducted experiments. In in-domain scenarios, RLoc outperforms the state-of-the-art methods [2, 26, 35, 52], achieving an average reduction in median localization error by 36.2%. In cross-domain scenarios, RLoc continues to demonstrate its robust performance, exhibiting an average reduction in median localization error by 20.40%. In our ablation analysis of the uncertainty modules, we note that their inclusion led to a significant reduction in both the mean average localization error and track error, by 20.40% and 7.12% respectively. Our extensive experiments demonstrate the effectiveness of RLoc as a prototype system for robust localization and tracking.

We summarize the main contributions of our work as follows:

- **Exploiting underutilized spatial spectrum information.** RLoc recognizes that the beamwidth in the spatial spectrum is not only related to angle estimation but also contains information about uncertainty. By employing the appropriate manual feature extraction using the ASP technique, RLoc effectively achieves both angle estimation and uncertainty quantification.
- **Exploring uncertainty in angle estimation tasks using deep learning.** RLoc incorporates uncertainty quantification in its network design, guiding us in determining when the network's angle estimations are reliable. To achieve this, RLoc first proposes using the KL divergence loss to evaluate the uncertainty

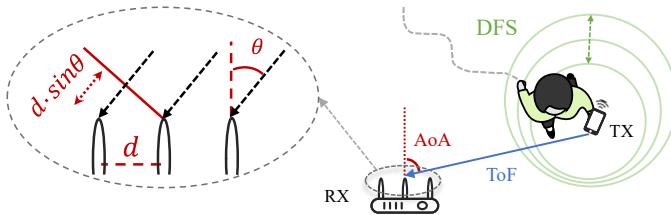


Fig. 2. Location or motion parameters of the direct path, including AoA, ToF, DFS. The expected phase differences across the receiving antennas vary with the AoA.

arising from noise and limited resolution. Subsequently, it employs the ensemble technique to measure the uncertainty caused by out-of-distribution test samples.

- **Utilizing uncertainty feedback to enhance localization and tracking.** RLoc utilizes uncertainty as prior knowledge to guide the localization and tracking process. This approach detects unreliable measurements and leverages the diversity of APs and the temporal continuous information to achieve reliable localization and tracking results.
- **The first human-held device WiFi localization and tracking dataset<sup>1</sup>.** We construct a human-held devices WiFi localization dataset, consisting of approximately 120k data points collected from ten volunteers across four classic indoor scenarios. We believe this dataset would facilitate future research on WiFi-based localization.

The rest of this paper is organized as follows. In Section §2, we establish an explicit model that relates position-dependent features to WiFi signals, which serves as the foundation for subsequent analyses. In Section §3, we explore the variability of uncertainty in learning-based AoA estimation task. In Section §4, we introduce RLoc, a system that integrates deep learning and uncertainty in the wireless localization field to achieve robust localization and tracking. In Section §5.1, we provide a comprehensive description of the human-held device dataset collected for evaluating our system. In Section §6, we evaluate the performance of RLoc on two datasets and conduct experiments in both in-domain and cross-domain scenarios. Section §7 provides an overview of the pertinent research. Section §8 discusses the limitations and we conclude our work in Section §9.

## 2 WIRELESS CHANNEL MODEL

In this section, we construct a model that explicitly links position-dependent features to WiFi signals. We begin by presenting a general and ideal wireless channel model. Next, we refine the wireless channel model to incorporate practical considerations. These channel models serve as the foundation for robust localization in our subsequent discussions.

### 2.1 General Wireless Channel Model

WiFi signals, which are emitted from the user-held device propagate through multiple paths and eventually arrive through superimposition at the APs. As users move around while seeking indoor navigation assistance, both the LoS and NLoS paths (i.e., direct and non-direct paths) undergo corresponding changes. The changes in these paths lead to variations in the final received signal. Thus, theoretically we are able to connect wireless channels and user's locations by separating superimposed signals and accurately estimating the location or motion parameters

<sup>1</sup>Our dataset is publicly available at [https://github.com/H-WILD/human\\_hold\\_device\\_wifi\\_inoor\\_localization\\_dataset](https://github.com/H-WILD/human_hold_device_wifi_inoor_localization_dataset).

of the direct path, such as AoA, Time of Flight (ToF), Doppler frequency shift (DFS) and complex attenuation, as shown in Fig. 2.

Commodity WiFi chips model the multiple propagation paths between a transmitter and receiver as the Channel State Information (CSI). The overall attenuation and phase shift measured at the granularity of the Orthogonal Frequency Division Multiplexing (OFDM) sub-carrier by each antenna is reported in a  $M \times N \times K$  format -  $M$  transmitting antennas,  $N$  receiving antennas and  $K$  subcarriers. To be scalable and inclusive of devices such as smartwatches, IoT sensors, and older phone models that can only support a single antenna, we set  $M = 1$  as the default value. We first express the received CSI  $H_k \in \mathbb{C}_{N \times 1}$  of the  $k$  subcarrier with frequency  $f_k$  as:

$$H_k = H_k^{LoS} + \sum_{l \in \mathcal{L}_N} H_k^l + W = \sum_{l=1}^L \alpha_l \Theta(\theta_l) T_k(\tau_l) \Gamma_k(\gamma_l) + W, \quad (1)$$

where  $\mathcal{L}_N$  denotes the NLoS paths set and  $W$  is the background noise which is usually assumed to be white Gaussian. Moreover, we model the multiple paths using the location and motion parameters, where  $\alpha_l$ ,  $\theta_l$ ,  $\tau_l$ ,  $\gamma_l$  are the complex attenuation, AoA, ToF and DFS of the  $l$ -th path, respectively.  $\Theta(\theta_l)$  expresses the expected phase differences across the receiving antenna array and can be written as:

$$\Theta(\theta_l) = [1, e^{-j2\pi d f_k \sin(\theta_l)/c}, \dots, e^{-j2\pi(N-1) d f_k \sin(\theta_l)/c}]^T, \quad (2)$$

where  $d$  denotes the inter-antenna spacing of the uniform linear array and  $c$  is the speed of light. The phase shift introduced by the ToF  $\tau$  can be denoted as:

$$T_k(\tau_l) = e^{-j2\pi f_k \tau_l}, \quad (3)$$

and the DFS caused by the motion of the transmitter can be expressed as:

$$\Gamma_k(\gamma_l) = e^{j\gamma_l} = e^{j \int_t 2\pi f_k \frac{v_l(t)}{c} dt}, \quad (4)$$

where  $v_l$  is the Doppler velocity (i.e., the change speed of the path length) and  $t$  is the time. Then by combining the models of Eq. 1, we can represent the CSI  $H \in \mathbb{C}_{NK \times 1}$  in terms of signal dimensions as<sup>2</sup>:

$$H = [H_1, H_2, \dots, H_K] = \sum_{l=1}^L \alpha_l \Theta(\theta_l) \otimes (\Omega(\tau_l) \odot G(\gamma_l)) + W, \quad (5)$$

where  $\otimes$  denotes the tensor product and  $\odot$  represents the Hadamard product. And we can express the phase shift  $\Omega(\tau_l)$  introduced by ToF, as well as the phase shift  $G(\gamma_l)$  caused by DFS at each subcarrier, in the following manner:

$$\Omega(\tau_l) = [T_1(\tau_l), T_2(\tau_l), \dots, T_K(\tau_l)]^T = T_1(\tau_l) [1, e^{-j2\pi f_\delta \tau_l}, \dots, e^{-j2\pi(K-1) f_\delta \tau_l}]^T, \quad (6)$$

$$G(\gamma_l) = [\Gamma_1(\gamma_l), \Gamma_2(\gamma_l), \dots, \Gamma_K(\gamma_l)]^T, \quad (7)$$

where  $f_\delta$  is the frequency interval of adjacent subcarriers.

## 2.2 Revisiting Wireless Channel Model

While the signal model for location and motion parameters may be elegant, separating multipath signals can often be challenging due to the practical constraints caused by WiFi devices. This includes:

**Challenges in Resolving DFS Due to Sampling Rate Limitations.** To obtain accurate Doppler frequency shifts in CSI, the sampling rate must exceed twice the maximum shift to satisfy the Nyquist sampling theorem. However, the WiFi CSI sampling rate depends on the CSI estimation models, which can be classified into three

<sup>2</sup>The phase shift introduced by the  $l$ th path's AoA across adjacent subcarriers can be expressed as  $2\pi(m-1)d f_\delta \sin(\theta_l)/c$ , which is small and negligible [26].

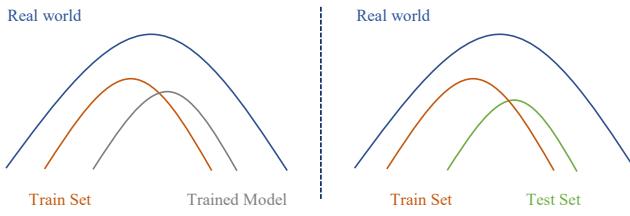


Fig. 3. Challenges in Assumptions: The trained model often struggles to perfectly fit the feature distribution of the training data due to factors such as noise and resolution. Additionally, since the training dataset is a subset of the real world, it is possible for the testing dataset to not conform to the feature distribution of the training dataset.

types according to the types of source frames [60]. In the beacon frame model, typically, the CSI measurement sampling rate ranges from 10-20Hz [1, 3]. In the data frame model, sampling of CSI can be unstable due to the natural randomness of data frames. Detecting the subtle DFS caused by human motion from such low or unstable sampling rates is challenging because the DFS can reach tens of hertz for a 5 GHz channel [29, 45]. While researchers can modify the transmitting interval in the injection frame model, it negatively impacts the communication of the current devices. Thus, we consider DFS as an inevitable noise, denoted as  $W_T$ , that cannot be effectively resolved under typical packet receiving conditions.

**Challenges in Estimating ToF Due to Imperfect Inertial Circuits Contamination.** Due to imperfect synchronization between the transmitter and receiver, the phase shift is contaminated by various factors such as symbol timing offset (STO), sampling frequency offset (SFO), carrier frequency offset (CFO), and carrier phase offset (CPO) [40, 50, 67]. This results in a time-varying random phase offset  $\Omega_n$  affecting the phase shift of Eq. 6, where  $\Omega_n \in \mathbb{C}_{K \times 1}$  represents the STO, SFO, CFO, and CPO noise of each subcarrier. Consequently, the absolute value of  $\tau_l$  is not determinable without special time synchronization techniques. However, since all paths in the channel have the same time-varying phase offset  $\Omega_n$ , the relative time of flight (rToF) values between propagation paths can be used to estimate the channel information [26, 39].

After taking into account these practical constraints, we can represent the CSI channel model in terms of signal dimensions as follows:

$$\widehat{H} = \sum_{l=1}^L \alpha_l \Theta(\theta_l) \otimes (\Omega(\tau_l) \odot \Omega_n) + W_T + W, \quad (8)$$

where all paths in the channel experience the same time-varying phase offset  $\Omega_n$  and the phase shift of Eq. 7 is treated as noise.

### 3 UNDERSTANDING THE UNCERTAINTY OF LEARNING-BASED AOA ESTIMATION

In this section, we explore the variability of uncertainty in learning-based AoA estimation, which influences the system's robustness. We start with an intuitive analysis of our motivation.

#### 3.1 Motivation of Uncertainty

To build a more robust localization and tracking system, it is crucial to achieve accurate AoA estimation. Intuitively, if we possess a precise model of the environment, accurate positioning can be achieved even in cases where the LoS path is blocked or multiple reflectors are present [2]. However, explicitly modeling the effects of complex environments on signal propagation is considerably difficult, due to the complex combination of diffraction, reflection, and refraction [30, 35]. Additionally, determining the environmental model can be challenging as well,

considering the variability in room layouts and continuous changes caused by human movements. Consequently, we seek to leverage the powerful complex mapping ability of neural networks, which have demonstrated outstanding performance in a wide range of applications, to model the relationship between received signals and AoAs. This kind of data-driven method enables us to establish an implicit representation of the environment and incorporate the environmental impacts on AoA into network parameters.

Despite our efforts to improve the accuracy of AoA estimation, uncertainties inevitably arise due to limited spatial resolution, as well as complex and dynamic interference. Specifically, commodity WiFi typically has a bandwidth of 20, 40, or 80 MHz, which corresponds to a distance resolution of 15m, 7.5m, and 3.75m, respectively. This means that we can only resolve the LoS path from multipath when they are separated by more than the corresponding distance resolution in the range domain. Furthermore, both the number of antennas and the configuration of antenna array inherently limit the angular resolution[26, 42]. As a result, due to the limited spatial resolution, closely-spaced multipath reflections cannot be resolved theoretically [5, 8, 11], leading to uncertainty about the accurate AoA extraction from the LoS path. In addition, in real-world scenarios, various occlusions can lead to NLoS conditions. For instance, when a user holds a device and walks around while seeking location services, the LoS conditions with different APs will vary based on the user's orientation and position. Even after deploying the trained neural network, such situations still occur and result in inaccuracies of AoA estimation. As such, the implicit mapping between WiFi received signals and AoAs might be corrupted or absent, which could lead to inaccurate parameters for neural network learning and unreliable estimates during neural network testing. Moreover, experiences from deployments have shown learning-based wireless localization methods to be sensitive to environmental changes [28, 58]. This means when deploying a trained model in a new environment, it often results in unreliable performance.

