Environment Adaptive RFID-Based 3D Human Pose Tracking With a Meta-Learning Approach

Chao Yang, *Student Member, IEEE*, Lingxiao Wang, Xuyu Wang, *Member, IEEE*, and Shiwen Mao, *Fellow, IEEE*

***Abstract*—With the development of Radio-Frequency (RF) sensing techniques, RF based 3D human pose estimation has attracted increasing interest recently. Unlike video camera based techniques, RF sensing has the unique strength of preserving user privacy. However, due to the complex wireless channels indoors, a well-trained RF sensing system is usually hard to generalize to new environments. In this paper, we propose an environment adaptive solution for Radio-Frequency Identification (RFID) based 3D human skeleton tracking systems. We first analyze the challenges in environment adaptation for RFID based sensing systems. Following the analysis, we then propose a meta-learning approach for RFID-based 3D human pose tracking, termed *Meta-Pose*. The system is implemented with off-the-shelf RFID devices and can well adapt to new environments with few-shot fine-tuning, thus greatly simplifying the deployment of the trained system. We conduct extensive experiments in different indoor scenarios to validate the high adaptability and accuracy of the Meta-Pose system.**

***Index Terms*—3D human pose tracking, few-shot fine-tuning, generalization, meta-learning, RFID sensing.**

I. Introduction

HUMAN pose tracking has attracted great interest in recent years, because it is useful for numerous applications such as human-computer interaction, video surveillance, and somatosensory games. The advances in human pose tracking have been mainly driven by the new developments in computer vision, from two-dimensional (2D) systems [1] to three-dimensional (3D) realtime systems [2]. However, the vision-based techniques often raise concerns of security and privacy. For example, many wireless security cameras are easily hacked by malicious users [3]. The collected video data for pose tracking could also be illegally intercepted. Several radio frequency (RF) sensing schemes have been proposed to address the privacy concern in human pose tracking, using various RF sensing techniques such as Frequency-Modulated Continuous Wave (FMCW) radar [4], millimeter wave (mmWave) radar [5], WiFi [6], [7], and RFID [8]–[10]. Compared with vision-based techniques, RF sensing-based human pose tracking requires neither sufficient lighting, nor a line-of-sight path between the subject and camera (i.e., capable of getting around obstacles or even through walls), and more important, the privacy of users can be better protected.

In RF sensing-based systems, deep learning has been widely adopted to translate captured RF data to human poses. However, such techniques usually have the generalization problem, i.e., the inference performance usually degrades greatly when applying a well-trained deep learning model to a new, un-trained environment. Since RF signals propagate in the open air, the received RF signal is usually highly sensitive to the specifics of the deployed environment, such as the placement of the antennas, the layout (e.g., walls) of the room, the obstacles in the surroundings, and the movement of objects and/or subjects nearby. When there are variations in the environment, the same human subject performing the same activity could generate considerably different RF features. It has become a great challenge to develop human pose tracking schemes that are adaptive to the environment.

Researchers have proposed several solutions to address the environment adaptation challenge. The most straightforward approach is to simply increase the size and variety of the training dataset, i.e., to train the deep learning model with vast amounts of data measured in many types of environments. When applying the trained model to a new environment, it is likely that new environment will be similar to an environment that exists in the training dataset, and thus the inference performance would not degrade much. However, this approach requires considerable efforts and incurs high costs in collecting large amounts of training data. In addition, more sophisticated schemes leverage the idea of *adversarial learning* to improve feature extraction [11], [12]. Rather than training using data collected from numerous RF environments, adversarial learning incorporates a *domain discriminator* to distinguish features from different environments (i.e., domains). When the model is trained such that it is capable of fooling the discriminator, the features that are common to all the domains will be extracted. The domain adversarial network can effectively reduce the requirement on training data.

Alternatively, when applying the well-trained (or, pre-trained) model to a new, unknown domain, we can fine-tune the model by further training it with new data collected from the new environment, such that the pretrained model can better capture the specific features of the new domain. The

Manuscript received 1 September 2021; revised 12 November 2021; accepted 1 January 2022. Date of publication 6 January 2022; date of current version 21 July 2022. This work was supported in part by NSF through the Wireless Engineering Research and Education Center at Auburn University under Grant ECCS-1923163 and Grant CNS-2107190. This work was presented in part at IEEE GLOBECOM 2021, Madrid, Spain, December 2021. *(Corresponding author: Shiwen Mao.)*

Chao Yang, Lingxiao Wang, and Shiwen Mao are with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: czy0017@auburn.edu; lzw0039@auburn.edu; smao@ieee.org).

