FaHo: Deep Learning Enhanced Holographic Localization for RFID tags

Huatao Xu, Dong Wang, Run Zhao, Qian Zhang*
Shanghai Jiao Tong University, China
{xuhuatao,wangdong,qwert3472}@sjtu.edu.cn,zhaorun@cs.sjtu.edu.cn

ABSTRACT

In recent years, radio frequency identification (RFID)-based approaches have been demonstrated to be a promising indoor localization techniques for many valuable applications, such as tracking tagged objects on the manufacturing lines, locating items in smart warehouses, and so on. In the near future, many applications will gain great benefits from knowing the positions of RFID-tagged objects. However, existing localization approaches often suffer from severe accuracy degradation in real-world environments due to the prevalent environmental interferences, such as the multipath effects. To this end, we designed an RFID-based localization system FaHo, which leverages a deep learning enhanced holographic technique for locating RFID tags accurately even in complex indoor environments. By carefully analyzing the features of the traditional holographic method, we created a new hologram-based algorithm called *joint hologram*, which yields a robust likelihood for each assumed position to be the true tag position. FaHo then adopts a deep convolutional neural network for analyzing the whole hologram, and subsequently estimate the true location of the RFID tag rather than simply seek for the largest-likelihood location. Furthermore, we implemented FaHo and evaluated its performance in several multipath-rich scenarios. The experimental results show that FaHo can achieve centimeter-level accuracy in both the lateral and radial directions using only one moving antenna. More importantly, our work also demonstrates that hologram-based localization is a highly effective technique for RFID indoor localization tasks.

CCS CONCEPTS

• **Networks** → *Location based services*; • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; • **Computer systems organization** → Sensors and actuators.

KEYWORDS

RFID, Localization, Synthetic Aperture, Joint Hologram

ACM Reference Format:
Huatao Xu, Dong Wang, Run Zhao, Qian Zhang. 2019. FaHo: Deep Learning Enhanced Holographic Localization for RFID tags. In *The 17th ACM Conference on Embedded Networked Sensor Systems (SenSys '19), November 10–13, 2019, New York, NY, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3356250.3360035

1 INTRODUCTION

With the widespread deployment of wireless devices, indoor localization has become a key technique in many Internet-of-Things (IoT) applications. As an important building brick of IoT, radio frequency identification (RFID) technology has received significant attention in both academia and industry. RFID-based localization systems could obtain accurate spatial information from many objects with low-cost and easily-deployed RFID tags attached. RFID technologies can help to track inventories as they move along the conveyor belt with several static antennas in next-generation manufacturing, and they can help find the misplaced books or commodities using an automated robot equipped with an RFID reader in smart libraries or supermarkets.

Inspired by these important IoT applications, many state-of-the-art works have proposed many fine-grained localization approaches for accurately estimating the locations of RFID tags. Pioneer works [16, 21, 33, 35] have adopted hyperbolic positioning methods or the angle-of-arrival (AoA) of the received signal-based methods for determining the positions of tagged objects. Later systems [19, 22, 24, 34, 38] have leveraged the concept of the synthetic aperture radar (SAR) or inverse SAR (ISAR) by simulating virtual antenna arrays to locate RFID tags and achieved promising positioning accuracy (*cm*-level and even *mm*-level in some cases).

However, there are inevitable multipath effects generated by nearby reflectors (especially metal objects) in real-world RFID application scenarios, which undermine the received signal metrics (e.g., RSSI and phase) and bring huge challenges for the localization accuracy of RFID localization systems. PinIt [34] employs reference tags and multipath profiles for misplace objects localization. Hologram-based method Tagoram [38] considers the effect of thermal noise and diminishes the impact of extra phase shifts introduced by different RFID tags through a differential advanced method, yet it may still fail to maintain high precision when prevalent multipath reflections exist. Later, MobiTagbot [24] proposes an innovative multipath suppression technique for creating relatively robust holograms with multiple carrier frequency data and obtaining certain accuracy improvement. However, to date, the fundamental challenge for the localization accuracy of the RFID systems is still uncertain environmental interferences, such as ubiquitous multipath effects.

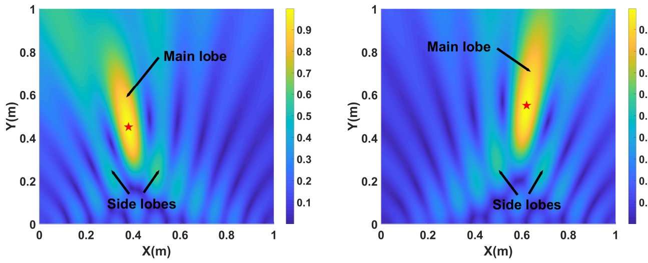
In general, we could combat this challenge and achieve higher precision by increasing the number of deployed or virtual RFID antennas for better robustness. However, this method would increase the cost and difficulty of deployment for RFID localization systems. Hence, we may need a more effective approach. *So how*

---

[*] Dong Wang is the corresponding author.

SenSys '19, November 10–13, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6950-3/19/11…$15.00
https://doi.org/10.1145/3356250.3360035



(a) Example 1 (b) Example 2

Figure 1: Radio frequency holograms. Each pixel value in the hologram represents the likelihood of how its corresponding position is likely to be the true position of the target tag. Yellow values represents larger likelihoods while blue ones represent smaller likelihoods. Red star indicates the true location of the target tag.

can RFID tags be located accurately with minimal resources in indoor environments where prevalent multipath effects exist?

In this paper, we present FaHo, a fine-grained RFID localization technique that combines the synthetic aperture concept and machine learning algorithm to locate RFID tags with high precision using only one antenna, even in multipath-prevalent indoor environments. The key insight is using a machine learning algorithm to suppress the impact of environmental interference and improve the positioning accuracy of the RFID-based holographic technique. To achieve this, we propose a new hologram and an effective hologram-based position estimation method for multipath suppression, which are inspired by several crucial findings from RFID-based holographic localization technique.

As illustrated in Figure 1, we give two examples of holograms generated by the phase measurements of RFID signals using the algorithm in [19]. The first clear finding is that the true position of the RFID tag can yield a larger likelihood than most other positions can. Second, not only the likelihood of the tag position but also its nearby positions could provide relatively larger likelihoods compared with the remaining positions. Third, in addition to the position with the largest corresponding likelihood, the overall distribution changes as the tag position changes. In other words, the distributions of probability in holograms, as well as the largest-likelihood position, could reveal the true position of the target tag. For example, there are always some symmetrical side lobes around the main lobe, and the true position is always located in the central area of the ellipse-like main lobe. These observations show that these holograms contain some common patterns that are highly related to the tag location.

Inspired by the key observations outlined above, we designed FaHo, where the name refers to taking full advantage of holograms. The key idea is to make full use of the valuable probability distribution patterns contained in the generated holograms to mitigate the effects of environmental interference. Based on this pivotal point, we propose several techniques and make the following contributions:

- We propose an original RFID-based holographic localization method, called *joint hologram*, which gives a more robust and effective likelihood for each assumed position by considering the performances of both the assumed position and its nearby positions. In addition, we combine phase profiles of multiple frequencies into a single hologram for further performance improvement;
- We treat holograms like images and construct a deep convolutional neural network (CNN) for analyzing the whole holograms and estimating the accurate position of the target tag. We transform the localization problem to a regression one and create massive low-cost simulation data instead of time-consuming experimental data to train the CNN. Some noise are also added in the simulation data to make our estimation method more robust;
- We implement a prototype of FaHo and systematically evaluate its localization performance under indoor environments. Two case studies, including ordering and reconstruction, are conducted to comprehensively measure FaHo's end-to-end capability in practical application scenarios. The experimental results prove that FaHo can remarkably outperform existing state-of-the-art schemes in terms of localization accuracy.

The rest of our paper is organized as follows: We first present the basic knowledge and preliminary studies in Section 2. The detailed design is described in Section 3, while the implementation is given in Section 4. Section 5 and Section 6 evaluate the comprehensive performance of FaHo. We make discussion in Section 7 and review the related work in Section 8. Finally, we conclude our work in Section 9.

2 BACKGROUND

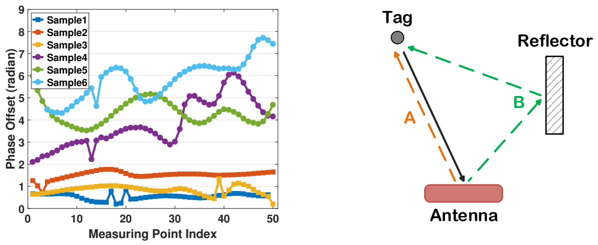
2.1 Challenge

RF Phase. In our paper, we mainly focus on the RF phase, which is a common indicator supported by COTS readers and is relatively sensitive to the signal propagation distance. When an antenna interrogates the UHF RFID tags, the RFID reader can capture the several information indicators of the received backscatter signals. As one of these indicators, the reported phase value can be expressed as [7]:

$$\theta = (2\pi \times \frac{2d}{\lambda} + \phi) \bmod 2\pi \tag{1}$$

where λ is the wavelength and the d is the distance between the reader antenna to the target tag. And ϕ is phase offset introduced by RFID tag, reader characteristics and environmental noise. It is easy to prove that phase value θ is a periodic function, with period 2π and it is related to the frequency of the radio signal.