Given the analysis above, uniformly treating all the AoAs from multiple APs would significantly affect the final location accuracy due to these unreliable AoA measurements. This, in turn, would inevitably amplify the impact of errors in triangulation. However, if we can quantify the uncertainties associated with AoA measurements, we would be able to utilize uncertainty feedback to enhance localization and tracking. Specifically, we could dynamically prioritize the APs based on their quantified uncertainty to diminish the impact of unreliable AoA estimates. Additionally, we could identify which segments in the AoA time series carry unreliable estimates and filter these outliers by utilizing the continuity of the track.

*Takeaway:* To achieve a ubiquitous and robust localization and tracking system, we focus not only on enhancing the accuracy of AoA estimation, but also on quantifying the uncertainties that arise when WiFi signals fail to provide precise AoA results.

### 3.2 Investigation of Learning-based AoA Estimation Uncertainty

To successfully employ neural networks for AoA estimation, it is essential to investigate how neural networks function in the AoA estimation task and how uncertainty arises. Specifically, suppose that the training dataset  $D = \{x_s, y_s\}_{s=1}^S$  comprises  $S$  inputs  $x \in \mathbb{X}_{train}$  and target values  $y \in \mathbb{Y}_{train}$ . These samples, adhering to the distribution  $p_{train}(x, y)$ , are also assumed to be independent and identically distributed (i.i.d) [44] and sampled from the true but unknown distribution  $p^*(x, y)$ . The input  $x \in \mathbb{X}_{train}$  typically corresponds to the feature of the received WiFi signal, such as the raw signal or signal covariance matrix. The label  $y$  is a regression estimation of the direct path AoA and the output space  $\mathbb{Y}_{train}$  is assumed to consist of continuous real-valued numbers. During training, the neural network is expected to model  $p_{\omega^*}(y|x)$ , where  $\omega^*$  represents the ideal set of model parameters that could perfectly map the input space  $\mathbb{X}_{train}$  to the target space  $\mathbb{Y}_{train}$ . Therefore, we can use the trained model and deploy it for practical use, provided that the test dataset follows the same distribution  $p^*(x, y)$ .

In practice, however, there are two issues with these assumptions, which are illustrated in Fig. 3. Firstly, it is impossible to ensure that the dataset is entirely free of noise, as measurement errors in the received WiFi

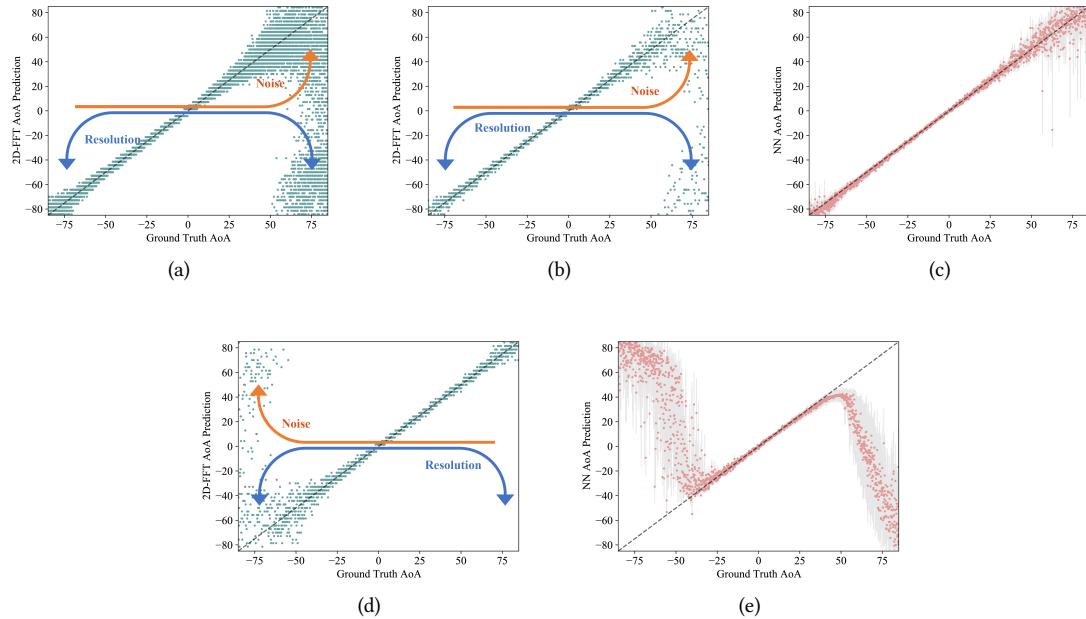


Fig. 4. An example illustrating the uncertainty of learning-based AoA estimation. The green and red scatterers symbolize the AoA estimations generated by the 2D-FFT method and the network, respectively. The black dotted line denotes the ground truth AoA, with the quantified uncertainty represented by the shaded gray area. Notably, the angular resolution diminishes as the absolute value of the ground truth AoA escalates, a phenomenon attributable to the linear array configuration. (a) A training set consisting of 17k samples, where the noise level remains constant in the negative half-axis, but increases in the positive half-axis. (b) An i.i.d test set with 1.7k samples that maintains the same noise setting as the training set. (c) The performance of the neural network on the i.i.d test set, where both a decrease in resolution and an increase in noise levels contribute to inaccurate angle estimation in certain regions. (d) An o.o.d test set with 1.7k samples where the noise settings for the positive and negative half axes are interchanged, thus differing from the training set. (e) The performance of the neural network on the o.o.d test set, where substantial errors manifest in regions where the noise conditions diverge from those of the training set.

signal and label observations can introduce noise. Moreover, the received WiFi signal may have incomplete information due to limited spatial resolution from the measurements. These inherent and irreducible data noises cause uncertainty in the mapping function parameter  $p_\omega(y|x)$ , where  $\omega$  is close to but not equal to  $\omega^*$ . This could result in the model predictions incorrect even when the test samples come from the same distribution as the training dataset. Secondly, the training samples cannot encompass all the multipath scenarios in practical settings, which means the training set is a subset of the real world. Thus the training and test samples are likely to come from different distributions, which is known as data drift [36]. Once the model is deployed, the indoor localization environment often undergoes changes, such as moving furniture and adjusting the position and orientation of APs. These changes can affect the multipath of signal propagation, which ultimately affects the input features of the model. These changes can affect the multipath of signal propagation, which further changes the distribution of input features  $p(x)$ , and ultimately leading to poor prediction accuracy [22].

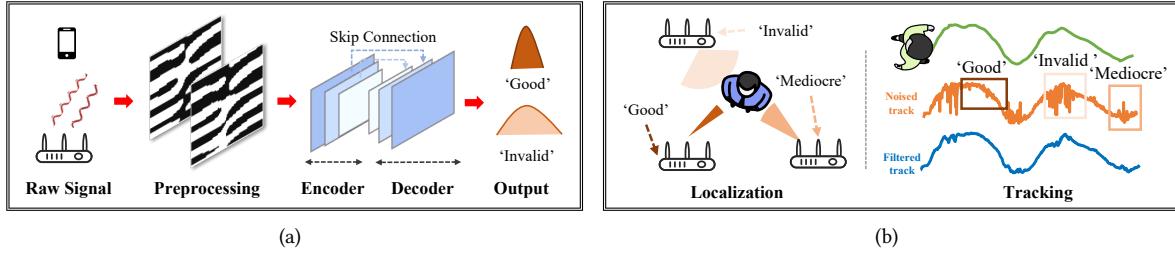


Fig. 5. System Framework of RLoc. **(a)** RLoc collects and processes the raw signals from all the APs, subsequently estimating the angle and its associated uncertainty through a learning-based method. **(b)** RLoc detects unreliable measurements and leverages the diversity of APs and the temporal continuous information to achieve reliable localization and tracking results.

To further illustrate this uncertainty, we generate a CSI simulation dataset while controlling the noise level, as shown in Fig.4(a). In order to simulate the received WiFi signal, we assume a simple case where  $M \times N \times K = 1 \times 2 \times 5$ ,  $L = 1$ ,  $W_\Gamma = 0$ , and the elements of  $\Omega_n$  are all 1s in Eq.8. We then control the noise level of  $W$  to obtain various signal-to-noise ratios (SNRs) for different ground truth angles, where the noise situation can be reflected in the AoA estimation results of the 2D-FFT [17]. In the positive half-axis, the noise level increases as the ground truth AoA changes from  $0^\circ$  to  $85^\circ$ . In the negative half-axis, the noise level is constant while angular resolution decreases [42] as the ground truth AoA changes from  $0^\circ$  to  $-85^\circ$ . Thus, we can observe the impact of measurement errors and incomplete information on the neural network from the positive and negative axes, respectively. As illustrated in Fig.4(b), we generate an i.i.d test dataset that samples from the same distribution as the training dataset. As shown in Fig.4(c), both measurement errors and incomplete information affect the learning of network parameters  $\omega$  and can lead to varying degrees of AoA estimation error. We also generate an out-of-distribution (o.o.d) test dataset that samples from  $q(x, y) \neq p_{train}(x, y)$ , as shown in Fig.4(d). We interchange the noise settings for the positive and negative half axes. As shown in Fig.4(e), the network tends to incorrectly estimate o.o.d inputs.

In summary, uncertainty arises due to the discrepancy between reality and assumptions. Firstly, it is not feasible to assume that the network parameter  $\omega$  equals  $\omega^*$  after model training. This is because noise interference and limited resolution hinder the network's ability to understand the mapping relationship between wireless signals and AoAs. Additionally, we cannot guarantee that  $p_{train}(x, y)$  is equivalent to  $p^*(x, y)$ . This implies that the training set does not entirely represent the real world. Although the test set may sample from  $p^*(x, y)$ , it can still be significantly different from the training set, which can result in a decline in the network's performance. Consequently, neural networks might confidently provide incorrect AoA estimations, and failure to identify these inaccuracies can lead to poor localization results.

#### 4 ROBUST INDOOR LOCALIZATION DESIGN

In this section, we introduce RLoc, a Single-Input Multiple-Output (SIMO) and snapshot-based system that integrates wireless localization domain knowledge and deep learning to achieve robust localization. First, we provide an overview of our approach in Section §4.1. Next, we detail the network architecture and discuss the approach to quantify the uncertainty of neural network output in Section §4.2. We then explore how to appropriately extract manual feature extraction using array signal processing techniques for enhanced performance in Section §4.3. Lastly, we explain the approach of utilizing uncertainty feedback to augment localization and tracking in Section §4.4.

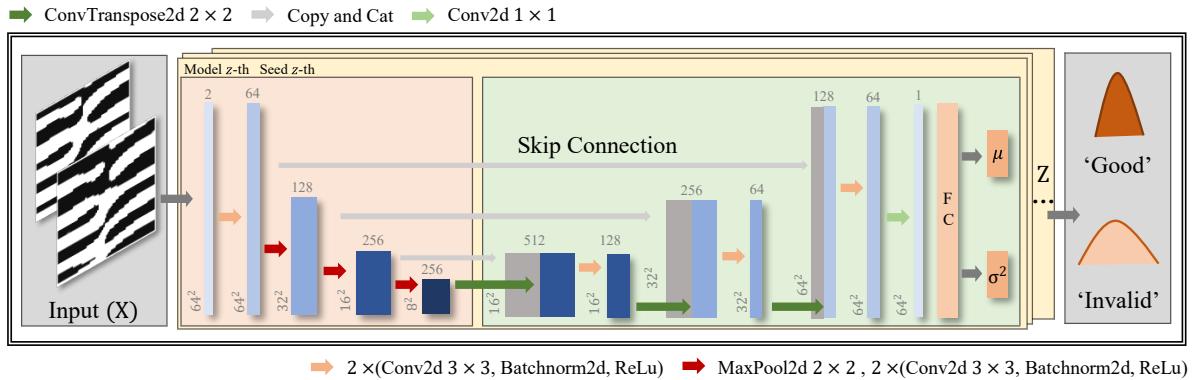


Fig. 6. RLoc network architecture. All the  $Z$  models share this uniform structure. Each trained model takes the real and imaginary components of the representation  $X$  as inputs, and outputs the mean  $\mu$  and variance  $\sigma$  of a Gaussian distribution  $\mathcal{N}(\mu, \sigma)$ . Varied shades of blue boxes denote multi-channel feature maps, with the channel count listed above and x-y dimensions in the lower left. Grey boxes represent copied feature maps and arrows indicate the operations carried out. Finally, RLoc merges the outputs of the  $Z$  models to indicate the predicted angle and its associated uncertainty.