Xuyu Wang is with the Department of Computer Science, California State University, Sacramento, CA 95819 USA (e-mail: xuyu.wang@csus.edu).

Digital Object Identifier 10.1109/JRFID.2022.3140256

fine-tuning technique has been shown effective to address the generalization problem found in other deep learning applications [13]. Fine-turning still requires new data measured from the unknown domain. However, such new data should be as few as possible; otherwise, it will still incur great efforts and a high cost, which hinder the easy deployment of the technique in practice. To this end, meta-learning, a.k.a. "learning to learn" [14], provides an effective solution. Meta-learning optimizes the deep learning model using different learning tasks or datasets [15], so that the model will be appropriately initialized and be amenable to adapt to new domains. When applied to a new RF environment, the meta-learning model will only require a few training examples from the new environment for fine-tuning (i.e., few-shot fine-tuning), while still achieving a satisfactory performance.

In this paper, we tackle the environment/domain adaptation challenge with a meta-learning approach [16]. We propose a novel environment-adaptive, RFID based 3D human skeleton tracking system termed *Meta-Pose*. As in our prior work RFID-Pose [9], RFID tags are attached to the human body and interrogated by an RFID reader, such that the movements of human joints will be captured by analyzing the phase information in received RFID responses. Meta-Pose is also a vision-assisted scheme, where Kinect captured video data of the same human activity is used for supervised training. Note that the vision data will only be used for training the deep learning model; it will not be needed in the inference stage. Therefore the use of Kinect does not cause privacy concerns. To address the generalization problem, we first analyze the main causes for the divergence of RFID data in different RF environments. Based on the analysis, we then propose a novel Meta-Pose initialization algorithm to pretrain the model with RFID data sampled from a few different environments. With few-shot fine-tuning, the Meta-Pose system will be able to accurately track 3D human skeleton in a new, unknown environment. Extensive experiments are conducted to validate the high environment adaptation ability and high accuracy of the proposed Meta-Pose system.

The main contributions of this paper are summarized below.

- To the best of our knowledge, Meta-Pose is the first environment-adaptive system for 3D human pose tracking, which is designed using off-the-shelf RFID reader and tags. Meta-Pose can be easily deployed to estimate and track 3D human poses with RFID data in different RF environments.
- We analyze the divergence of RFID data measured in different propagation environments and identify the main challenges to the generalization problems, including sensitivity divergence of RFID tags and phase distortion in different sampling environments.
- We propose a novel Meta-Pose initialization algorithm based on meta-learning algorithms (i.e., model-agnostic meta-learning (MAML) and Reptile) to pretrain the deep learning model with a limited number of training datasets sampled from several known environments. We develop the initialization approaches based on both Reptile and MAML. A domain fusion technique is incorporated to generate more synthesized (or, fake) environments for model pretraining, to allow the pretrained model be quickly adapted to a new environment.
- We develop a prototype with off-the-shelf RFID tags and reader, and use Kinect 2.0 to measure the ground truth data for training the model and for performance evaluation. The performance of Meta-Pose is validated with extensive experiments as well as a comparison study with a baseline scheme termed RFID-Pose developed in our prior work [9]. The experimental results show that the proposed Meta-Pose system can accurately track 3D human poses while achieving high environmental adaptability simultaneously.

In the remainder of this paper, Section III briefly summarizes and contrasts with related works. The background of the proposed system is presented in Section III. Section IV examines the challenges of the domain adaptation problem. Section V presents our meta-learning based solution to these challenges. Our implementation and experimental study are presented in Section VI. Section VII summarizes this paper.

II. Related Work

In this section, we examine the related work on human pose estimation and tracking, which can be roughly classified into video camera-based schemes, WiFi-based schemes, radar-based schemes, and RFID-based schemes.

A. Traditional Pose Tracking Systems

A strength of the traditional camera, WiFi, and radar-based systems is that they are "markerless" methods, which are less intrusive. Video camera was first used to detect human poses in [17], [18]. With deep learning models, such systems localized the coordinates of human joints in the captured video frames, using, e.g., 2D RGB cameras [1], [19] or 3D depth cameras [20]. The most accurate 3D pose tracking performance was achieved, so far, by the Vicon system [21], which has been widely used for production of 3D movies. However, such video based schemes usually raise privacy concerns, as discussed, and their performance is usually limited by poor illumination, cluttered background, or poor camera angles.