Empirical Study. We conduct an empirical study for probing into the effects of the environment on the phase offset ϕ. We adopt the same RFID devices (including RFID tags, antenna, and reader) under the same setting in two different scenarios. As shown in Figure 8, the first scenario is an open space, while the second one is filled with various objects (including metal objects, which introduce strong multipath effects). The antenna interrogates RFID tags with known locations at 50 consecutive measuring points and collects about 50 corresponding phase values. All the phase offsets in Figure



(a) Phase offsets comparison

(b) Multipath environment

Figure 2: Impact of multipath refection. (a) Samples $1-3$ are collected in a clean environment, while samples $4-6$ is gathered in a multipath-rich environment.

2(a) equal to unwrapped measured phases minus the corresponding theoretical ones with an extra offset for a better demonstration. We find that the phase offsets in the clean environment are highly stable and approximately equal to a constant. However, there are significant differences between the calculated phase offsets in the multipath-rich environment.

The reason is that the received signal is a superposition of radio signals from both path A and path B when multipath reflection exists as illustrated in Figure 2(b). The signal of propagation path B subsequently causes an extra offset on the received phase value. Therefore, environmental interference such as multipath reflection could have a big impact on received signals, and the uncertain phase offsets bring enormous challenges for accurate RFID-based indoor localization. In this paper, we focus on locating RFID tags in the scenarios where the most received line of sight (LOS) signals are stronger than the non-LOS (NLOS) ones.

Another curse in RFID-based SAR localization technique is that the radial accuracy is much poorer than the lateral one. This can be seen intuitively in Figure 1, in which the resolution in the radial direction (y-axis) is much worse than that in the lateral direction (x-axis). According to our experimental results, existing state-of-the-art works can achieve *cm*-level lateral positioning accuracy but *dm*-level accuracy in the radial direction. Thus, the second challenge is how to reduce the radial localization error and mitigate such accuracy mismatch between the two direction?

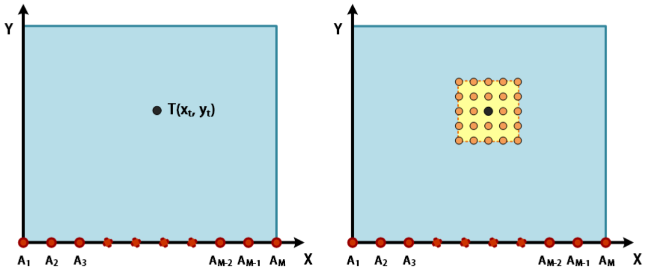
2.2 Convolutional Neural Networks

The convolutional neural network (CNN) is one of the most commonly used deep neural network structures, and is widely used in many applications, including image recognition [29, 31], object detection [5, 23], semantic parsing [6, 17], and many other valuable tasks [11, 40]. Different from fully-connected deep neural networks, CNN applies small kernels to capture the spatial dependencies in images efficiently, which accelerates computation and improves the training efficiency significantly.

3 DESIGN

3.1 Problem Definition

First, we review the basic mechanism of the holographic SAR technique and give the definition of RFID-based localization problem.



(a) SAR illustration

(b) Joint hologram

Figure 3: Illustrations for holographic SAR based localization. Black and red points indicate the position of target tag and virtual antennas respectively. Blue region represents the surveillance plane. (b) The yellow region represents the kernel region of a supposed point and orange points are the centroids of its neighboring girds.

Note that we only consider locating target tags in a two-dimensional plane since our system only adopts one antenna. Suppose an antenna moves along a specific trajectory, which is known as a synthetic aperture, to interrogate the target tag $T(x_t, y_t)$ at different measuring points A_m on multiple frequencies, as illustrated in Figure 3(a). In our paper, we define the positive direction of x and y axis as the lateral and radial direction respectively, while the positive direction of x is the moving direction of the antenna. Let a vector $\boldsymbol{\theta}$ with a length of M denote the phase measurements, where the item $\boldsymbol{\theta}^{(m)}$ represents the received phase collected at the m^{th} measuring point A_m on one frequency. We then define a $P \times Q$ hologram $\boldsymbol{H}$ using the following image:

$$\boldsymbol{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,Q} \\ \vdots & \ddots & \vdots \\ h_{P,1} & \cdots & h_{P,Q} \end{bmatrix} \tag{2}$$

where each pixel $h_{p,q}$ maps to a corresponding grid $S_{p,q}$ in the two-dimensional surveillance plane S. If we use $Z_{p,q}$ to represent the centroid of grid $S_{p,q}$, then the theoretical phase value of $Z_{p,q}$ collected at A_m can be expressed as

$$\boldsymbol{\theta}_{p,q}^{(m)} = \frac{4\pi}{\lambda} \times d(Z_{p,q}, A_m) \tag{3}$$

where $d(Z_{p,q}, A_m)$ stands for the Euclidean distance between $Z_{p,q}$ and A_m. Let $h_{p,q}$ indicate the likelihood that the target tag T is located at $Z_{p,q}$, which can be formulated as:

$$h_{p,q} = \mathcal{L}(\boldsymbol{\theta}\,;\, Z_{p,q}), \; Z_{p,q} \in S \tag{4}$$

where $\mathcal{L}$ is an abstract likelihood function. Therefore, we could formally define two problems in RFID-based holographic localization.

Problem 1. *Given $\boldsymbol{\theta}$, A and S, how to design an effective function $\mathcal{L}$ that can output the maximal value for the nearest grid to the target tag location T and give smaller values for other grids?*

After solving the first problem, we can use $\mathcal{L}$ to generate a hologram $\boldsymbol{H}$, whose element indicates the possibility that its corresponding position is the true position of tag T. We then face the second problem.

PROBLEM 2. *Given* $\boldsymbol{H}$ *with corresponding* S*, how to estimate the target tag position* T*?*

In the remaining part of the design, we will elaborate on the solutions of these two problems in FaHo.

3.2 Joint Hologram

For the first problem of holographic localization for RFID tags, we propose *joint hologram* based on the traditional hologram.

Hologram Comparison. Previous studies [19, 24, 38] have proposed many fine-grained functions that can be formulated in the following equation:

$$h_{p,q} = \mathcal{F}(Z_{p,q},\ \boldsymbol{\theta},\ A) \tag{5}$$

The key insight of these functions is measuring the similarity between measured phases and theoretical ones of each supposed position since they should be similar if the supposed position is the true tag position. However, we also observe that not only the likelihood of tag position but also its nearby positions could get relatively larger likelihoods. Based on this point, we design a joint hologram, whose function $\mathcal{L}$ considers the performance of both $Z_{p,q}$ and its nearby locations when calculating likelihoods:

$$h_{p,q} = \mathcal{F}(Z_{p,q},\ \boldsymbol{\theta},\ A) \cdot \mathcal{G}(\widehat{Z_{p,q}},\ \boldsymbol{\theta}_{p,q},\ A) \tag{6}$$

where $\mathcal{G}$ is a function of $\widehat{Z_{p,q}}$, which represents the centroids of grids around $Z_{p,q}$. As shown in Figure 3(b), all the orange points are centroids included in $\widehat{Z_{p,q}}$, and the yellow region is called *kernel region* in this paper.

Basic Hologram. The definition of function $\mathcal{F}(Z_{p,q},\ \boldsymbol{\theta},\ A)$ is first introduced. We start with a set of functions defined to calculate the likelihood of $Z_{p,q}$:

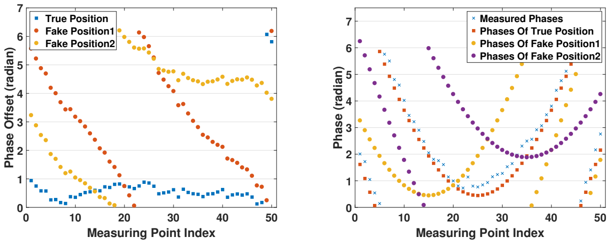
$$\mathcal{H}(\boldsymbol{\vartheta}, Z_{p,q}) = \mathcal{R}(\boldsymbol{\vartheta}, Z_{p,q}) \times \mathcal{W}(\boldsymbol{\vartheta}, Z_{p,q}) \tag{7}$$

$$\mathcal{R}(\boldsymbol{\vartheta}, Z_{p,q}) = \frac{1}{M} \Big| \sum_{m=1}^{M} e^{j(\boldsymbol{\vartheta}^{(m)} - \theta_{p,q}^{(m)})} \Big| \tag{8}$$

$$\mathcal{W}(\boldsymbol{\vartheta}, Z_{p,q}) = \frac{\boldsymbol{\vartheta} \cdot \boldsymbol{\theta}_{p,q}}{\|\boldsymbol{\vartheta}\|_2 \times \left\|\boldsymbol{\theta}_{p,q}\right\|_2} \tag{9}$$

where j denotes the imaginary number and $\boldsymbol{\vartheta}$ indicates either measured or theoretical phase values. We skip the explanation of the former formulas first and directly define a *basic hologram* using following equation:

$$hb_{p,q} = \mathcal{H}(\boldsymbol{\theta}, Z_{p,q}) \tag{10}$$



(a) Phase offsets comparison (b) Phase values comparison

Figure 4: Comparison between the performances of true and fake positions. (a) Phase offsets equal to the remainder of the difference between the measured and theoretical phases divided by 2π. (b) Values of the measured phases, theoretical phases of true and fake positions.