#### 4.1 System Overview

As shown in Fig. 5, RLoc begins with pre-processing on the raw WiFi sensing data, then estimates the angle and its associated uncertainty using a learning-based method. Finally, it applies non-learning algorithms to integrate the AoAs and their uncertainties from multiple APs for accurate target localization and tracking. In detail, RLoc comprises three components: (i) *Raw Data Pre-processing*. Initially, we incorporate wireless localization knowledge into the input representation by pre-processing the raw WiFi signal and manually extracting features. These features contain both AoA and uncertainty information, as detailed in Subsection §4.3. (ii) *Uncertainty Quantification via Neural Networks*. We adjust an encoder-decoder architecture with skip-connections [37] to fit our requirements. Furthermore, we incorporate uncertainty estimation into the design of neural networks based on the two issues of the assumption mentioned in Subsection §3.2. We achieve this by replacing point outputs with probabilistic outputs and designing an ensemble module, as described in Subsection §4.2. (iii) *Uncertainty-assisted Localization and Tracking*. We exploit the estimated uncertainty as prior knowledge to guide the localization and tracking process, which detects unreliable AoA measurements and leverages the diversity of APs and the temporal continuous information of AoAs to achieve reliable localization and tracking results.

#### 4.2 Incorporating Uncertainty in Neural Network

Similar to beams, ground-truth AoA values should not be considered as exact points. Due to noise in the wireless signal and label, ground-truth AoA values are inherently ambiguous, which is further exacerbated when the direct path is occluded. In regression, the most common approach is to use a neural network to directly predict the AoA  $y$  from a given test input  $x$ , such that  $y = f_\omega(x)$ , where  $\omega$  is learned by minimizing the L1 or L2 comparative loss over the training dataset. However, this direct regression does not model the ambiguity.

In contrast, recent studies [19, 27, 32] have explored predicting the distribution  $\mathcal{N}(\mu, \sigma)$ , where the network predicts the mean  $\mu$  and variance  $\sigma$  of the target  $y$ . This method not only predicts the angle value but also captures the uncertainty associated with it. Given a Dirac delta function  $\delta_f$  of target label  $y$ , we can minimize the KL-Divergence between  $\mathcal{N}(\mu, \sigma)$  and  $\delta_f(y)$  by:

$$L_{KL} = D_{KL}(\mathcal{N}(\mu, \sigma) || \delta_f(y)) \propto \frac{\log(\sigma^2)}{2} + \frac{(y - \mu)^2}{2\sigma^2}. \quad (9)$$

KL-Divergence captures the aleatoric uncertainty in training process, which represents the gap between the actual  $p_\omega(y, x)$  and the ideal  $p_{\omega^*}(y, x)$ , as described in Section §3. However, the gap between  $p_{train}(y, x)$  and  $p^*(x, y)$ , also known as epistemic uncertainty, is not within its scope. To address this, we use the technique of deep ensemble [15, 27] to capture the uncertainty when the deployed model encounters an o.o.d test set. Deep ensemble is based on the intuition that when multiple independently trained models yield consistent outputs, their predictions tend to be more reliable. Conversely, if the models produce inconsistent results, it often indicates that the predictions are unreliable. As demonstrated in Fig. 6, we train  $Z$  models independently with the same structure and on the same training set, but with different random seeds. We treat the final AoA estimate  $\hat{\theta}$  as a uniformly-weighted mixture result of  $Z$  model outputs, which can be expressed as

$$\hat{\theta} = \sum_{z=1}^Z \mu_z / Z, \quad (10)$$

where we set  $Z = 5$  by default and  $\mu_z$  denotes the mean prediction of the  $z$ -th model. However, as the training dataset better captures the real-world feature distribution [41],  $Z$  can be reduced. This is because, with an increasing dataset size, the proximity between  $p_{train}(x, y)$  and  $p^*(x, y)$  also increases, leading to a gradual reduction in epistemic uncertainty. As a result, the final variance  $\hat{\sigma}$  can be expressed as:

$$\hat{\sigma} = \sqrt{\sum_{z=1}^Z (\hat{\theta} - \mu_z)^2 / Z + \sum_{z=1}^Z \sigma_z / Z}, \quad (11)$$

where  $\sigma_z$  denotes the variance prediction of the  $z$ -th model and the quantified uncertainty is denoted as the shaded gray in Fig. 4. Besides, as depicted in Fig. 6, our network model, rooted in the U-Net architecture [37], undergoes modifications to fit our requirements. We modify both the depth and width of the network model and append a Fully Connected (FC) layer to fit the design of uncertainty estimation.

### 4.3 Incorporating Wireless Localization Knowledge for Input Representation

So far, we are able to quantify the output uncertainty of neural networks, however, there still remains prior knowledge in the wireless localization field that we can incorporate and leverage in neural network framework. Previous studies [9, 30, 35, 66] in AoA estimation typically input a complex-valued matrix, such as the raw signal or the signal covariance matrix, to a neural network. However, this representation does not fully leverage the prior knowledge in WiFi localization to extract features and may not be compatible with most state-of-the-art deep learning algorithms [2]. In contrast, appropriate feature extraction can prevent unnecessarily complexity for the neural network. To assist with AoA estimation, we distill the essential information which can be commonly reduced to two dimensions: AoA and ToF, as mentioned in Section §2. Recall, angle-based localization systems typically utilize ASP algorithms [17] to transform the CSI matrix  $\tilde{H}$  into a 2D spatial heatmap  $X \in \mathbb{C}_{g_A g_D \times 1}$  of AoA and ToF, which can be expressed as:

$$X = \mathbf{A}^H(\theta, \tau) \tilde{H} \tilde{H}^H \mathbf{A}(\theta, \tau), \quad (12)$$

with  $\mathbf{A}(\theta, \tau) \in \mathbb{C}_{g_A g_D \times NK}$  defined as follows:

$$\mathbf{A} = [\Theta(\theta_1) \otimes \Omega(\tau_1), \Theta(\theta_2) \otimes \Omega(\tau_1), \dots, \Theta(\theta_{g_A}) \otimes \Omega(\tau_1), \dots, \Theta(\theta_{g_A}) \otimes \Omega(\tau_{g_D})], \quad (13)$$

where  $g_A$  and  $g_D$  denote the number of candidate AoA and ToF, respectively, with default values set to  $g_A = 64$  and  $g_D = 64$ . Then they choose the highest peak or the peak with the smallest ToF from this heatmap to identify the direct path and use triangulation or trilateration to locate the user's device. However, relying solely on the 2D heatmap's special peak may not extract its full potential information. Moreover, selecting the correct location is not as simple as merely choosing the special point in the 2D heatmaps, as a particular peak may result from the

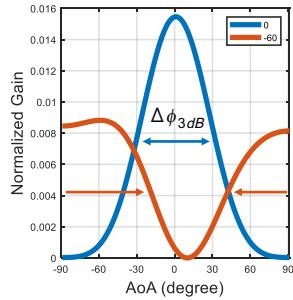


Fig. 7. A diagram of beamwidth. The beam pattern corresponding to a true angle of  $-60$  degrees is depicted in orange, while the beam pattern associated with a true angle of  $30$  degrees is represented in blue.

Table 1. Comparison of Pearson correlation coefficient [20] (PPC) between two variables. The PPC ranges between  $-1$  and  $1$ , with a value closer to  $1$  indicating a stronger positive linear correlation.

PPC	S1	S2	S3	S4	S5	S6
$\theta_{max} - \theta_{label}$	0.70	0.47	0.55	0.62	0.27	0.59
$\theta_{beam} - \theta_{label}$	<b>0.71</b>	<b>0.49</b>	<b>0.58</b>	<b>0.64</b>	<b>0.30</b>	<b>0.61</b>
$\theta_{max} - \theta_{error}$	-0.06	-0.37	0.04	0.05	0.23	0.25
$\Delta\phi_{3dB} - \theta_{error}$	<b>0.41</b>	<b>0.40</b>	<b>0.29</b>	<b>0.26</b>	<b>0.35</b>	<b>0.28</b>

superposition of multiple paths due to limited resolution. Luckily, we uncover a strong correlation between the beamwidth in the 2D heatmap and AoA estimation, which guides us to leverage insights from AoA estimation knowledge in the network input design. Specifically, beamwidth is defined as the angle between the half-power points of the main lobe. It can be written in a simple form of the angle-space expression [33]:

$$\Delta\phi_{3dB} = \begin{cases} 107.86^\circ \sqrt{\frac{\lambda}{Nd}} b, & \text{for } \theta = 90^\circ, -90^\circ \\ \frac{50.76^\circ}{\cos\theta} \frac{\lambda}{Nd} b, & \text{for } -90^\circ < \theta < 90^\circ, \end{cases} \quad (14)$$

where  $b$  denotes the so-called broadening factor, whose value is constant. As shown in Fig. 7, the beamwidth  $\Delta\phi_{3dB}$  will broaden as the AoA varies from  $0^\circ$  to  $\pm 90^\circ$  due to the effect of  $\cos\theta$ . This implies a significant relationship between beamwidth and AoA, suggesting that the angle can be estimated through  $\Delta\phi_{3dB}$ . More importantly, beamwidth also maps uncertainty information, as that the absolute value of  $\theta$  approaches  $90^\circ$ , the beamwidth expands, leading to less accurate AoA estimations. More specifically, as demonstrated in Tab. 1, the angle estimated through beamwidth  $\theta_{beam}$  displays a stable linear correlation with the true angle,  $\theta_{label}$ . This correlation is similar with the angle determined by simply taking the maximum peak value  $\theta_{max}$ . In addition, beamwidth  $\Delta\phi_{3dB}$  displays a stable linear correlation with the error  $\theta_{error}$  between  $\theta_{label}$  and  $\theta_{max}$ . Consequently, we realize that the often overlooked beamwidth information in the spatial spectrum holds significance not only for angle estimation but also for uncertainty quantification. As such, we decide to employ  $X$ , which contains both the underutilized beamwidth feature and the special peak information, as the input to our network. To incorporate the existing techniques with angle estimation tasks, we decompose  $X$  into its real and imaginary components, which can be represented as  $[\Re(X), \Im(X)] \in \mathbb{R}_{2 \times g_A \times g_D}$ .

#### 4.4 Uncertainty-assisted Localization and Tracking

So far, we have demonstrated how RLoc can estimate the uncertainty of the AoA output of the neural network. As a bonus, the uncertainty can be considered as prior knowledge to guide the localization and tracking. In this section, we discuss how RLoc uses these uncertainties to pinpoint the target location and tracking.

**Improving Localization with Uncertainty:** RLoc is designed based on a key observation that, although the direct path connecting an AP and a user's device might be obstructed, it is unlikely that its location happens to be within the NLoS zones of all other APs. Therefore, if we can dynamically prioritize the APs based on their levels of uncertainty, we should be able to diminish the impact of unreliable network output on localization results. To

achieve more robust effect, we propose an uncertainty-aware algorithm that integrates AoA estimation and its uncertainty from multiple APs to locate the target. Given the probability estimation network output from  $I$  APs, we can compute the likelihood  $L(\chi)$  of the user's device being at a specific location  $\chi$  as follows:

$$L(\chi) = \prod_{i=1}^I \frac{1}{\hat{\sigma}_i \sqrt{2\pi}} e^{-\frac{(\theta_i - \hat{\theta}_i)^2}{2\hat{\sigma}_i^2}}, \quad (15)$$

where  $\theta_i$  is the AoA from the candidate location  $\chi$  to  $i$ -th AP. And the target location can be the one with the maximal likelihood, which can be expressed as:

$$\chi^* = \arg \max L(\chi). \quad (16)$$

or, alternatively, the average coordinate of the locations with the top few likelihoods.

**Improving Tracking with Uncertainty:** RLoc decomposes trajectory tracking into two stages: AoA smoothing and trajectory smoothing. This is important since, for a moving target, both the AoAs and locations represent continuous information in the temporal dimension. Therefore, if we can identify which segments in the AoA and location time series carry unreliable estimates, we should be able to filter these outliers by utilizing the continuity of the track. To achieve this, we propose an AoA filter algorithm that incorporates the quantified uncertainty into its filtering process. Given the probability estimation network output  $(\hat{\theta}_t, \hat{\sigma}_t)$  of time-series  $t$ , we model the AoA smoothing problem using the Kalman filter [24]. The true state at time  $t$  evolves from the state at  $(t-1)$  according to:

$$x_t = x_{t-1} + W_p, \quad (17)$$

an observation  $z_k$  of the true state  $x_k$  is made according to:

$$z_t = x_t + W_o, \quad (18)$$

where  $W_p$  and  $W_o$  are zero-mean normal distributions. During each time step of the Kalman filter, we treat the observation  $z_k$  as the AoA estimation  $\hat{\theta}_t$  and the observation noise as the uncertainty output  $\hat{\sigma}_t$ . After obtaining the filtered angle track, we perform triangulation to obtain positions in time series and then apply Kalman filtering to obtain the final trajectory.