To address the privacy concerns and mitigate the dependency on lighting and background, several RF pose tracking techniques have been proposed. Since such systems record no vision data and the RF data is not visible, user privacy can be better preserved. Furthermore, RF sensing systems perform well in poorly lighted environments and are able to detect human poses through obstacles and walls [5], [22]. FMCW Radar was first utilized to construct both 2D and 3D human poses by incorporating a vision-aided teacher-student deep learning model [4], [22]. As another type of non-intrusive sensor, WiFi channel state information (CSI) has also been analyzed to extract 2D and 3D human poses [6], [7]. Most existing RF sensing systems incorporate a deep learning model with vision data supervised training. Furthermore, due to the relatively wide transmission range of the radio signals, such systems are susceptible to interference from the operating environment. Usually radar-based techniques are more resistant to

environmental interference than WiFi-based schemes, but their customized hardware, e.g., the FMCW radar implemented on the Universal Software Radio Peripherals (USRP) platform, usually incurs a higher cost.

B. RFID-Based Pose Estimation Systems

RFID tags can serve as low-cost and light-weight wearable sensors to attach to the human body, which provides a promising solution for human pose estimation. Several RFID sensing techniques have been developed in recent years, such as human vital sign monitoring [23]–[26], mechanical vibration sensing [27], user authentication [28], material identification [29], and temperature sensing [30], Furthermore, RFID has also been utilized for indoor localization [31]–[34] and drone navigation [35]–[37].

Using RFID tags as wearable sensors, such systems are usually more robust to interference from the operating environment than other RF sensing techniques (e.g., WiFi). This feature inspires the development of several RFID based human pose tracking systems as well. For example, RF-Wear [38] and RF-Kinect [39] were developed to track the movements of a single human limb, while RFID-Pose [9] and Cycle-Pose [10] were developed to track 3D human poses in realtime. However, although the near-field RFID communications are more resilient to environmental interference, the locations of the tags and antennas still have a big impact on how the tags are sampled by the reader, and thus on the performance of the human pose tracking system.

In [40], the authors presented a domain adversarial technique to adapt to changes in the environment by utilizing a domain discriminator, which can constrain the unnecessary feature extraction from different environments. However, the proposed learning model may not be able to obtain the optimal variables when applied in a new RF environment, because all the training variables are determined by the datasets from a limited number of environments.

Inspired by the existing human pose tracking systems, we propose the Meta-Pose system in this paper, which is based on the meta-learning framework for greatly enhanced environmental adaptability. The proposed system incorporates a novel initialization algorithm to pretrain the deep learning model using a limited amount of training data, so that the system can be quickly fine-tuned with a small amount of new data when applied to a new environment, while still achieving a satisfactory performance.

III. Preliminaries of RFID-Based Human Pose Tracking

The Meta-Pose system is proposed to estimate 3D human pose with RFID data collected from the passive RFID tags attached to the human subject. An overview of the Meta-Pose system is shown in Fig. 1. The Meta-Pose system comprises three key components, i.e., (i) RFID phase data collection, (ii) RFID phase preprocessing, and (iii) a deep neural network.

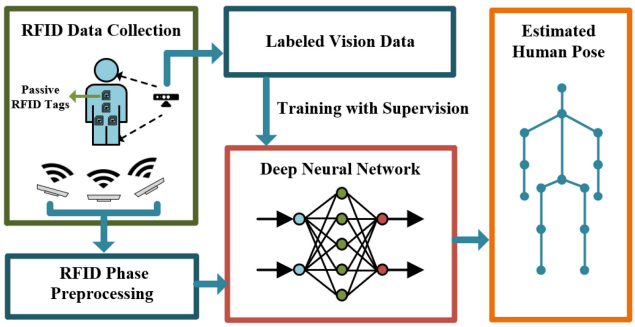


Fig. 1. Overview of the proposed RFID pose tracking system.