For more intuitive explanation, we compare the performance of the true tag position and several fake positions in Figure 4. As shown in Figure 4(a), since LOS signals received at the most of aperture points are stronger than the NLOS ones, all $\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}_{p,q}^{(m)}$ would be evenly distributed between 0 and 2π if the centroid of a grid is not the actual tag position despite the received phases θ contains the phase offsets introduced by multipath reflections. As a result, all complex signals $e^{j(\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}_{p,q}^{(m)})}$ will have different directions, and $\mathcal{R}(\boldsymbol{\theta}, Z_{p,q})$ is a small value. In contrast, all $\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}_{p,q}^{(m)}$ would be distributed in a limited range if the target tag is located at the grid, resulting in a large $\mathcal{R}(\boldsymbol{\theta}, Z_{p,q})$. As Figure 4(b) shows, if $Z_{p,q}$ is the ground truth, the values of the measured phases and theoretical ones would be similar and have the same trend despite extra phase offsets exist in the measured phases, leading to a larger $\mathcal{W}(\boldsymbol{\vartheta}, Z_{p,q})$. Otherwise, the huge difference between the measured and theoretical phases would cause a small $\mathcal{W}(\boldsymbol{\vartheta}, Z_{p,q})$. Therefore, we could also easily find out which grid cells are most likely to cover the real position of the target tag.

We provide two samples of different holograms in Figure 5. The first is generated by Equation 10 without extra weights $\mathcal{W}(\boldsymbol{\theta}, Z_{p,q})$, and it is called naive hologram in this paper. The largest-likelihood position could indicate the rough position of the ground truth but there are still some localization errors, especially in radial direction. Different from the naive hologram, the resolution of basic hologram has been advanced remarkably. There is another interesting phenomenon in that the hologram seems to be sliced into different blocks. Moreover, the effect of segmentation is pretty outstanding in the radial direction around the ground truth, which results in better radial resolution for basic hologram.

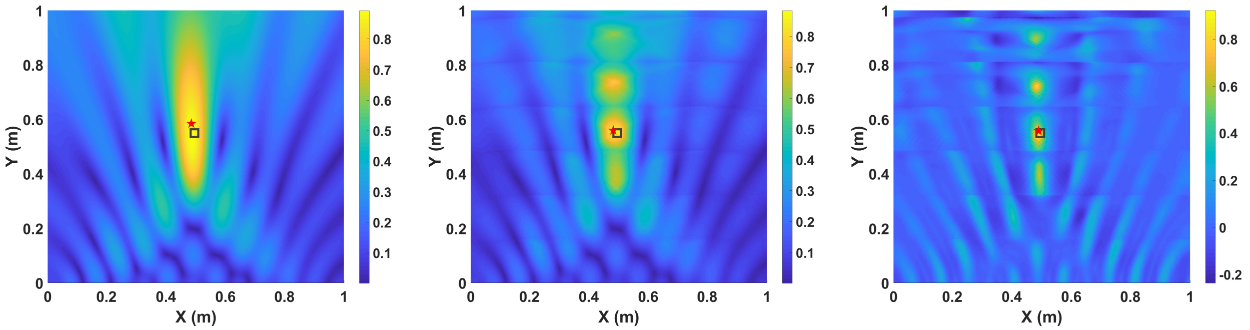
Joint Hologram. Now we give the definition of a more robust likelihood for each grid by utilizing the performance of its neighboring grid cells. Suppose a kernel region centered on $Z_{p,q}$ covers $(2K + 1) \times (2L + 1)$ cells. For instance, K and L are both 2 in Figure 3(b). We use a $(2K + 1) \times (2L + 1)$ matrix $\boldsymbol{CA}_{p,q}$ to denote the real likelihoods of all grids in the kernel region:

$$\boldsymbol{CA}_{p,q} = \begin{bmatrix} \mathcal{H}(\boldsymbol{\theta},\ Z_{p-K,q-L}) & \cdots & \mathcal{H}(\boldsymbol{\theta},\ Z_{p-K,q+L}) \\ \vdots & \ddots & \vdots \\ \mathcal{H}(\boldsymbol{\theta},\ Z_{p+K,q-L}) & \cdots & \mathcal{H}(\boldsymbol{\theta},\ Z_{p+K,q+L}) \end{bmatrix} \tag{11}$$

Then, we define a matrix $\boldsymbol{CT}_{p,q}$ of the same size for representing the theoretical likelihoods of the grids in the kernel region if $Z_{p,q}$ is the true tag location, which is expressed by:

$$\boldsymbol{CT}_{p,q} = \begin{bmatrix} \mathcal{H}(\boldsymbol{\theta}_{p,q},\ Z_{p-K,q-L}) & \cdots & \mathcal{H}(\boldsymbol{\theta}_{p,q},\ Z_{p-K,q+L}) \\ \vdots & \ddots & \vdots \\ \mathcal{H}(\boldsymbol{\theta}_{p,q},\ Z_{p+K,q-L}) & \cdots & \mathcal{H}(\boldsymbol{\theta}_{p,q},\ Z_{p+K,q+L}) \end{bmatrix} \tag{12}$$

Note that $\boldsymbol{CA}_{p,q}$ and $\boldsymbol{CT}_{p,q}$ are different because the input of $\mathcal{H}$ in $\boldsymbol{CA}_{p,q}$ compromises measured phases, whereas that in $\boldsymbol{CT}_{p,q}$ involves theoretical phases. Next, we treat $\boldsymbol{CA}_{p,q}$ and $\boldsymbol{CT}_{p,q}$ as two random variables, and each element of $\boldsymbol{CA}_{p,q}$ corresponds to



(a) Naive hologram

(b) Basic hologram

(c) Joint hologram

Figure 5: Comparison of different holograms. Red stars indicate largest-likelihood location, while gray squares stand for the real position of tag. (a) Naive hologram generated using Equation 10 without $\mathcal{W}$, (b) basic hologram. (c) joint hologram with a kernel region of $11cm \times 11cm$.

the element of $\boldsymbol{CT}_{p,q}$ with the same corresponding grid. We then define joint hologram using $\boldsymbol{CA}_{p,q}$ and $\boldsymbol{CT}_{p,q}$:

$$h_{p,q} = \rho(\boldsymbol{CA}_{p,q}, \boldsymbol{CT}_{p,q}) \times hb_{p,q} \tag{13}$$

where $\rho(\boldsymbol{CA}_{p,q}, \boldsymbol{CT}_{p,q})$ is the Pearson correlation coefficient [2] between $\boldsymbol{CA}_{p,q}$ and $\boldsymbol{CT}_{p,q}$, which is a value ranging between -1 and 1, measuring the strength of the linear relationship between two random variables. Similar to the former explanation, if $Z_{p,q}$ is the true tag position, the correlation coefficient and $hb_{p,q}$ would theoretically have a maximum 1 since there is a positive correlation between $\boldsymbol{CA}_{p,q}$ and $\boldsymbol{CT}_{p,q}$. Conversely, they would be relatively smaller values when the position of the target tag and $Z_{p,q}$ are far apart, leading to a much smaller $h_{p,q}$. Note that, for the grids near the edges that don't have enough neighbor grids, we adopt kernel regions with dynamically adjusted sizes.

Different from the traditional holograms, each likelihood in the joint hologram utilizes both the similarity between the theoretical and measured phase values of its corresponding grid and the similarity between theoretical and actual likelihoods of all grids in the kernel region. In other words, not only the performance of the assumed position is considered in the joint hologram, but also the performance of its neighbor positions. Figure 5(c) depicts an example of a joint hologram. It is obvious that the resolution in the joint hologram has been further improved, especially in lateral direction compared to basic hologram shown in Figure 5.