## 5 DATASETS

In this section, we begin by providing a comprehensive description of the dataset collected for evaluating our model. Subsequently, we analyze the accuracy of the ground truth labels collected.

### 5.1 Dataset Overview

The public Robot-based Dataset<sup>3</sup> employs a robot equipped with Simultaneous Localization And Mapping (SLAM) technology to gather ground truth location estimates corresponding to wireless channels. However, real-world usage scenarios often involve users holding their devices in their hands, introducing additional effects such as occlusion caused by human bodies [64] that are not modeled in existing localization system architectures. To address this gap, we generalize the scenario from robot-held to human-held devices. We use an Ultra-Wideband (UWB) based localization system with an accuracy of ten of centimeters to collect ground truth location data, and we build a Network Time Protocol (NTP) server to achieve millisecond-level synchronization between all devices. Our First Human-held Devices Dataset comprises approximately 120k datapoints from ten volunteers across four classic indoor scenarios. We believe this dataset would facilitate future research on WiFi-based localization.

We collect datasets in four typical indoor scenarios: conference, laboratory, office, and lounge, as illustrated in Fig. 8. In these scenarios, the laboratory scene mainly consists of a simple line-of-sight environment, with the

---

<sup>3</sup>The description of robot-based dataset is accessible at <https://wcsng.ucsd.edu/wild/>.

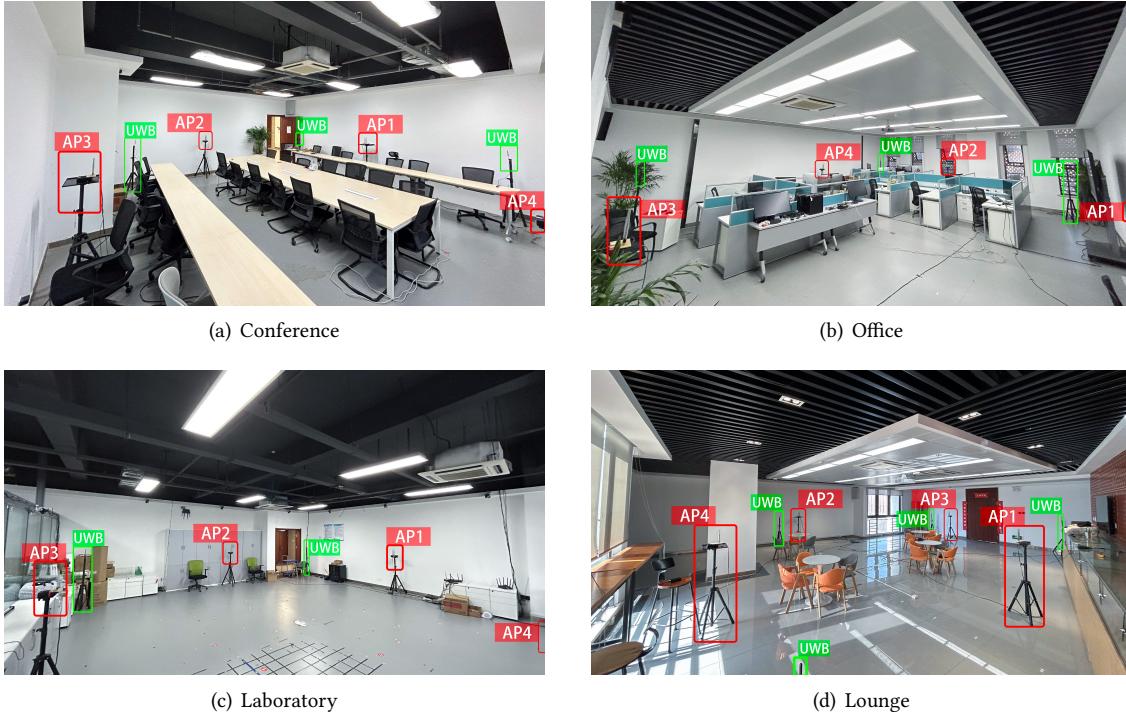


Fig. 8. Scenario Deployment.

primary reflections originating from walls and ceilings. In contrast, the conference features more tables and chairs as reflectors. The office scene also contains numerous tables and chairs, along with additional display screens and large monitors. The lounge scene has many tables and chairs as well, with interference from non-line-of-sight areas formed by wall occlusion and reflections from decorative metal walls. We deploy five mini-computers equipped with off-the-shelf Intel 5300 WiFi NICs in each space, four as receivers and one as transmitter. All the receivers are equipped with three omni-directional antennas in a uniform linear array, while the transmitter is equipped with one. The antenna spacing is 2.6cm, which is nearly half the wavelength of 5.3GHz band. And the receivers are positioned on tripods that stand at the same height. In order to facilitate users' mobility with their devices, we equip the transmitter with an uninterrupted power supply. Moreover, we estimate time-stamped CSI data for each AP every 5 ms using the Linux 802.11n CSI tool [16]. The experiments are conducted on the channel 60 at 5.3GHz, utilizing 40 MHz bandwidth. In addition, the captured data is pre-processed using the phase calibration and ToF sanitization algorithms [26, 59]. Besides, we employ the low-power low-cost UWB transceiver DW1000 [7], which supports a sampling rate of 10Hz, in all scenarios to collect the ground truth labels. Specifically, we mount four UWB anchors on tripods at a uniform height of either 1.3m or 1.5m, ensuring clear LoS conditions above any potential obstacles. The UWB tag, which is battery-powered, is fastened to the WiFi transmitter and can move freely. During data collection, volunteers are instructed to walk freely around the scenario while holding the transmitter in their hands. They can walk slowly, walk quickly, or stop, just as they normally would do during their daily activities. Each volunteer walks alone for 6 minutes and then with other persons, who could potentially cause interference, for 3 minutes. During data processing, we down-sample the

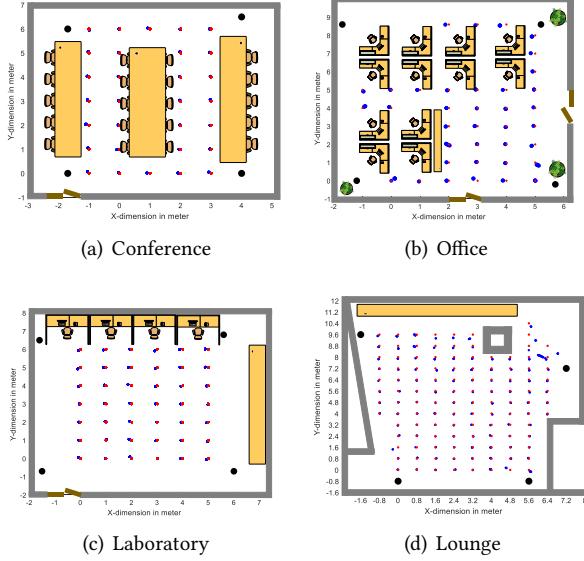


Fig. 9. Evaluation setup and results in a 2D Plane: Uniformly distributed evaluation points are depicted as red dots across four rooms. The UWB anchors are denoted by grey dots, while the 2D localization results are designated by blue dots.

CSI data to 10 Hz to approximate the practical sampling rate of the WiFi devices. Additionally, we implement a simple linear interpolation for the data recorded by the UWB localization system to match the timestamps of the WiFi signals after temporal alignment.

## 5.2 Ground Truth Accuracy

We evaluate the accuracy of the ground truth labels collected across four scenarios and conduct a targeted experiment to meticulously examine the effects of human body occlusion on accuracy.

Firstly, in each scenario, we mount the tag on a tripod at the same height as four anchors and position it at evenly distributed evaluation points, marked using a commodity laser range finder. As illustrated in Fig. 9, we collect measurements from 116 uniformly distributed points in the lounge, with a consistent 0.8-meter distance between each point. Similarly, we gather data from the laboratory, office, and conference room at 42, 37, and 30 points, respectively, ensuring a uniform 1-meter interval between each point. At every evaluation point, we perform 100 position estimations and statistically analyze the 2D localization errors using all the points, as depicted in Fig. 10. We find the median error to be 6.5cm, 7.6cm, 8.5cm, 5.3cm for lounge, laboratory, office and conference, respectively, which highlights the precision of our ground truth labels across various environment types. Secondly, we execute a specialized experiment at the same 42 distinct positions within our laboratory environment to investigate the effects of human body occlusion on the data collection process. We intentionally obstruct the LoS path between a UWB anchor and tag with a human body, and our findings affirm that the ground truth system maintains its accuracy under such conditions, with a median error of 8.2 cm, as shown in Fig. 10. This can be attributed to the deployment of UWB anchors, which go beyond the standard quantity requirements

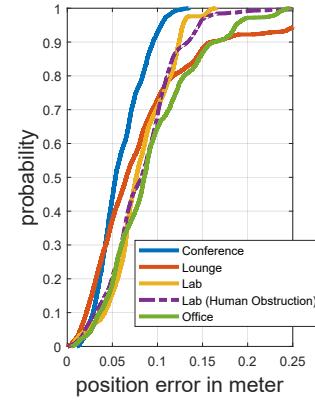


Fig. 10. This figure depicts the accuracy of the ground truth labels collected across four different environments.

for 2D trilateration and ensure clear LoS conditions above other potential obstacles, and the ultra bandwidth of the UWB localization system, which endows it with the ability to resist multipath effects.

## 6 EVALUATION

### 6.1 Experiment Setup

To verify the overall performance of our system, we conduct evaluations using the two datasets mentioned above. We divide these datasets into eight parts based on scenarios for effective training and testing, as shown in Tab. 2. In all experiments, each dataset is split into five equal portions, with four as the training set and the remaining one as the test set. This results in five training and testing combinations for each scenario set. For the human-held device scenario sets, the distinction between the training and testing sets mainly lies in being collected by different users. For the robot-based scenario sets, in the S1/S2 scenarios, the distinction lies in the different trajectories where the robot collected the data. In the S1\*/S2\* scenarios, the difference stems primarily from the combination of four different setups achieved by adjusting the deployment of furniture and reflector. These setups [2] simulate the occasional movement of furniture over time, as typically observed in daily life. To ensure fairness, identical parameter settings are employed for each of the five training runs. In addition, we evaluate the performance of the RLoc model from two main dimensions: in-domain and cross-domain. The in-domain refers to training and testing within the same scenario, while the cross-domain involves training in one or more scenarios and testing in another different scenario.

Table 2. Scenario sets description.

Abbreviation	Scenarios	Space Size	AP Number	Human-held	Description
S1	Atkison Hall	8m × 5m	3	No	Simple LoS based environment for data collected on July 16 and 18, 2019.
S1*	Atkison Hall*	8m × 5m	3	No	Simple LoS based environment with 2 different setups.
S2	Jacobs Hall	18m × 8m	4	No	Complex High-multipath and NLoS environment for data collected on July 28, 2019.
S2*	Jacobs Hall*	18m × 8m	4	No	Complex High-multipath and NLoS environment with 4 different setups.
S3	Conference	8m × 8m	4	Yes	Simple LoS based environment in conference room.
S4	Laboratory	9m × 10m	4	Yes	Simple LoS based environment in laboratory.
S5	Office	9m × 11m	4	Yes	Complex High-multipath environment in office.
S6	Lounge	11m × 14m	4	Yes	Complex High-multipath and NLoS environment in lounge.

### 6.2 Comparison with State-of-the-Art Methods

To evaluate the overall performance of RLoc, we initially select a subset of AoA-based methods, such as SpotFi [26], IAA-based Method [52], and CNN-based Method [35], for their accuracy in angle estimation. We then compare the localization performance of these systems with that of RLoc. Additionally, we reproduce the fingerprint-based method DLoc [2] for further comparison. For the trajectory tracking task, we compare with the filtering algorithm, LOWESS [9] and UAT [42]. The baselines are described as follows:

- SpotFi [26]: SpotFi employs a super-resolution algorithm that jointly estimates the AoA and ToF of each path by combining CSI values across subcarriers and antennas. This approach overcomes the constraints posed by limited antennas and enables highly accurate AoA estimation.
- IAA-based Method [52]: This system introduces a non-parametric iterative approach (IAA) to estimate AoA using sparse recovery and proposes a multiple APs co-localization algorithm to improve the localization performance.
- CNN-based Method [35]: This method introduces a convolutional neural network (CNN) that predicts angular directions using the estimate of the sample covariance matrix, which exhibits improved robustness even in the presence of noise.