Finally, we can extend joint hologram to a multi-frequency one based one for better robustness [24] using following equation:

$$h'_{p,q} = \frac{1}{F}\sum_{f=1}^{F} h^{(f)}_{p,q} \tag{14}$$

where $h^{(f)}_{p,q}$ means the likelihood of $Z_{p,q}$ calculated by f^{th} frequency phase measurements.

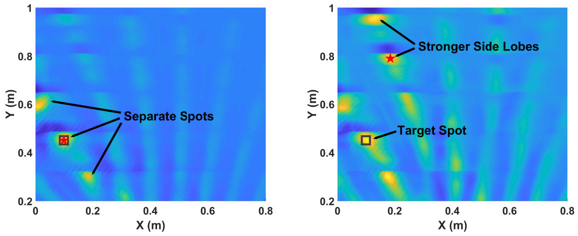
Kernel Region. Intuitively, the performance of joint hologram would boost as the kernel region enlarges because each likelihood takes into account more information about its neighbor grids. But it turned out to be incorrect in practice. In fact, the real goal of joint hologram is to optimize the likelihood of the grids around the ground truth and make sure that the true grid can obtain a higher likelihood than others. For those grids around the ground truth, larger kernel region would cover more unstable likelihoods when strong environmental disturbances exist, leading to unstable weights for $h_{p,q}$. Therefore, a larger kernel region might cause a counterproductive outcome. Based on the experiment results in Section 5, we prove that a smaller kernel region is more suitable for a multipath-rich environment.

Computation Optimization. The computational complexity of calculating likelihoods in a basic hologram is $O(M)$, while the computational complexity in a joint hologram is $O(MKL)$. The time cost (minute-level calculating time for a $1m^2$ surveillance plane with 100×100 grids and 5×5 kernel region) of creating a joint hologram is unacceptable. Fortunately, we find most of the time is spent on calculating the theoretical phase values and likelihoods in the kernel region for each grid, which could be reused to reduce time cost. Although there may exist signal loss, which means the number of phase measurements is smaller than we expected, it only occurs at a few aperture points according to our experimental results. Therefore, we store the theoretical phase profiles and likelihoods in the kernel regions in a hash table and reuse them to generate holograms for a new tag. This is a classic optimization method that reduces the time complexity by sacrificing space complexity.

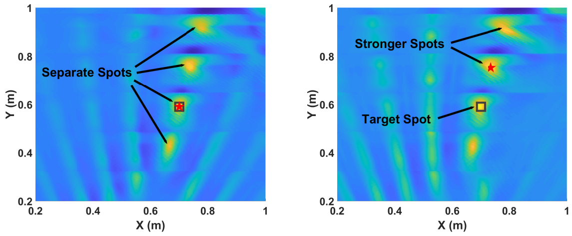
3.3 Position Estimation

Next, we introduce the solution for the second problem on how to estimate the target tag position based on generated holograms.

Estimation Comparison. Previous research [19, 24, 38] has selected the position with the largest likelihood as the true position, which is the typical maximum likelihood estimation [37] and only focuses on the position with the largest likelihood. It utilizes little information in the hologram and does not consider how other positions perform that may be susceptible to environmental interference because the likelihood of a single position is probably unreliable. We could find that there are still some estimate errors in Figure 5(c).



(a) Ideal hologram 1 (b) Real hologram 1



(c) Ideal hologram 2 (d) Real hologram 2

Figure 6: Comparison between ideal holograms and real ones. Ideal holograms are generated from simulation phases with no noise, while real holograms are generated from experimental data. Red stars indicate largest-likelihood locations, while black squares stand for the true positions.

Based on the aforementioned analysis in Section 1, we could design a more effective estimate method by fully utilizing valuable patterns in holograms. Inspired by applications of deep neural networks in computer vision, we treat holograms as images and construct a CNN to analyze the whole hologram for accurate position estimation.

CNN Functionality. By carefully observing the likelihood distribution in the joint hologram, we find two common patterns under our experimental setting, which can be seen intuitively in Figures 5(c) and 6:

- The unique main lobe surrounded by the symmetrical side lobes is partitioned into three or more strong spots;
- Only the spot second closest to the synthetic aperture covers the true tag location.

Meanwhile, although the joint hologram yields optimized likelihoods, the extra phase offsets introduced by multipath reflections can still affect the likelihood distribution in holograms. As shown in Figure 6, we observe two possible effects:

- *Side lobe disturbance*. Compared with ideal hologram 1, the side lobes in real hologram 1 are stronger;
- *Main lobe ambiguity*. Other spots in the main lobe may have larger likelihoods than the target spot does.

To tackle these problems, we could train the CNN and let it learn to recognize the main lobe and locate the spot second closest to the synthetic aperture by learning the common patterns. For example, after learning the first pattern, CNN can recognize the main lobe, and the first effect *side lobe disturbance* can be suppressed. CNN can also handle the issue of *main lobe ambiguity* after learning the second pattern. In fact, the functionality of CNN in FaHo is similar to classical tasks in object detection (e.g. recognize human faces and locate them in images) and extensive literature [5, 23] have proven that CNNs can achieve super performances in these tasks.

Neural Network Structure. As illustrated in Figure 7, we design a 12-layer deep neural network. The estimate problem is transformed into a regression problem in FaHo. Therefore, the CNN is trained to predict the exact coordinates of the target tag instead of the grid where the tag is located. In FaHo, the input layer is an image-like fixed-size joint hologram generated by multiple-frequency data using Equation 14. Constructed upon input layer are three convolution layers, which adopts sets of learnable filters to extract the high-level representation of the input data. Their kernel sizes are $5 \times 5 \times 16$, $5 \times 5 \times 32$ and $3 \times 3 \times 16$, with the same stride of 1 in both the vertical and horizontal directions. Zero paddings are added in the convolutional layers for maintaining the dimensions of output. Rectified Linear Unit (ReLU) layer is built upon each convolution layer, which provides fast and effective training for the network and is widely used in CNN[14]. Different from ReLU layers, which don't change the size of the input, the Max-pooling layers with the size of 3×3 kernel and strides of 2, 2, and 1, significantly simplify the connections to the following layers.

A dropout layer is constructed upon the last convolutional layer to help prevent overfitting [13]. In the training stage, each unit in the dropout [30] layer will be dropped with probability $1 - p$ for the network simplification. At the test times, the outgoing weights of this unit will be multiplied by p. After dropout layer, there are four fully-connected layers, which connect all neurons to those of the previous layer. There is also ReLU layer between each two fully-connected layers. The last layer contains two units, which stand for estimated lateral and radial coordinates.

Training Method. The training objective of the neural network is to reduce the error between the estimated coordinates and the ground truth. Since we regard the estimation position problem as a regression one, mean square error (MSE) is leveraged to define the loss function:

$$loss = \frac{1}{N} \sum_{i=0}^{N} \|\hat{T}_i - T_i\|^2 \tag{15}$$

where N is the number of mini-batch, while $\hat{T}$ and T represent the estimated position and true position respectively. In the training stage, we adopt Adam [12] optimization algorithm for parameters updates.

Instead of conducting a large number of experiments for gathering training dataset, we make it using simulation phase measurements. To make our model obtain good robustness and tolerate noise well, we add random Gaussian phase offsets to the theoretical phases of each target tag, whose position is randomly selected from a fixed surveillance plane. To mimic the effects of environmental noise as much as possible, we gather a small amount of experimental data and calculate the variance of phase offsets collected at each experiment. We then refer to that variance and determine the parameters of the Gaussian noise distribution. Moreover, the model is selected through a small validation dataset, which is a subset of the experimental dataset. The selected model is eventually tested on the remaining experimental data.

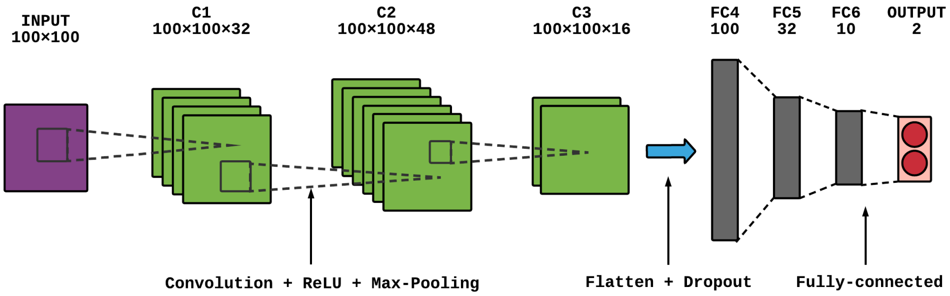


Figure 7: Deep neural network structure in FaHo.

4 IMPLEMENTATION

Hardware. The prototype is implemented with COTS UHF RFID devices, including a set of Alien[1] ALN9662 passive RFID tags, one Laird[2] S9028PCR RFID antenna, and an ImpinJ[3] Speedway R420 RFID reader. The reader works at a frequency band $920.625 - 924.375MHz$ and is programmed to hop over eight channels with a step of $0.5MHz$. The antenna is fixed at a linear guide moving at a speed of $0.05m/s$ with a length of $1m$. The reader connects to a PC equipped with Intel(R) $i5 - 8500$ CPU, 16 GB of RAM and an NVIDIA GeForce GTX1060 6GB GPU.

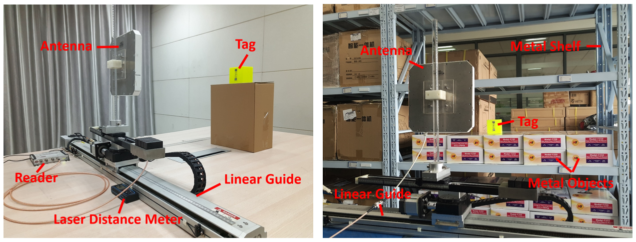
Software. There are two components in the software of FaHo. The data collection component is implemented in C#, which communicates with the RFID reader through a low-level reader protocol (LLRP) and controls the movement of the linear guide simultaneously. The data analysis component implemented in Python processes the raw data and estimates the positions of scanned tags. The CNN in the data analysis component is constructed on Tensorflow [1] 1.9.0 with cuDNN.

5 EVALUATION

In this section, we make an exhaustive comparison between FaHo and three state-of-the-art works, followed by several benchmarks presented to provide insights into the effects of different factors on the performance of FaHo.

5.1 Experimental methodology

Scenario. We conduct experiments in two different indoor environments, as shown in Figure 8. The first experimental environment is a laboratory, which is a multipath-low scenario, since there are few objects nearby generating multipaths. In contrast, the second experiment is conducted at an RFID laboratory where there are many metal objects, representing a classic multipath-prevalent scenario. In addition, the antenna mounted at the linear guide interrogates tags at 50 aperture points. We adopt the same RFID devices with identical setting in these two scenarios and collect about 100 samples for each scenario.



(a) Scenario 1 (b) Scenario 2

Figure 8: Experiment scenarios. (a) Multipath-low scenario. (b) Multipath-rich scenario.

Dataset. As mentioned before, we adopt simulation data for training our neural network. As in the experimental setting, we simulate phase measurements for 50 aperture points distributed uniformly in a $1-m$ synthetic aperture. Moreover, random Gaussian noises added to all the points have identical averages and standard deviations in each simulation experiment. The standard deviation of noise for each experiment is also stochastic, and the average standard deviation of all experiments is 0.5. In contrast to previous multiple-frequency model, we use a single-frequency simulation holograms. And four frequencies are selected to simulate the phase profiles. The position of target tag is also randomly selected in the $1m \times 1m$ surveillance plane, with a precision of $0.001m$. In total, we create a training dataset with 10, 000 samples (100×100 holograms).

Validation and test dataset are both holograms generated by the real data collected in the former environments. The validation dataset is a randomly selected subset of the initial experimental dataset, and the sample size is 10. The remaining samples belong to test dataset. The datasets of two scenarios are independent of each other, and we train two different models for them since the effect of environmental interference is different in each scenario.

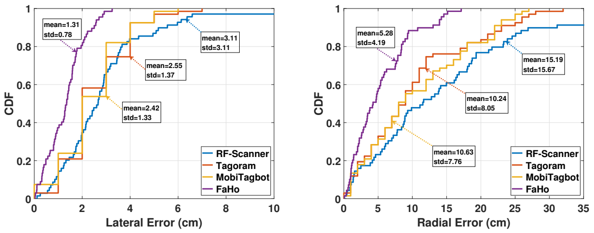
Metric. In the evaluation part, we mainly focus on *localization error* or *positioning error*, which is the absolute value of distance difference between estimated and true position. There two types of error, namely, *lateral error* and *radial error*, denoted as $e_x = ||\hat{p_x} - p_x||$ and $e_y = ||\hat{p_y} - p_y||$ respectively.

5.2 Accuracy among different schemes

We also implement RF-Scanner [15], Tagoram [38] (DAH algorithm), and MobiTagbot [24], which are fine-grained RFID localization

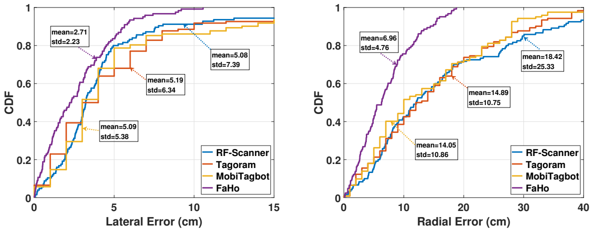
---

[1] https://www.alientechnology.com/
[2] https://connectivity.lairdtech.com/
[3] https://www.impinj.com/



(a) Lateral accuracy comparison (b) Radial accuracy comparison

Figure 9: Accuracy comparison in scenario 1.



(a) Lateral accuracy comparison (b) Radial accuracy comparison

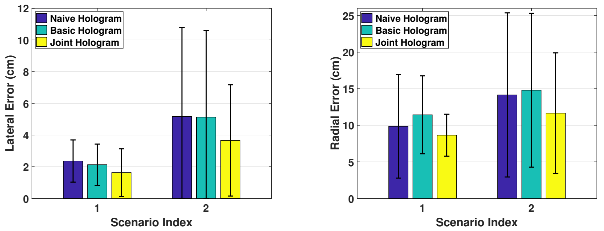
Figure 10: Accuracy comparison in scenario 2.

schemes. RF-Scanner [15] derives the tag position by curve fitting the hyperbola pattern on the phase profiles while DAH [38] and MobiTagbot [24] are both holographic SAR schemes. Since RF-Scanner [15] and DAH [38] are single-frequency localization models, we regard multiple frequency data as multiple virtual antennas and extend them to multiple-frequency models. We run these three algorithms and FaHo on the same datasets for fair comparisons.

It is worth mentioning that Tagoram and MobiTagbot can achieve *cm*-level precision since we adopt *cm*-level discretization holograms. However, RF-Scanner and FaHo can achieve higher precision because the localization problem is transformed into a parameter estimation or regression problem. Therefore, there are apparent zigzag patterns in the cumulative distributed function (CDF) of Tagoram and MobiTagbot's localization errors, as shown in Figure 9 and 10.

Figure 9 plots the results of the positioning accuracy in the first scenario. Previous three schemes achieve 90th-percentile lateral errors of 5.40*cm*, 4.00*cm*, 4.00*cm*, and 90th-percentile radial errors of 29.44*cm*, 23.00*cm*, and 22.00*cm* in the first scenario. At the same time, FaHo has a 90th-percentile error of 2.48*cm* in the lateral direction, outperforming RF-Scanner, Tagoram and MobiTagbot by 2.17×, 1.61×, and 1.61× respectively and a 90th-percentile error of 12.13*cm* in the radial direction, which is 17.31*cm*, 10.87*cm*, and 9.87*cm* lower than others, respectively. These results clearly demonstrate that FaHo successfully achieves better performance in the multipath-low scenario.

Next, we analyze the localization accuracies of the four schemes in the second scenario. As illustrated in Figure 10, RF-Scanner achieves 90th-percentile errors of 7.78*cm* and 36.32*cm* in the two directions. Tagoram considers the thermal noises and achieves average errors of 10.00*cm* and 32.30*cm*, but these accuracies are far lower than the results reported in [38]. The reasons for this include the presence of prevalent multipath effects and reduction in the number of aperture points. MobiTagbot obtains some accuracy improvements through a novel multipath suppression algorithm and has similar accuracy values to Tagoram, at 13.30*cm* and 28.00*cm* respectively. However, FaHo achieves a 90th percentile lateral error of 5.64*cm*, which is smaller than the errors of the other schemes. Moreover, the radial average error of FaHo is 14.53*cm*, outperforming previous schemes by 2.50×, 2.22×, and 1.92× respectively. This improvement is due to its ability to handle the disturbance caused by the multipath reflections.



(a) Lateral accuracy comparison (b) Radial accuracy comparison

Figure 11: Accuracy comparison among different holograms. The error bar specifies the average error and the standard deviation.

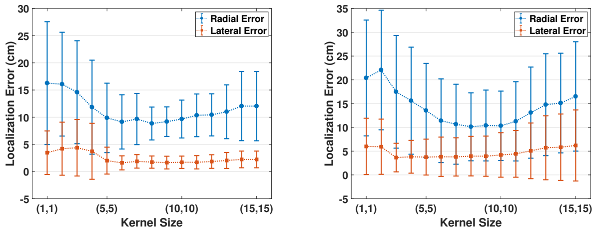
The average lateral and radial errors of FaHo are 1.31*cm* and 5.28*cm*, respectively in scenario 1. In scenario 2, FaHo achieves positioning accuracies of 2.71*cm* and 6.96*cm* in two directions. Based on these results, we find that FaHo can achieve *cm*-level accuracies in both the lateral and radial directions, which significantly weakens the accuracy mismatch. Comparing the positioning accuracies of FaHo in two scenarios, we also find that the differences of accuracy are relatively smaller (1.40*cm* and 1.68*cm* in the lateral and radial direction) compared with the accuracy differences of other schemes. This demonstrates that FaHo is more insensitive to environmental interference.