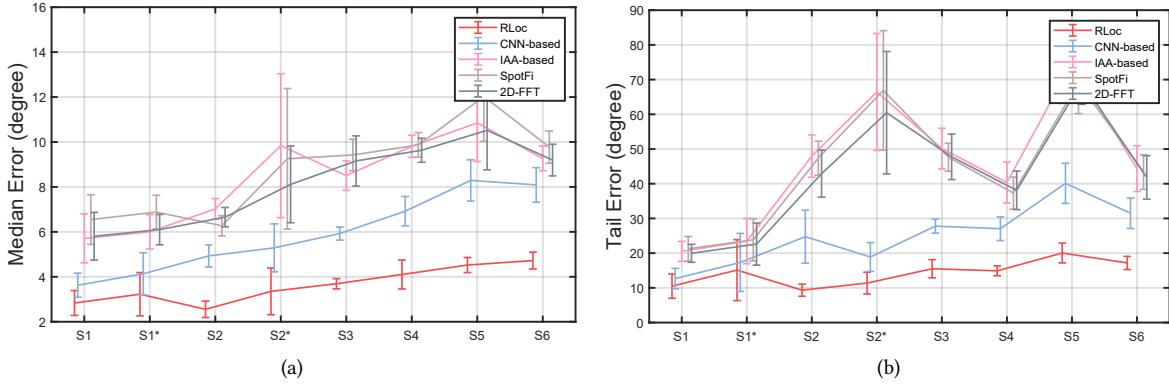


Fig. 11. In-domain evaluation of AoA estimation accuracy across different scenario sets. (a) Comparison of median errors. (b) Comparison of tail errors.

- DLoc [2]: DLoc utilizes deep learning to implicitly model the environment and uses this model to localize the target with sub-meter accuracy in absence of the direct path.
- UAT [42]: Unequal AoA Tracking (UAT) leverages the understanding that direct paths in multiple packets maintain stability across sequential packets. Owing to its dependence on time-domain information, we evaluate and compare its performance based solely on tracking metrics.
- LOWESS [9]: This method proposes a Locally Weighted Scatterplot Smoothing (LOWESS) method to smooth the WiFi-based trajectory.

### 6.3 In-Domain Evaluation

**6.3.1 AoA Estimation Accuracy.** This subsection focuses on evaluating the AoA estimation modules of the RLoc system, which is a critical component for achieving robust and accurate localization. To evaluate the accuracy of the AoA estimation, we employ the median error (represented by the 50th percentile of accuracy), and the tail error (indicated by the 90th percentile of accuracy) as our evaluation metrics. Fig. 11 illustrates the in-domain AoA estimation performance of RLoc, pertaining to training and testing within the same scenario. The results demonstrate that RLoc achieves high performance in terms of angle estimation in different scenarios. Meanwhile, we observe that in the  $S2^*$ , due to the existence of NLoS conditions and additional furniture placed randomly with an added reflector, the angle estimation of existing channel modeling-based methods [26, 52] fails to maintain robustness and exhibits significant variance fluctuations and large errors. In contrast, the deep learning approach employed by RLoc, with its implicit modeling, maintains robustness in such challenging situations, showcasing the advantages of our method in complex environments.

**6.3.2 Localization Accuracy.** To evaluate localization performance, we utilize median and tail error as our evaluation metrics. Fig. 12 illustrates the in-domain localization estimation performance of RLoc, involving training and testing within the same scenario. Our observations indicate that RLoc's framework achieves the best performance across both evaluation metrics. RLoc outperforms the state-of-the-art method [2], achieving an average reduction in median localization error by 36.2% cross eight scenario sets. Regarding the median error of RLoc,  $S3$  shows the best performance with a value of 0.33m, while  $S2^*$  has the worst value of 0.68m. In terms of the tail error,  $S1$  demonstrates the best performance with a value of 0.80m, while  $S2^*$  has the worst performance at 1.52m.

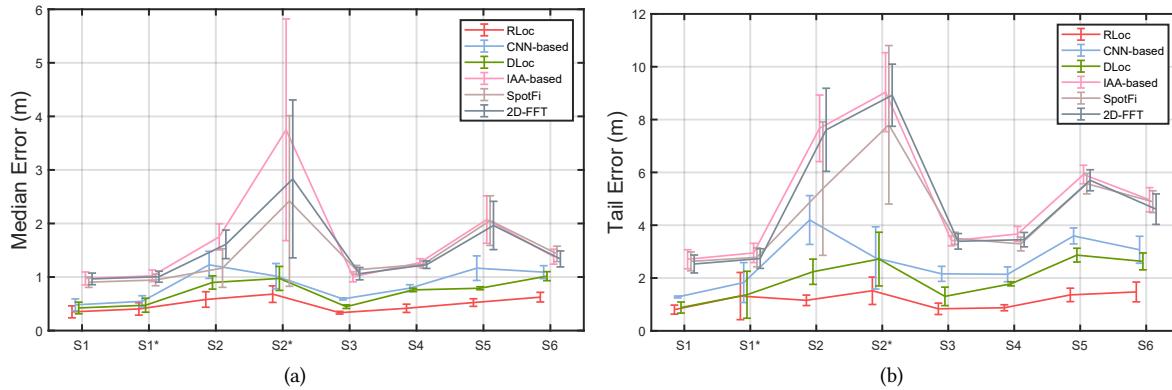


Fig. 12. In-domain evaluation of localization estimation accuracy across different scenarios sets. (a) Comparison of median errors. (b) Comparison of tail errors.

Table 3. In-domain evaluation of tracking performance across various scenarios.

Metrics	Method	S1	S1*	S2	S2*	S3	S4	S5	S6
1-LEO(1m) $\uparrow$	LOWESS	$0.90 \pm 0.05$	$0.83 \pm 0.13$	<b><math>0.83 \pm 0.12</math></b>	<b><math>0.71 \pm 0.15</math></b>	<b><math>0.97 \pm 0.04</math></b>	$0.96 \pm 0.03$	$0.89 \pm 0.03$	$0.88 \pm 0.05$
	UAT	$0.92 \pm 0.03$	$0.85 \pm 0.14$	$0.84 \pm 0.12$	$0.70 \pm 0.16$	$0.96 \pm 0.04$	$0.94 \pm 0.05$	$0.80 \pm 0.08$	$0.86 \pm 0.06$
	RLoc	<b><math>0.93 \pm 0.02</math></b>	<b><math>0.87 \pm 0.14</math></b>	$0.82 \pm 0.13$	<b><math>0.71 \pm 0.16</math></b>	<b><math>0.97 \pm 0.05</math></b>	<b><math>0.97 \pm 0.03</math></b>	<b><math>0.92 \pm 0.03</math></b>	<b><math>0.93 \pm 0.04</math></b>
MAE $\downarrow$	LOWESS	$0.49 \pm 0.13$	$0.63 \pm 0.27$	$0.66 \pm 0.14$	$0.87 \pm 0.25$	$0.36 \pm 0.12$	$0.44 \pm 0.03$	$0.56 \pm 0.07$	$0.64 \pm 0.11$
	UAT	$0.46 \pm 0.11$	$0.61 \pm 0.27$	<b><math>0.63 \pm 0.12</math></b>	$0.86 \pm 0.28$	$0.37 \pm 0.08$	$0.46 \pm 0.05$	$0.69 \pm 0.14$	$0.64 \pm 0.11$
	RLoc	<b><math>0.43 \pm 0.10</math></b>	<b><math>0.57 \pm 0.29</math></b>	$0.66 \pm 0.13$	<b><math>0.83 \pm 0.22</math></b>	<b><math>0.32 \pm 0.11</math></b>	<b><math>0.40 \pm 0.05</math></b>	<b><math>0.42 \pm 0.06</math></b>	<b><math>0.44 \pm 0.09</math></b>

**6.3.3 Tracking Accuracy.** Regarding tracking, relying solely on localization error indicators, which only focus on the accuracy of the position of each point and do not evaluate the time dimension, is not comprehensive. To correct for that, we introduce the Mean Absolute Error (MAE) and the Localization Error Outage [6] (LEO) to measure the tracking accuracy, which can be expressed by:

$$LEO(e_{th}) = \mathbb{P}(\text{error} > e_{th}), \quad (19)$$

where  $\mathbb{P}$  denotes the probability of the outage event  $e < e_{th}$  for one trajectory. Here, the outage event is the localization error exceeds the threshold of the maximum allowable error  $e_{th}$ . We use the equivalent metric  $1 - LEO(e_{th})$ , which denotes the percentage of the total trajectory in which the error is less than  $e_{th}$ . To ensure fairness, UAT, LOWESS and RLoc utilize the same AoA outputs from our network. As demonstrated in Tab. 3, the filtering algorithm of RLoc that takes quantified uncertainty into account exhibits its superiority in terms of overall performance.

**6.3.4 Impact of Uncertainty.** RLoc considers uncertainty as an essential aspect, as it detects unreliable AoA measures and guides us to leverage the diversity of APs and the temporal continuous information of AoAs to achieve robust localization and tracking results. To evaluate the quantified uncertainty, we use the PCC between uncertainty and the AoA error as our evaluation index, which measures the linear correlation between two variables. Furthermore, we conduct ablation experiments on uncertainty-assisted localization, comparing systems with and without uncertainty modeling. As shown in Tab. 4, the quantified uncertainties exhibit a stable correlation with the AoA errors in different scenarios. Moreover, ablation experiments on uncertainty-assisted

Table 4. In-domain evaluation of the quantified uncertainty for different scenarios. To evaluate the impact of uncertainty, ablation experiments are conducted using localization and tracking MAE metrics, with (w) and without (w/o) considering uncertainty.  $Z$  represents the total number of models that RLoc utilizes.

Metrics	Description	S1	S1*	S2	S2*	S3	S4	S5	S6
PCC ↑	Z=1	$0.39 \pm 0.16$	$0.33 \pm 0.14$	$0.48 \pm 0.11$	$0.37 \pm 0.14$	$0.34 \pm 0.07$	$0.37 \pm 0.06$	$0.42 \pm 0.02$	$0.37 \pm 0.02$
	Z=5	<b><math>0.54 \pm 0.13</math></b>	<b><math>0.49 \pm 0.14</math></b>	<b><math>0.58 \pm 0.08</math></b>	<b><math>0.51 \pm 0.13</math></b>	<b><math>0.48 \pm 0.10</math></b>	<b><math>0.50 \pm 0.06</math></b>	<b><math>0.56 \pm 0.01</math></b>	<b><math>0.53 \pm 0.02</math></b>
L-MAE ↓	w/o	$0.52 \pm 0.12$	$0.67 \pm 0.26$	$0.70 \pm 0.13$	$0.90 \pm 0.30$	$0.54 \pm 0.08$	$0.63 \pm 0.05$	$0.91 \pm 0.12$	$0.94 \pm 0.15$
	w( $Z=1$ )	$0.49 \pm 0.12$	$0.71 \pm 0.13$	$0.66 \pm 0.27$	$0.92 \pm 0.31$	$0.51 \pm 0.06$	$0.57 \pm 0.05$	$0.81 \pm 0.11$	$0.92 \pm 0.15$
	w( $Z=5$ )	<b><math>0.47 \pm 0.10</math></b>	<b><math>0.61 \pm 0.28</math></b>	<b><math>0.65 \pm 0.12</math></b>	<b><math>0.84 \pm 0.26</math></b>	<b><math>0.44 \pm 0.07</math></b>	<b><math>0.51 \pm 0.05</math></b>	<b><math>0.71 \pm 0.09</math></b>	<b><math>0.81 \pm 0.15</math></b>
T-MAE ↓	w/o	<b><math>0.43 \pm 0.11</math></b>	$0.59 \pm 0.28$	$0.70 \pm 0.15$	$0.86 \pm 0.21$	$0.33 \pm 0.11$	<b><math>0.40 \pm 0.08</math></b>	$0.48 \pm 0.05$	$0.45 \pm 0.10$
	w( $Z=5$ )	<b><math>0.43 \pm 0.10</math></b>	<b><math>0.57 \pm 0.29</math></b>	<b><math>0.66 \pm 0.13</math></b>	<b><math>0.83 \pm 0.22</math></b>	<b><math>0.32 \pm 0.11</math></b>	<b><math>0.40 \pm 0.05</math></b>	<b><math>0.42 \pm 0.06</math></b>	<b><math>0.44 \pm 0.09</math></b>

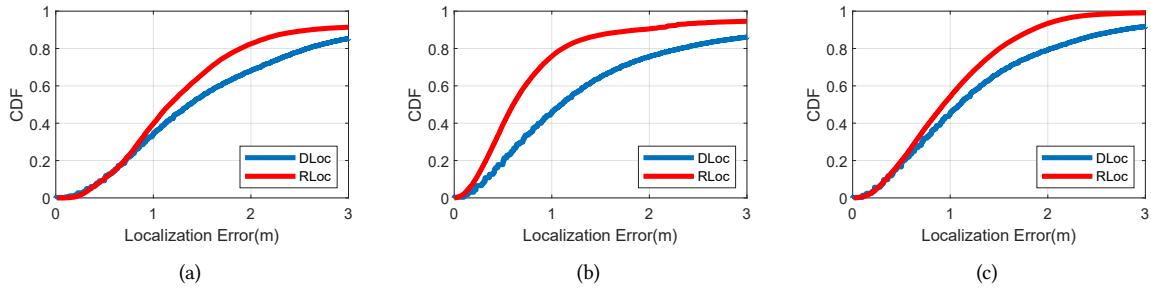


Fig. 13. Evaluation of RLoc ( $Z = 1$ ) and DLoc localization performance under environmental changes using 5-fold cross-validation. (a) Training on datasets labeled as July28, Aug16-1, and Aug16-3, while testing on Aug16-4. (b) Training on datasets labeled as July28, Aug16-1, and Aug16-4, while testing on Aug16-3. (c) Training on datasets labeled as July28, Aug16-3, and Aug16-4, while testing on Aug16-1.

localization also demonstrate its role in enhancing the robustness of the system. In  $S_2$ , we reduce the MAE error by 6.31%, while in  $S_5$ , we achieve the most substantial reduction of 21.64%. In addition, the same ablation experiments and conclusions are applicable to the tracking experiments, which we achieve the most substantial reduction of 12.65% in  $S_5$ .

**6.3.5 Impact of Environmental Changes.** In day-to-day living, the rearrangement of home furnishings is a commonplace occurrence. Recognizing this, we conduct an evaluation to further examine how changes to objects or obstacles, such as the movement of furniture and reflectors, affect our system, RLoc. With the valuable facilitation provided by the robot-based dataset [2], we craft three experimental settings utilizing dataset parts labeled as July28<sup>4</sup> (excluding July28-2), Aug13-1, -3, -4. As shown in Fig. 13, RLoc consistently performs well, proving its ability to handle changes in furniture layout and reflector positions.

#### 6.4 Cross Domain Evaluation

So far, we evaluate the overall performance of RLoc in the in-domain. Nevertheless, it should be noted that this high performance in in-domain does not account for the model’s ability to generalize to new environments. To address this concern, we firstly carry out a cross-domain experiment involving all 20 scenario combinations, which allows us to evaluate the RLoc’s adaptability to different scenarios as detailed in Section §6.4.1. Then we

<sup>4</sup>These labels correspond to specific segments of the dataset and more details can be found at <https://wcsng.ucsd.edu/wild/>.

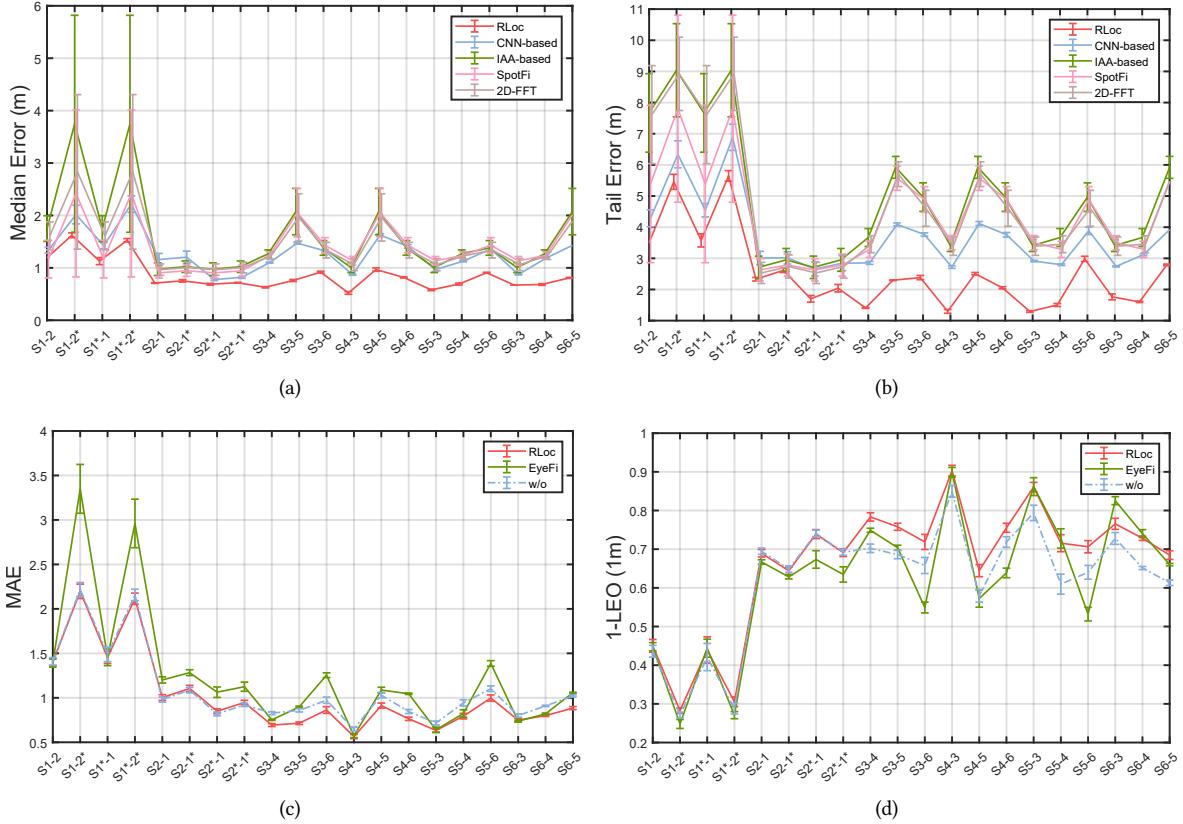


Fig. 14. Cross-scenario evaluation of localization and tracking results. "SA-B" denotes training on Scenario A and testing on Scenario B. **(a)** Comparison of median errors in localization. **(b)** Comparison of tail errors in localization. **(c)** Comparison of MAE in tracking. **(d)** Comparison of LEO in tracking.

investigate the impact of data size on cross-domain performance in Section §6.4.2. Finally, as an exploratory effort, we conduct a cross-domain experiment between human-held and robot-based datasets in Section §6.4.3.

**6.4.1 Cross Scenario Evaluation.** As shown in Fig. 14, we observe that despite a decrease in performance compared to training and testing within the same domain, RLoc continues to demonstrate its robust performance. Importantly, across these scenario combinations, RLoc exhibits an average reduction in median localization error by 20.40%, a notable improvement compared with the state-of-the-art method [35]. It is worth noting that S1 has a relatively simple line-of-sight situation, and its sampling as a training set does not fully cover the feature distribution of S2. As a result, its performance does not reach the decimeter level of localization and tracking accuracy when tested in S2 and S2\*. However, when the training set sampling becomes more diverse, such as in S2\* and other scenarios collected from handheld devices, RLoc's performance can be ensured, demonstrating its adaptability to various environmental conditions. In our evaluations, S4-3 exhibits outstanding performance, with the lowest median localization error at 0.55m, the lowest mean average tracking error at 0.56m, and the highest 1 – LEO(1m) score at 0.90, while S5-3 shows the best performance of tail error with a value of 1.29m.

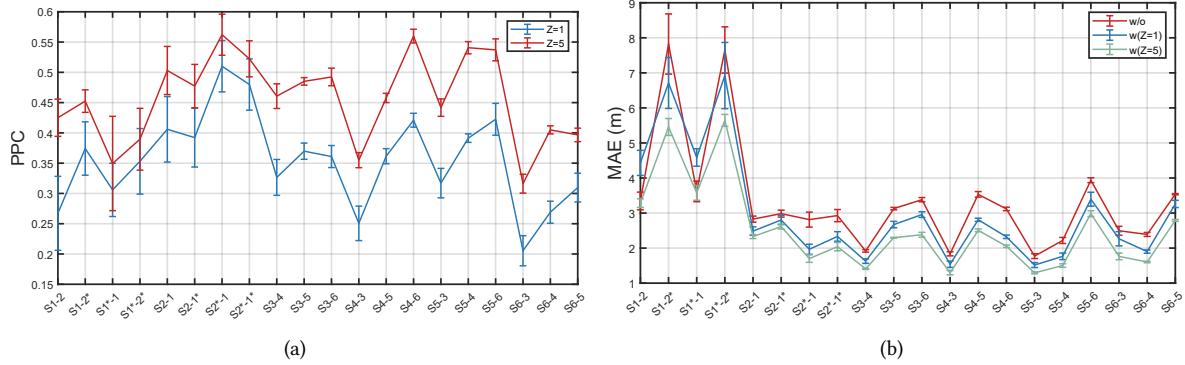


Fig. 15. Evaluation of quantified uncertainty in cross-scenario. (a) Evaluation of PPC between AoA error and quantization uncertainty in different scenarios. (b) Comparative analysis of the localization system’s performance with and without the incorporation of uncertainty.

Our analysis also extends to the impact of uncertainty, as shown in Fig. 15. This shows the impact of quantified uncertainty on cross-domain localization results, revealing a stable correlation between quantified uncertainties and AoA errors across various scenario combinations. Further, ablation experiments demonstrate the role of uncertainty-assisted localization in enhancing system robustness, with scenario combination  $S2^*-S1^*$  witnessing the most significant reduction of the MAE error by 39.54%. Additionally, Fig. 14(c) and Fig. 14(d) present the cross-domain tracking results, demonstrating the robustness of our filtering method that incorporates uncertainty, especially in scenarios involving handheld devices. Our results show that RLoc, once trained, can be effectively applied to new environments and maintain performance, especially when using a comprehensive dataset that better reflects the feature distribution of real-world.

**6.4.2 Impact of Training Data Set.** As mentioned earlier, the training set is a critical component of deep learning methods, which should closely reflect the distribution of real-world features so that the trained model can exhibit better robustness and adaptability to different data scenarios and environmental changes. To examine this, we conduct training in two to three scenarios and perform testing in one or two different scenarios. As illustrated in Fig. 16, the performance of deep learning models trained on datasets from a single scenario may suffer significant errors when tested in another scenario due to the differences in feature distributions between the two scenarios. However, when the training set is expanded to include data from two or three scenarios, the model’s performance becomes more robust, and the error level in testing gradually decreases. The testing results of the model trained on data from three scenarios consistently perform at a relatively good level, indicating improved adaptability to different environments. Alongside these observations, we also conduct tracking experiments, as shown in the Tab. 5. Specifically in  $S3$ , RLoc is able to achieve a remarkable 93% of trajectories with an error within 1m, and the MAE across all scenarios are at the decimeter level. These results imply that in subsequent actual deployment, we can enhance the effectiveness of RLoc by enriching the training set.

**6.4.3 Cross Dataset Evaluation.** A cross-domain evaluation across human-held and robot-based dataset is an important step in validating RLoc’s performance regarding cross-domain adaptability. However, executing such an evaluation presents considerable challenges: (i) *Signal Dimension Mismatch*: Our CSI configuration has a dimensionality of  $M \times N \times K = 1 \times 3 \times 30$ , while the robot-based counterpart employs  $1 \times 4 \times 236$ . (ii) *Center Frequency Variation*: We utilize a center frequency of 5.3 GHz, contrasting with the robot-based dataset, which

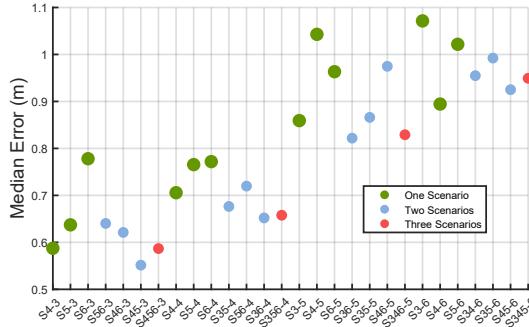


Fig. 16. Localization median error comparison of testing performance with different scenario combinations. The red, blue, and green dots represent the results trained with three scenarios, two scenarios, and one scenario, respectively. All test results are conducted under the condition of  $Z = 1$ .