It is obvious that the localization error standard deviation of FaHo is much smaller than those of the previous three schemes in these cases, which also demonstrates the stable performance of FaHo. In summary, our experimental results show that FaHo is superior to the other schemes in the two scenarios.

5.3 Accuracy among different holograms

In this part, we compare the localization accuracy among three holograms, namely, the naive, basic, and joint holograms. It should be noted that we estimate the tag position by directly selecting the largest-likelihood position for all the hologram in this evaluation.

As shown in Figure 11(a), the joint hologram achieves lateral average errors of 1.63*cm* and 3.66*cm* in the two scenarios, outperforming the naive hologram by 1.44× and 1.30× and basic hologram by 1.30× and 1.40×. For the radial errors illustrated in Figure 11(b), the joint hologram also obtains higher radial localization accuracy than the other two holograms in both two scenarios based on the



(a) Accuracy in scenario 1 **(b) Accuracy in scenario 2**

Figure 12: Impact of kernel region size. The error bar specifies the average error and standard deviation.

maximum likelihood estimation method. The accuracy improvement in scenario 2 is more outstanding since the joint hologram is more capable of suppressing multipath effects. Moreover, the standard deviations of the joint hologram are smaller than those of the other two in all cases, which shows that the joint hologram has better robustness. Thus, it is clear that the joint hologram achieves the best performance based on the maximum likelihood estimation method. The reason is that joint hologram gives a more robust likelihood for each grid by taking full advantages of the performances of its neighbor grids.

There is a seemingly abnormal phenomenon that the basic hologram achieves lower positioning accuracy than the naive hologram does in the radial direction under two scenarios. In our opinion, it cannot prove that the basic hologram is less effective than the naive hologram is. Based on the previous analysis, the basic hologram seems to be sliced into several slots, and the largest likelihood is likely located at the neighboring slot of ground truth. Hence, the maximum likelihood estimation may introduce an extra localization error in the radial direction. Therefore, we think that maximum likelihood estimation is not suitable for basic hologram. And we show how our convolutional neural network solution outperforms the maximum likelihood estimation later.

5.4 Accuracy among different kernel region

Next, the impact of kernel region size in the joint hologram is investigated. We define a tuple (K, L) to represent the values of K and L in Equation 11. Figure 12 depicts the relationship between the localization error and size of the kernel region in two scenarios. It is worth mentioning that we only consider square area kernel region when comparing different region sizes here. The final positioning accuracies in the two scenarios show that the overall localization error decreases and then increases as we continuously enlarge the kernel region. In other words, either a too large or too small region will reduce the positioning accuracy. Hence, we need a trade off for selecting a suitable kernel size.

Another finding is that the best-fit kernel regions in the two scenarios are varied due to the different environmental interferences. Therefore, we fully utilize the validation dataset for comparing the positioning accuracy in the combined direction of different kernel regions and select the best-fit kernel region for the two scenarios. In our experimental setting, the kernel regions are $(10, 7)$ and $(5, 4)$ in two scenarios, which are equivalent to $21cm \times 15cm$ and $11cm \times 9cm$ kernel region in our experimental setting. Based on our results, a relatively smaller kernel region (e.g., from $(4, 4)$ to $(9, 9)$) is more suitable for a multipath-rich environment.

Table 1: Accuracy definition

Element	Value
a_1	N: naive hologram B: basic hologram J: joint hologram
a_2	M: maximum likelihood estimation C: convolutional neural network method
a_3	X: lateral direction Y: radial direction

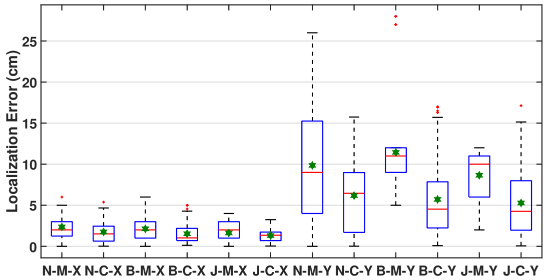
5.5 Accuracy among different position estimation methods

In this evaluation, we make a comprehensive comparison between two hologram-based position estimation methods, namely, maximum likelihood estimation method and our deep learning method. To better understand their performance difference, we apply the two methods on the three holograms used above, including the naive, basic and joint holograms. We define a symbol a_1-a_2-a_3 to represent the positioning accuracy obtained by different method combinations. The first two elements indicate the type of hologram and position estimation method, while the last element represents the direction of localization accuracy. Table 1 provides the possible values of each letter. For instance, *N-C-X* means the lateral accuracy obtained using our convolutional neural network for analyzing the naive hologram and subsequently estimating the tag position.

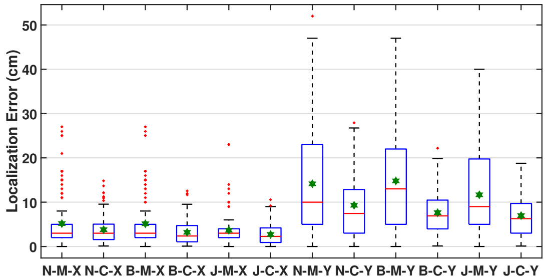
Figure 13 depicts the positioning accuracies of varied method combinations in two scenarios. We can compare the performance of maximum likelihood estimation with our deep convolutional neural network solution through two adjacent cases. The results in Figure 13 show that the average positioning accuracies, as well as their distribution of our deep learning estimation method, are better than those of maximum likelihood estimation in most cases. Furthermore, the accuracy improvements in the radial direction are outstanding.

We also find that the deep learning method remarkably outperforms traditional maximum likelihood estimation, especially in the second scenario, which is a classic multipath-rich environment. For joint hologram, deep learning method achieves median errors of $2.27cm$ and $6.28cm$ in two direction separately, outperforming maximum likelihood estimation by $1.32\times$ and $1.43\times$, respectively. Moreover, it is clear that the outliers are reduced significantly if the CNN estimation method is adopted. Such improvement is due to the full utilization of likelihood distribution in the hologram and the robustness of our estimation method since we have adopted noise-added dataset to train it. These results prove that estimating the tag position by merely relying on the largest likelihood is not a robust estimation method in multipath-rich environments.

Significantly, we find that the radial errors of the basic hologram are smaller than those of the naive hologram, which is different



(a) Accuracy comparison in scenario 1



(b) Accuracy comparison in scenario 2

Figure 13: Accuracy comparison of estimation methods. The green asterisk indicates the average error in each case.

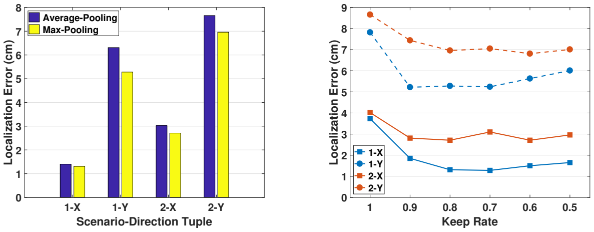
from the former result obtained by maximum likelihood estimation. Therefore, our experimental results prove that the hologram-based localization method has the potential for more accurate indoor localization, while the maximum likelihood estimation obviously fails to live up to such potential.

5.6 Impact of neural network component

Here, we investigate the performance of different neural network components, including pooling, dropout layer, and normalization. We tried different pooling methods, and the results are presented in Figure 14(a). We chose max-pooling in our deep learning model since it achieves relatively better performance. As shown in Figure 14(b), different keep rates in dropout layer are also compared. We set keep rate to 0.8 in FaHo. Furthermore, we also tried to add several Batch Normalization [8] layers to our model. Unfortunately, the model cannot converge after a Batch Normalization layer is adopted, so it is highly unsuitable for our localization task. The possible reason is that it may change the original likelihood distributions in holograms; accordingly, we have not used it in our model.

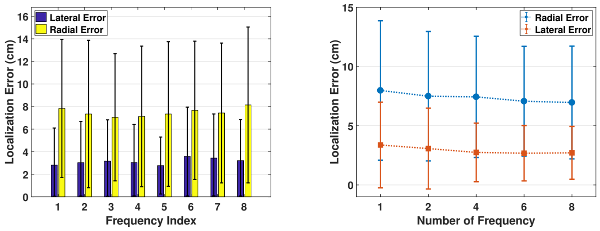
5.7 Impact of frequency

Since we utilize multiple-frequency data, it is necessary to check the performance of FaHo on different frequencies (channels). In this evaluation, we take the dataset gathered at the second scenario as an example. We use single frequency phase profiles to generate joint holograms and then apply our CNN for the position estimation.