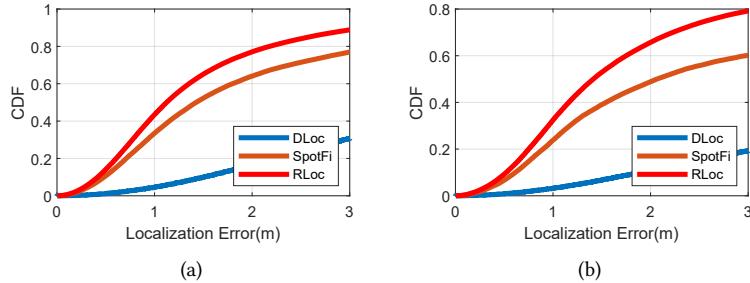


Fig. 17. Evaluation of cross dataset among RLoc ( $Z=1$ ), SpotFi, DLoc using 5-fold cross-validation. (a) Training on datasets labeled as S2, while testing on S6. (b) Training on datasets labeled as S6, while testing on S2.

Table 5. Tracking performance of RLoc trained on data from three scenarios, with and without considering uncertainty.

Metrics	Description	S3	S4	S5	S6
T-MAE ↓	w/o	$0.52 \pm 0.02$	$0.57 \pm 0.01$	$0.75 \pm 0.02$	$0.89 \pm 0.03$
	w( $Z=1$ )	<b><math>0.49 \pm 0.02</math></b>	<b><math>0.55 \pm 0.01</math></b>	<b><math>0.68 \pm 0.02</math></b>	<b><math>0.84 \pm 0.04</math></b>
1-LEO (1m) ↑	w/o	$0.92 \pm 0.01$	$0.87 \pm 0.01$	$0.74 \pm 0.02$	$0.71 \pm 0.02$
	w( $Z=1$ )	<b><math>0.93 \pm 0.01</math></b>	<b><math>0.89 \pm 0.01</math></b>	<b><math>0.79 \pm 0.01</math></b>	<b><math>0.73 \pm 0.02</math></b>

operates close to 5.7 GHz. (iii) *Potential Hardware Interference Noise Due to Different Wi-Fi Chip Manufacturers:* We employ Intel 802.11n 5300 WiFi NICs while the robot-based dataset utilizes the Quantenna 802.11ac WiFi cards. For a fair comparison, we standardize data formats between the two datasets by disregarding one antenna and a 40 MHz bandwidth. As illustrated in Fig. 17, our system's performance does decline across datasets. However, it is noteworthy that it remains more robust compared to both SpotFi and DLoc, even without extra data collections for these scenarios.

Table 6. Average time consumption of RLoc ( $Z = 5$ ) with  $I = 4$  APs in S2. Entry with an asterisk (\*) denotes GPU measurements.

Procedure	Average Time Consumption (ms)
Preprocessing (I=4)	0.85
Neural networks* (Z=5)	7.61
Triangulation (I=4)	2.25
<b>Total</b>	<b>10.71</b>

## 6.5 Latency Evaluation

The computation cost of RLoc mainly comes from three aspects: signal preprocessing, neural network prediction, and triangulation computation. We first evaluate RLoc using a commercial off-the-shelf computer equipped with an Intel Core i7-10700K 3.80 GHz processor and 32 GB RAM. We measure the average computation time of signal preprocessing and localization computation over 1,000 packets using MATLAB R2023a. Subsequently, on the same computer equipped with a NVIDIA GeForce RTX 2080 Ti, we deploy and assess the computation time of  $Z = 5$  trained models, with the average calculated over 1,000 runs. As shown in Tab. 6, the computation time for each signal preprocessing operation is 0.85ms, while each triangulation computation requires 2.25ms. The average computation time of RLoc with  $Z = 5$  models is 7.61ms. These computations lead to a total localization time of 10.71ms, demonstrating the applicability of our system for real-time localization.

## 7 RELATED WORK

### 7.1 Channel-Modeling Indoor Localization

In general, wireless localization schemes model the channel measurements from wireless signals based on geometric parameters, such as distance or direction, to localize the target with respect to one or multiple reference devices using trilateration or triangulation. Early attempts focused on combining Received Signal Strength Indicator (RSSI) measurements with a path loss model, such as the Log-Normal Distance Path Loss (LDPL) [38]. However, power-based ranging suffers from significant performance degradation in complex situations due to insufficient modeling of distance and environment [53]. Fortunately, the availability of finer-grained channel state information (CSI) makes the modeling more accurate, which can be categorized into two types: distance-based [13, 43, 51, 63] and angle-based [12, 26, 39, 52, 62]. Distance-based modeling relies on the phase difference of the subcarrier dimension caused by the time-of-flight (ToF), which is limited by the narrow bandwidth. Although the virtual wide-band can be achieved by channel hopping, these methods pose constraints to communications. In contrast, RLoc only requires CSI measurements from one single channel, similar to most existing angle-based solutions. These solutions model the phase difference between antennas caused by the angle of arrival (AoA) or departure (AoD) without requiring any hardware or firmware changes. Among these approaches [12, 26, 52], the focus is on estimating the direct path angle to subsequently perform triangulation, which limits their robustness performance when some direct paths are heavily attenuated or obstructed. In contrast, RLoc quantifies the uncertainty level arising in the angle estimation task and then exploits the uncertainty to enhance the reliability of localization and tracking. Recent solutions, such as MonoLoco [39] and NLoc [62], exploit multipath information to model the reflection path, thereby achieving more robust performance. However, these methods rely on multiple-input, multiple-output (MIMO) technology, while RLoc offers an effective countermeasure when only a single transmit antenna is available. A recent solution, UAT [42], designs an AoA-based localization system that considers the unequal reliability issue. This method considers the distribution of AoA estimation across sequential packets. Intuitively, the estimation would be more confident if the distribution is concentrated, which means consistent estimation among different packets. In contrast, we investigate how neural networks generate AoA estimations, where the uncertainty arises on the learnable parameters of the neural network. Intuitively, RLoc

is designed to identify instances where the neural network might confidently provide incorrect AoA estimation, thereby mitigating the impact of unreliable angle estimations on localization. Furthermore, since UAT is based on sequential AoA estimations, it requires consecutive CSI measurements for uncertainty estimation, while RLoc is still effective even with only a single packet, as demonstrated in our evaluation results.

## 7.2 Data-Driven Indoor Localization

Another approach to wireless localization is through fingerprint-based methods [2, 4, 31, 46, 47, 55], where landmarks are prelabeled to enable direct localization of the target in space. Fingerprint-based localization systems have been extensively studied, and earlier systems, such as Horus [55], have achieved satisfactory localization performance. With the growing interest in deep learning methods, many fingerprint localization systems based on deep learning techniques have emerged in the field [4, 46, 47, 66]. DLoc [2] stands out among fingerprint systems for its substantial enhancement of their potential. However, the labor-intensive process of fingerprint collection stands as a barrier to the widespread adoption of these approaches. To address this issue, DLoc [2] incorporates an automated mapping platform, thereby mitigating the associated labor costs. Some systems turn to weakly supervised [10, 14, 25, 28, 34, 57] and unsupervised learning [23, 65], which can mitigate the cost of data acquisition by learning from partially labeled or unlabeled data. Other approaches leverage additional information sources [31, 54], such as inertial measurement units (IMU) or user feedback, to obtain more accurate fingerprints. Nevertheless, the necessity for data collection when deployed in new environments still limits their applications. In contrast, RLoc requires data primarily during the initial stages of neural network training. In other words, it becomes possible to readily deploy the trained model in various new scenarios without the need for further data collection. Our results show that RLoc, once trained, can be applied to new environments and maintain performance.

## 8 DISCUSSION

In this section, we briefly discuss some limitations and practical approaches in a more realistic context that have not been directly addressed in this paper.

### 8.1 Manual Initial Phase Calibration

In our experiments, we calibrate all receivers using coaxial cables each time they restart to ensure that the estimation of AoA is not affected by the initial phase of the phase-locked-loop (PLL). However, this calibration process is time-consuming and requires a significant amount of labor, which is a common challenge with angle-based systems [12, 26, 39, 52, 62]. Therefore, automatic calibration remains an important topic for further research.

### 8.2 No Line-of-Sight Localization

In the design of RLoc, the system prioritizes Angle of Arrival (AoA) measurements based on their respective uncertainty levels, in order to cope with instances where some direct paths may be heavily attenuated or obstructed. Notably, while scenarios without a Line-of-Sight (LoS) path are infrequent, they do present limitations for RLoc. Currently, all other decimeter-level WiFi localization systems that utilize Single-Input, Multiple-Output (SIMO) configurations also have this limitation. Nevertheless, the widespread adoption of Multi-Input, Multi-Output (MIMO) communication allows us to extend RLoc from SIMO to MIMO systems. This advancement enhances the use of multipath reflections to localize target devices under no LoS path conditions by extracting additional channel features such as the Angle of Departure (AoD) [39, 62]. We leave this as a topic for future work.

### 8.3 Large-Scale Building Localization

Our RLoc system has been evaluated in various spacious indoor scenarios, such as hall [2], conference room, office, laboratory, lounge. As we plan to extend our work to large-scale, real-time location-based applications encompassing entire buildings, we recognize the unique potential of environments with narrow dimensions, like corridors and staircases. In these areas, one-dimensional localization is often sufficient to address practical needs. Leveraging this, constraints from 2D floor plans can be adapted to suit these unique spatial conditions. Furthermore, one can leverage additional sensor information [21, 48, 49], such as those obtained from accelerometers, gyroscopes, and magnetometers. Exploring multimodal data fusion techniques and analyzing the uncertainty associated with these multimodal data sources emerges as a promising prospect for enhancing the system's accuracy and applicability to ubiquitous indoor environments. We leave this as a topic for future work.

### 8.4 From Experimental to Universal Platforms

Similar to many existing WiFi localization systems [12, 26, 39, 52, 62], our experiments utilize off-the-shelf WiFi 5300 NICs to validate RLoc's efficacy as a prototype for robust localization and tracking. When expanding our considerations to more practical platforms, we encounter an increasing amount of uncertainties. For instance, the orientation of smartphones can dramatically influence the quality of the received signal. This is primarily due to resistive losses encountered on the LoS path between the smartphone and the AP. Such losses may arise from components like the camera, printed circuit board (PCB), and screen glass, located at different positions within the smartphone. However, RLoc's design is built on quantifying and utilizing these uncertainties. Its adaptive nature offers a promising solution to these challenges, making the framework applicable for various environments. We look forward to further exploring our uncertainty modules' performance on more universal platforms, building on our current groundwork.

### 8.5 From Addressing to Preventing Uncertainty

Our results show that RLoc, once trained, can be effectively applied to new environments or datasets while maintaining its performance. However, similar to other learning-based systems [21], there is still a decrease in performance compared to training and testing within the same environment or dataset. This occurs because RLoc is not designed to prevent the generation of such unreliable angle estimates. Instead, it provides a way to mitigate the effect of unreliable angle estimates by prioritizing them based on their uncertainty level. Thus, we plan to leverage the prior constraint from wireless localization knowledge to design unsupervised learning approaches in order to enhance network performance after deployment in new environments or datasets. For instance, we can leverage the triangulation's convergent properties [65], the trajectory continuity of CSI measurements [23], or the other prior information [31] gathered from crowdsourced WiFi signals. Moreover, unifying CSI feature extractions across different wireless modes (e.g., 802.11n/ac/ax) and communication modes (e.g., diversity/multiplexing mode [18]) for upper-layer applications, or utilizing the characteristics of various modes to design data augmentation strategies [61], represents exciting possibilities for more practical work.

## 9 CONCLUSION

In this work, we introduced RLoc, a system designed to combat performance degradation due to environmental changes and to mitigate significant localization errors arising from unreliable AoA estimations. RLoc achieved these goals by quantifying the uncertainty associated with the estimated AoAs and exploiting this uncertainty to enhance the reliability of localization and tracking. Extensive experiments using two datasets collected from diverse indoor settings demonstrated that the proposed approach significantly outperformed state-of-the-art methods. In creating RLoc, our aim was to strike a balance between accuracy and reliability when designing

deep learning-based solutions for general wireless localization applications. Given RLoc's robust advantages, we believe it has the potential to catalyze further advancements in the field of location-based applications.

## ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable comments for improving the quality of the paper. This work was supported by National Key R&D Programmes under Grant 2022YFC0869800 and 2022YFC2503405, National Natural Science Foundation of China under Grant 62201542 and 62172381, fellowship of China Postdoctoral Science Foundation under grant 2022M723069, and the Fundamental Research Funds for the Central Universities.