(a) Pooling (b) Dropout

Figure 14: Impact of different neural network components. The notation 1-X and 2-Y represent the lateral accuracy in scenario 1 and radial accuracy in scenario 2, respectively. The same scheme are applied to the other tuples.



(a) Accuracy comparison among different frequency (b) Accuracy comparison among number of utilized frequency

Figure 15: Impact of frequency. The error bar specifies the average error and standard deviation.

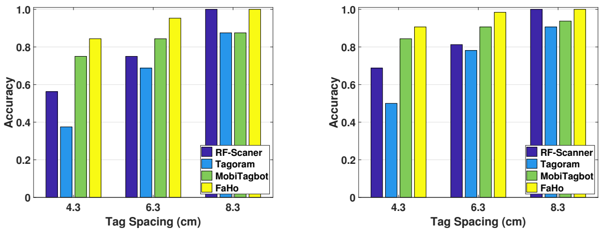
Figure 15(a) depicts the accuracies for eight frequencies. It is obvious that there is only a slight difference between them. Therefore, FaHo is irrelevant to frequency and can support frequency hopping without accuracy loss. Next, the effect of the number of frequencies is investigated. We randomly choose one, two, four, six channels out of eight channels and utilize their data to create holograms. As Figure 15(b) shows, the average localization error decreases substantially as the number of utilized frequencies increases. Moreover, the standard deviation reduces stably at the same time, which also proves that we can advance the robustness of localization model by combining multiple-frequency data into a single model.

5.8 Time consumption

Finally, we evaluate the execution time of the two processes, including hologram creation and position estimation. Table 2 presents the time consumption of different holograms generated from eight-frequency data. Time consumption includes the time reading the raw phase data, data preprocessing, and likelihood calculation. It is easy to find that the time consumed for each hologram is reduced significantly with optimization. In addition, the CNN needs little time (about 50*ms* for 500 holograms) to predict tag position after creating holograms. To summarize, it takes about 1.5*s* to estimate the tag position through raw phase data, which shows that FaHo can achieve good realtime performance.

Table 2: Time consumption of different holograms (unit: s). BH and JH-(5,5) represent the basic hologram and joint hologram with $K = 5$ and $L = 5$.

Model	BH	JH-(5, 5)	JH-(10, 10)	JH-(15, 15)
w.o. optimization	0.78	72.38	247.30	503.06
w. optimization	0.74	1.28	1.36	1.44



(a) Hard accuracy comparison (b) Soft accuracy comparison

Figure 16: Ordering tagged objects.

6 CASE STUDY

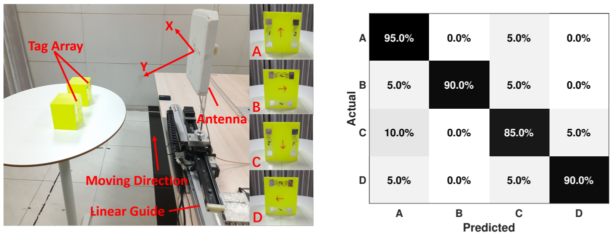
In this section, we investigate the performance of FaHo in two RFID-based localization application scenarios, including ordering and reconstruction.

6.1 Ordering

Shelf scanning is an important process for inventory management in a smart warehouse or library; it can also help find the misplaced objects. Moreover, it is a classic application scenario for RFID-based localization systems [15, 24, 26, 27, 38]. Therefore, we probe into FaHo's performance for ordering RFID tags along the lateral direction. We conduct ordering experiments in the multipath-prevalent scenario in Figure 8. In each experiment, we scan eight tags with spacings varying from 4.3*cm* to 8.3*cm* and obtain over 120 orders of tags. The other experimental configurations are the same as those in Section 5.

In this paper, we define two metrics of ordering accuracy, named as *hard ordering accuracy* and *soft ordering accuracy*. Hard ordering accuracy is the most common metric used in the previous studies [15, 24, 26, 27], which is defined as the ratio of number of correctly ordered tags to the total number of tags. However, the hard accuracy is easily affected by outliers, so we define soft ordering accuracy. Suppose we have a correct sort sequence s_c and a predicted one s_p. Soft ordering accuracy is the ratio of the length of the longest common subsequence between s_c and s_p to the length of s_c. For example, if the correct sequence is *1-2-3-4-5* and the predicted one is *1-3-4-2-5*, the corresponding hard ordering accuracy is 0.4 while the soft one is 0.8.

Figure 16 provides the ordering accuracies of RF-Scanner [15], Tagoram [38], MobiTagbot [24], and FaHo. As the results show, FaHo identifies the right orders with success rates of 84.4%, 95.3%, and 100.0% on hard ordering accuracy, outperforming the other approaches in most cases. Moreover, FaHo achieves soft ordering accuracies of 90.6%, 98.4% and 100.0%, which are also basically higher than the accuracies of others. Our experimental results clearly demonstrate that FaHo is superior to the other schemes in ordering tagged objects.



(a) Reconstruction scenario (b) Recognition accuracy

Figure 17: Reconstruct tagged objects. (a) On the right are four placements of each cube based on our assumption. (b) the classification accuracy for different placements.

Table 3: Position feature comparison.

Placement	Lateral Position	Radial Position
A	$x_2 - x_1 \geq \Delta x$	$\lvert y_1 - y_2 \rvert < \Delta y$
B	$\lvert x_1 - x_2 \rvert < \Delta x$	$y_2 - y_1 \geq \Delta y$
C	$x_1 - x_2 \geq \Delta x$	$\lvert y_1 - y_2 \rvert < \Delta y$
D	$\lvert x_1 - x_2 \rvert < \Delta x$	$y_1 - y_2 \geq \Delta y$

6.2 Reconstruction

Recognizing how tagged objects are placed is another important application for RFID-based localization systems [3, 36]. These systems can determine the package placement in smart warehouses [3] or make real-time observations for orientation-sensitive cargoes in intelligent factories [36]. In this case study, we design an RFID-based reconstruction method for determining the placement of tagged objects using the location information of RFID tag arrays.

As shown in Figure 17(a), two tags with a spacing of 8.3*cm* are attached on the surface of a $10 \times 10 \times 10cm^3$ cube. The rest of experimental setting is the same as those in Section 5. We assume the tagged surfaces of the cubes always face the antenna. Hence, there are four types of placements as illustrated in the Figure 17(a). The horizontal distance between them is 45*cm*, and the antenna is 40*cm* higher than the plane where cubes are placed is. We summarize the position features of the tag array for different placements in Table 3. Here, (x_1, y_1) and (x_2, y_2) are the coordinates of the left and right tag, respectively in placement A. For placements A and D, there are significant differences in the lateral coordinates of two tags, while the radial coordinates are similar. In the 3D space, two tags are on different XoY planes in placement B and D, leading to some differences in radial distance. Δx and Δy are two thresholds utilized to tolerate some localization errors, which are set to 4*cm* and 3*cm*. Hence, we can discriminate the placement of cubes by analyzing the difference between the coordinates of the tag array. We collect 20 experiment samples for each case.

As illustrated in Figure 17(b), the confusion matrix implies the high classification accuracy of FaHo that only few cubes are erroneously classified into the wrong placement. This high average accuracy of 90.00% is attributed to the high localization accuracy in the lateral and radial directions. Meanwhile, the overall classification accuracies of RF-Scanner, Tagoram, and MobiTagbot are 56.25%, 75.00%, 75.00%, respectively. Therefore, FaHo also remarkably outperforms other schemes in this case study.

7 DISCUSSION

In FaHo, we present a joint hologram and a new hologram-based position estimation method. We believe the basic idea of the joint hologram that utilizing the performances of the assumed position and its neighboring positions could be extended to other holographic SAR techniques (e.g., generating angle-based hologram). Moreover, it is considered that estimating target indicators by analyzing the whole hologram could be extended to other holographic methods. However, there are several limitations and future works that could be done to extend and refine our research.

Antenna Motion. In our experiments, the antenna is programmed to scan tags at 50 static aperture points along a linear trajectory, which introduces frequent acceleration or deceleration and causes continual antenna vibration. Hence, the antenna is designed to spends 2 seconds interrogating tags at each point per frequency to collect enough data. We use the average phase measurement to represent the phase value at the corresponding aperture points for weakening the impact of the vibration. This approach significantly increases the consumed time in the experiments.

Environmental Dependence. FaHo currently requires a validation set for selecting the size of the kernel region and determine the CNN model, which increases its deployment difficulty. Moreover, the final model may suffer accuracy degradation confronting environmental changes. To tackle this problem, we plan to design a component to automatically adjust kernel size before we generate joint holograms. As shown in Figure 4(b), the theoretical unwrapped phase values received at aperture points are parabola-like. If there are strong multipath effects, the phase offsets will introduce large deviations to this parabola-like pattern. Thus, we can use this feature to measure the multipath effects (a simple way is to do curve-fitting on the unwrapped raw phases and calculate the deviation between the fitted and raw phases). According to the previous analysis that larger kernel region is more suitable for a multipath-low environment described in Section 3.2, we can adjust the size of kernel region in a limited range by the calculated deviation.