## REFERENCES

- [1] 2016. IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)* (2016), 1–3534. <https://doi.org/10.1109/IEEESTD.2016.7786995>
- [2] Roshan Ayyalasomayajula, Aditya Arun, Chenfeng Wu, Sanatan Sharma, Abhishek Rajkumar Sethi, Deepak Vasisht, and Dinesh Bharadia. 2020. Deep learning based wireless localization for indoor navigation. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
- [3] Hongkai Chen, Sirajum Munir, and Shan Lin. 2022. RFCam: Uncertainty-aware Fusion of Camera and Wi-Fi for Real-time Human Identification with Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–29.
- [4] Zhenghua Chen, Han Zou, JianFei Yang, Hao Jiang, and Lihua Xie. 2019. WiFi fingerprinting indoor localization using local feature-based deep LSTM. *IEEE Systems Journal* 14, 2 (2019), 3001–3010.
- [5] Nicolai Czink, Markus Herdin, Hüseyin Özcelik, and Ernst Bonek. 2004. Number of multipath clusters in indoor MIMO propagation environments. *Electronics letters* 40, 23 (2004), 1498–1499.
- [6] Davide Dardari, Pau Closas, and Petar M Djurić. 2015. Indoor tracking: Theory, methods, and technologies. *IEEE Transactions on Vehicular Technology* 64, 4 (2015), 1263–1278.
- [7] Decawave. 2023. DW1000 Product Data Sheet. <https://www.qorvo.com/products/d/da007946>
- [8] Vinko Erceg. 2004. TGn channel models. <atftp://ieee:wireless@ftp.802wirelessworld.com/11/03/11-03-0940-04-000n-tgn-channel-models.doc> (2004).
- [9] Shiwei Fang, Tamzeed Islam, Sirajum Munir, and Shahriar Nirjon. 2020. Eyefi: Fast human identification through vision and wifi-based trajectory matching. In *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 59–68.
- [10] Farhad Ghazvinian Zanjani, Ilia Karmanov, Hanno Ackermann, Daniel Dijkman, Simone Merlin, Max Welling, and Fatih Porikli. 2021. Modality-agnostic topology aware localization. *Advances in Neural Information Processing Systems* 34 (2021), 10457–10468.
- [11] Stuart A Golden and Steve S Bateman. 2007. Sensor measurements for Wi-Fi location with emphasis on time-of-arrival ranging. *IEEE Transactions on Mobile Computing* 6, 10 (2007), 1185–1198.
- [12] Wei Gong and Jiangchuan Liu. 2018. RoArray: Towards more robust indoor localization using sparse recovery with commodity WiFi. *IEEE Transactions on Mobile Computing* 18, 6 (2018), 1380–1392.
- [13] Wei Gong and Jiangchuan Liu. 2018. SiFi: Pushing the limit of time-based WiFi localization using a single commodity access point. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–21.
- [14] Baoshen Guo, Weijian Zuo, Shuai Wang, Wenjun Lyu, Zhiqing Hong, Yi Ding, Tian He, and Desheng Zhang. 2022. Wepos: Weak-supervised indoor positioning with unlabeled wifi for on-demand delivery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–25.
- [15] Fredrik K Gustafsson, Martin Danielljan, and Thomas B Schon. 2020. Evaluating scalable bayesian deep learning methods for robust computer vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 318–319.
- [16] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review* 41, 1 (2011), 53–53.
- [17] Simon Haykin. 1985. Array signal processing. *Englewood Cliffs* (1985).
- [18] Yinghui He, Jianwei Liu, Mo Li, Guanding Yu, Jinsong Han, and Kui Ren. 2023. SenCom: Integrated Sensing and Communication with Practical WiFi. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–16.
- [19] Yihui He, Chenchen Zhu, Jianren Wang, Marios Savvides, and Xiangyu Zhang. 2019. Bounding box regression with uncertainty for accurate object detection. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. 2888–2897.
- [20] Dennis E Hinkle, William Wiersma, and Stephen G Jurs. 2003. *Applied statistics for the behavioral sciences*. Vol. 663. Houghton Mifflin college division.

- [21] Yuming Hu, Xiubin Fan, Zhimeng Yin, Feng Qian, Zhe Ji, Yuanchao Shu, Yeqiang Han, Qiang Xu, Jie Liu, and Paramvir Bahl. 2023. The Wisdom of 1,170 Teams: Lessons and Experiences from a Large Indoor Localization Competition. (2023).
- [22] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018. Towards environment independent device free human activity recognition. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 289–304.
- [23] Suk-hoon Jung, Byung-chul Moon, and Dongsoo Han. 2015. Unsupervised learning for crowdsourced indoor localization in wireless networks. *IEEE Transactions on Mobile Computing* 15, 11 (2015), 2892–2906.
- [24] Rudolph Emil Kalman. 1960. A new approach to linear filtering and prediction problems. (1960).
- [25] Ilya Karmanov, Farhad G Zanjani, Ishaque Kadampot, Simone Merlin, and Daniel Dijkman. 2021. Wicluster: Passive indoor 2d/3d positioning using wifi without precise labels. In *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 1–7.
- [26] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadwaj, and Sachin Katti. 2015. Spotfi: Decimeter level localization using wifi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*. 269–282.
- [27] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems* 30 (2017).
- [28] Danyang Li, Jingao Xu, Zheng Yang, Yumeng Lu, Qian Zhang, and Xinglin Zhang. 2021. Train once, locate anytime for anyone: Adversarial learning based wireless localization. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 1–10.
- [29] Xiang Li, Daqing Zhang, Qin Lv, Jie Xiong, Shengjie Li, Yue Zhang, and Hong Mei. 2017. IndoTrack: Device-free indoor human tracking with commodity Wi-Fi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–22.
- [30] Zhang-Meng Liu, Chenwei Zhang, and S Yu Philip. 2018. Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections. *IEEE Transactions on Antennas and Propagation* 66, 12 (2018), 7315–7327.
- [31] Jiazhi Ni, Fusang Zhang, Jie Xiong, Qiang Huang, Zhaoxin Chang, Junqi Ma, BinBin Xie, Pengsen Wang, Guangyu Bian, Xin Li, et al. 2022. Experience: Pushing indoor localization from laboratory to the wild. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 147–157.
- [32] David A Nix and Andreas S Weigend. 1994. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 ieee international conference on neural networks (ICNN'94)*, Vol. 1. IEEE, 55–60.
- [33] Sophocles J Orfanidis. 2002. Electromagnetic waves and antennas. (2002).
- [34] Jeffrey Junfeng Pan, Sinno Jialin Pan, Jie Yin, Lionel M Ni, and Qiang Yang. 2011. Tracking mobile users in wireless networks via semi-supervised colocalization. *IEEE transactions on pattern analysis and machine intelligence* 34, 3 (2011), 587–600.
- [35] Georgios K Papageorgiou, Mathini Sellathurai, and Yonina C Eldar. 2021. Deep networks for direction-of-arrival estimation in low SNR. *IEEE Transactions on Signal Processing* 69 (2021), 3714–3729.
- [36] Joaquin Quinonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. 2008. *Dataset shift in machine learning*. Mit Press.
- [37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. Springer, 234–241.
- [38] Scott Y Seidel and Theodore S Rappaport. 1992. 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE transactions on Antennas and Propagation* 40, 2 (1992), 207–217.
- [39] Elahe Soltanaghaei, Avinash Kalyanaraman, and Kamin Whitehouse. 2018. Multipath triangulation: Decimeter-level wifi localization and orientation with a single unaided receiver. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*. 376–388.
- [40] Navid Tadayon, Muhammed Tahsin Rahman, Shuo Han, Shahrokh Valaei, and Wei Yu. 2019. Decimeter ranging with channel state information. *IEEE Transactions on Wireless Communications* 18, 7 (2019), 3453–3468.
- [41] Natasa Tagasovska and David Lopez-Paz. 2019. Single-model uncertainties for deep learning. *Advances in Neural Information Processing Systems* 32 (2019).
- [42] Tzu-Chun Tai, Kate Ching-Ju Lin, and Yu-Chee Tseng. 2019. Toward reliable localization by unequal AoA tracking. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 444–456.
- [43] Deepak Vasishth, Swaran Kumar, and Dina Katabi. 2016. Decimeter-level localization with a single WiFi access point. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*. 165–178.
- [44] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. 2005. *Algorithmic learning in a random world*. Vol. 29. Springer.
- [45] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking*. 65–76.
- [46] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. 2016. CSI-based fingerprinting for indoor localization: A deep learning approach. *IEEE transactions on vehicular technology* 66, 1 (2016), 763–776.
- [47] Chenshu Wu, Jingao Xu, Zheng Yang, Nicholas D Lane, and Zuwei Yin. 2017. Gain without pain: Accurate WiFi-based localization using fingerprint spatial gradient. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 2 (2017), 1–19.

- [48] Chenshu Wu, Zheng Yang, Zimu Zhou, Kun Qian, Yunhao Liu, and Mingyan Liu. 2015. PhaseU: Real-time LOS identification with WiFi. In *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 2038–2046.
- [49] Chenshu Wu, Feng Zhang, Beibei Wang, and KJ Ray Liu. 2019. EasiTrack: Decimeter-level indoor tracking with graph-based particle filtering. *IEEE Internet of Things Journal* 7, 3 (2019), 2397–2411.
- [50] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2015. Precise power delay profiling with commodity WiFi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 53–64.
- [51] Jie Xiong, Karthikeyan Sundaresan, and Kyle Jamieson. 2015. Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 537–549.
- [52] Shuai Yang, Dongheng Zhang, Ruiyuan Song, Pengfei Yin, and Yan Chen. 2023. Multiple WiFi Access Points Co-Localization Through Joint AoA Estimation. *IEEE Transactions on Mobile Computing* (2023).
- [53] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor localization via channel response. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 1–32.
- [54] Xuehan Ye, Shuo Huang, Yongcai Wang, Wenping Chen, and Deying Li. 2019. Unsupervised Localization by Learning Transition Model. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 65 (jun 2019), 23 pages. <https://doi.org/10.1145/3328936>
- [55] Moustafa Youssef and Ashok Agrawala. 2005. The Horus WLAN location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. 205–218.
- [56] Faheem Safari, Athanasios Gkelias, and Kin K Leung. 2019. A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials* 21, 3 (2019), 2568–2599.
- [57] Farhad G Zanjani, Ilia Karmanov, Hanno Ackermann, Daniel Dijkman, Simone Merlin, Ishaque Kadampot, Brian Buesker, Vamsi Vegunta, and Fatih Porikli. 2022. Deep Learning Frameworks for Weakly-Supervised Indoor Localization. In *NeurIPS 2021 Competitions and Demonstrations Track*. PMLR, 349–354.
- [58] Shuangjiao Zhai, Zhanyong Tang, Petteri Nurmi, Dingyi Fang, Xiaojiang Chen, and Zheng Wang. 2021. RISE: Robust wireless sensing using probabilistic and statistical assessments. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 309–322.
- [59] Dongheng Zhang, Yang Hu, Yan Chen, and Bing Zeng. 2019. Calibrating phase offsets for commodity wifi. *IEEE Systems Journal* 14, 1 (2019), 661–664.
- [60] Daqing Zhang, Dan Wu, Kai Niu, Xuanzhi Wang, Fusang Zhang, Jian Yao, Dajie Jiang, and Fei Qin. 2022. Practical issues and challenges in CSI-based integrated sensing and communication. In *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 836–841.
- [61] Jin Zhang, Fuxiang Wu, Bo Wei, Qieshi Zhang, Hui Huang, Syed W Shah, and Jun Cheng. 2020. Data augmentation and dense-LSTM for human activity recognition using WiFi signal. *IEEE Internet of Things Journal* 8, 6 (2020), 4628–4641.
- [62] Xianan Zhang, Lieke Chen, Mingjie Feng, and Tao Jiang. 2022. Toward reliable non-line-of-sight localization using multipath reflections. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–25.
- [63] Xianan Zhang, Wei Wang, Xuedou Xiao, Hang Yang, Xinyu Zhang, and Tao Jiang. 2020. Peer-to-peer localization for single-antenna devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–25.
- [64] Zengbin Zhang, Xia Zhou, Weile Zhang, Yuanyang Zhang, Gang Wang, Ben Y Zhao, and Haitao Zheng. 2011. I am the antenna: Accurate outdoor ap location using smartphones. In *Proceedings of the 17th annual international conference on Mobile computing and networking*. 109–120.
- [65] Yonghao Zhao, Wai-Choong Wong, Tianyi Feng, and Hari Krishna Garg. 2019. Calibration-free indoor positioning using crowdsourced data and multidimensional scaling. *IEEE Transactions on Wireless Communications* 19, 3 (2019), 1770–1785.
- [66] Zhipeng Zhou, Jihong Yu, Zheng Yang, and Wei Gong. 2020. MobiFi: Fast deep-learning based localization using mobile WiFi. In *Globecom 2020-2020 IEEE Global Communications Conference*. IEEE, 1–6.
- [67] Hongzi Zhu, Yiwei Zhuo, Qinghao Liu, and Shan Chang. 2018.  $\pi$ -splicer: Perceiving accurate CSI phases with commodity WiFi devices. *IEEE Transactions on Mobile Computing* 17, 9 (2018), 2155–2165.