8 RELATED WORK

In this section, we make a review for existing state-of-the-art RFID localization works, which can be classified into two categories.

RSSI-based Schemes. Early works adopted the received signal strength indicator (RSSI) which is an indication of the signal power level. LANDMARC [20] utilizes the concept of reference tags to improve the overall accuracy of localization for RFID tags whereas OTrack [26] establishes a probabilistic model for recognizing the transient critical region and proposes a specific protocol to monitor the order of tags. Frogeye [39] tackles the issue caused by the weak stability of strength via the Mixture of Gaussian Model.

Phase-based Schemes. In recent years, there is a growing interest in utilizing phase profiles of received signals for estimating the precise location of RFID tags. The reason is that compared to RSSI, phase is more sensitive to the distance changes and has higher precision. Pioneer works [21, 34, 35] have utilized the phase difference collected at several measuring points to estimate the angle of arrival (AoA) for the received signal. Similar as AoA based approaches, geometry-based localization techniques [15, 16, 33] and proximity-based methods [34] were also proposed.

Later fine-grained researches [19, 22, 24, 25, 34, 38] applied SAR for localization. By utilizing the relative motion between RFID antenna and tags, we could simulate multiple virtual antennas to extract more spatial information of RFID tags. Former original works [24, 38] constructed different holographies, which reveal the likelihoods of all possible locations where the target tag may be at, to locate RFID tags and achieve high precision. Tagoram [38] introduces an advanced hologram DAH by considering the impact of thermal noise. MobiTagbot [24] further proposes a more robust holography by leveraging multiple channels. Our work is inspired by above holography imaging based localization schemes but advanced in two ways. We design a more robust hologram to suppress the influences of environmental interferences and propose an innovative position estimation method for accurate position prediction based on the generated hologram. Different from RFind [18] which leverages the underlying physical properties to emulate a large bandwidth and uses it for accurate localization, FaHo could also work on single frequency mode and achieve good performance.

Furthermore, there are also lots of valuable RFID based tracking works [4, 9, 10, 28, 32] recently. Pantomime [28] recognizes the user's gestures by replacing the conventional multiple antennas single tag tracking framework with an equivalent multiple tags single antenna system. OmniTrack [9] proposes an orientation-aware localization model to explicitly quantify the respective impact of the read-tag distance and the orientation of tag.

9 CONCLUSION

In this paper, we presented a fine-grained RFID-based localization system, which a fusion approach of the holographic technique and machine learning algorithm. FaHo utilizes a robust self-designed hologram and an effective position estimate method to suppress multipath-rich effects and achieve high positioning accuracy. We implemented the prototype of FaHo and conducted extensive experiments for its performance evaluations. The experimental results showed that FaHo can achieve good positioning accuracy with strong robustness (*cm*-level accuracies in both the lateral and radial directions). Furthermore, our approach proves that holographic method is a highly effective for RFID-based indoor localization. Furthermore, our approach proves that holographic method is highly effective for RFID-based indoor localization.

ACKNOWLEDGMENTS

We sincerely thank our shepherd and anonymous reviewers for their helpful and valuable suggestions. This work is supported by Artificial Intelligence Institute, Shanghai Jiao Tong University.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 265–283.

[2] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.

[3] Yanling Bu, Lei Xie, Jia Liu, Bingbing He, Yinyin Gong, and Sanglu Lu. 2017. 3-Dimensional Reconstruction on Tagged Packages via RFID Systems. In *Sensing, Communication, and Networking (SECON), 2017 14th Annual IEEE International Conference on*. IEEE, 1–9.

[4] Liqiong Chang, Jie Xiong, Ju Wang, Xiaojiang Chen, Yu Wang, Zhanyong Tang, and Dingyi Fang. 2018. RF-Copybook: A Millimeter Level Calligraphy Copybook based on commodity RFID. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 128.

[5] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

[6] Seunghoon Hong, Junhyuk Oh, Honglak Lee, and Bohyung Han. 2016. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3204–3212.

[7] J Impin. 2010. Speedway revolution reader application note: Low level user data support. *Speedway Revolution Reader Application Note* (2010).

[8] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).

[9] Chengkun Jiang, Yuan He, Xiaolong Zheng, and Yunhao Liu. 2018. Orientation-aware RFID Tracking with Centimeter-level Accuracy. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN '18)*. IEEE Press, Piscataway, NJ, USA, 290–301. https://doi.org/10.1109/IPSN.2018.00057

[10] Haojian Jin, Zhijian Yang, Swarun Kumar, and Jason I Hong. 2018. Towards Wearable Everyday Body-Frame Tracking using Passive RFIDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 145.

[11] Andrej Karpathy and Li Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3128–3137.

[12] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *Computer Science* (2014).

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

[15] Jia Liu, Feng Zhu, Yanyan Wang, Xia Wang, Qingfeng Pan, and Lijun Chen. 2017. RF-Scanner: Shelf scanning with robot-assisted RFID systems. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1–9.

[16] Tianci Liu, Lei Yang, Qiongzheng Lin, Yi Guo, and Yunhao Liu. 2014. Anchor-free backscatter positioning for RFID tags with high accuracy. In *IEEE International Conference on Computer Communications, 2014 Proceedings IEEE*. 379–387.

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.

[18] Yunfei Ma, Nicholas Selby, and Fadel Adib. 2017. Minding the billions: Ultra-wideband localization for deployed rfid tags. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 248–260.

[19] Robert Miesen, Fabian Kirsch, and Martin Vossiek. 2011. Holographic localization of passive UHF RFID transponders. In *IEEE International Conference on Rfid*. 32–37.

[20] Lionel M. Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P. Patil. 2004. LANDMARC: indoor location sensing using active RFID. *Wireless Networks* 10, 6 (2004), 701–710.

[21] P. V. Nikitin, R. Martinez, S. Ramamurthy, and H. Leland. 2010. Phase based spatial identification of UHF RFID tags. In *IEEE International Conference on Rfid*. 102–109.

[22] A. Parr, R. Miesen, and M. Vossiek. 2013. Inverse SAR approach for localization of moving RFID tags. In *IEEE International Conference on Rfid*. 104–109.

[23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.

[24] Longfei Shangguan and Kyle Jamieson. 2016. The Design and Implementation of a Mobile RFID Tag Sorting Robot. In *International Conference on Mobile Systems*.

[25] Longfei Shangguan and Kyle Jamieson. 2016. Leveraging electromagnetic polarization in a two-antenna whiteboard in the air. In *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*. ACM, 443–456.

[26] Longfei Shangguan, Zhenjiang Li, Zheng Yang, and Mo Li. 2013. OTrack: Order tracking for luggage in mobile RFID systems. In *IEEE International Conference on Computer Communications, 2013 Proceedings IEEE*. 3066–3074.

[27] Longfei Shangguan, Zheng Yang, Alex X Liu, Zimu Zhou, and Yunhao Liu. 2015. Relative localization of RFID tags using spatial-temporal phase profiling. In *Usenix Conference on Networked Systems Design and Implementation*.

[28] Longfei Shangguan, Zimu Zhou, and Kyle Jamieson. 2017. Enabling gesture-based interactions with objects. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 239–251.

[29] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[30] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.

[31] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[32] Chuyu Wang, Jian Liu, Yingying Chen, Lei Xie, Hong Bo Liu, and Sanclu Lu. 2018. RF-Kinect: A Wearable RFID-based Approach Towards 3D Body Movement Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 41.

[33] Jue Wang, Fadel Adib, Ross Knepper, Dina Katabi, and Daniela Rus. 2013. RF-compass:robot object manipulation using RFIDs. In *International Conference on Mobile Computing & NETWORKING*. 3–14.

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

[35] Jue Wang, Deepak Vasisht, and Dina Katabi. 2014. RF-IDraw: Virtual touch screen in the air using RF signals. In *Acm Conference on Sigcomm*.

[36] Teng Wei and Xinyu Zhang. 2016. Gyro in the air: tracking 3D orientation of batteryless internet-of-things. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 55–68.

[37] Halbert White. 1982. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society* (1982), 1–25.

[38] Lei Yang, Yekui Chen, Xiang Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: real-time tracking of mobile RFID tags to high precision using COTS devices. In *International Conference on Mobile Computing & Networking*.

[39] Lei Yang, Yong Qi, Jianbing Fang, Xuan Ding, Tianci Liu, and Mo Li. 2014. Frogeye: Perception of the slightest tag motion. In *IEEE International Conference on Computer Communications, 2014 Proceedings IEEE*. 2670–2678.

[40] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. ACM, 267–281.