A. RFID Phase Data Collection and Preprocessing

In the RFID pose tracking system, the human pose is learned from RFID phase data, which is obtained by interrogating the tags attached to the human body using the RFID Low Level Reader Protocol (LLRP) [41]. The received RFID signal on a channel c can be written as [41]:

$$H = \sum_{m=1}^{M} \alpha_m e^{j\{2\pi 2R_m f_c/v + \Theta_c\}}, \tag{1}$$

where v represents the speed of light, M is the total number of RF signal propagation paths, α_m and R_m represent the signal strength and distance of each multipath component, respectively, f_c is the current channel frequency, and Θ_c is the initial phase offset caused by the circuit of both the antenna and the tag on channel c. Due to the limitation of the Gen2 protocol used in the current commodity RFID systems, only one phase value of H could be directly sampled by the system. Due to the multipath effect, deriving the phase value of the line-of-sight (LOS) component from (1) is difficult. The reported phase value may not accurately depict the relationship between propagation distance and received phase. Fortunately, the polarized reader antenna operates as both the transmitter and receiver, and the interference caused by multipath reflections is not strong. Thus, we can assume that each propagation environment has at least one dominant path, and the received phase is given by:

$$\Theta = 2\pi 2Rf_c/v + \Theta_c, \quad c = 1, 2, \ldots, 50, \tag{2}$$

where R is the distance of the dominant path between the reader antenna and tag. while the channel index c changes from 1 to 50 for every 200ms on each channel following the FCC regulation [41].

The LOS typically contributes significantly to the received signal for the following two reasons. First, in a passive RFID system, the only source of power utilized to send a response to the RFID reader is the tag antenna. The signal strength from reflection paths is typically much weaker than that of the LOS path. Additionally, the RFID reader uses a power threshold for packet detection, which means that if there is no LOS path between the antenna and the tags, the interrogation is likely to fail. Second, the RFID phase data shall be prepossessed to mitigate the impact of the randomness in Θ_c on different channels. To this end, using the phase variation Φ between two

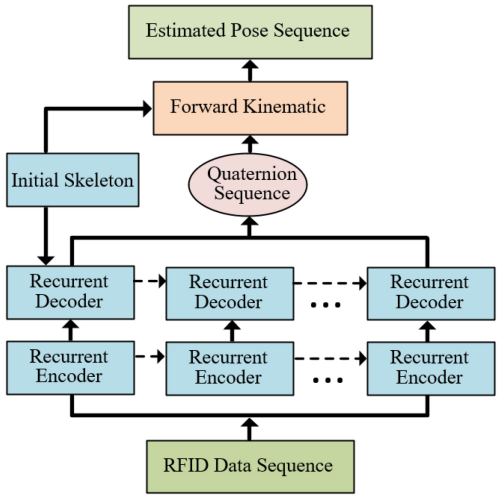


Fig. 2. Structure of the deep learning model used in RFID based 3D human pose tracking.

adjacent samples from the same channel would be effective to cancel most of the randomness, which is given by:

$$\begin{aligned}\Phi(n) &= \Theta(n) - \Theta(n-1) \\ &= 2\pi 2(R(n) - R(n-1))f_c/v,\ c = 1, \ldots, 50, n > 1,\end{aligned} \tag{3}$$

where n is the sample index on each channel and $R(n)$ is the propagation distance corresponding to the nth sample on channel c. As (3) shows, the impact of the random channel hopping offset Θ_c (see (1)) is effectively canceled, except for the first sample on each channel (which is discarded). The phase variation only depends on the changes in the range of the dominant propagation path $\Delta R(n) = R(n) - R(n-1)$. Therefore, the sequence of phase variations $\{\Phi_2, \Phi_3, \ldots, \}$ can be translated into a sequence of antenna-tag distance variations $\{\Delta R_1, \Delta R_2, \Delta R_3, \ldots, \}$, which captures the realtime movements of the RFID tags. Consequently, with the RFID phase variations for the attached tags can be leveraged to reconstruct the human skeleton and track 3D human poses in realtime.

B. Multi-Modal Deep Neural Network

Although phase variation can effectively capture the movements of the tags attached to human body, the translation from phase variation data to 3D human pose is still a challenge. In the few existing RFID based human pose tracking systems, the transformation is mostly accomplished using deep learning techniques [9], [10], which is mainly composed of a recurrent autoencoder and a forward kinematic layer. The brief structure of the deep learning model is presented in Fig. 2. As the figure shows, the network is designed to generate a sequence of 3D human poses, consisting of coordinate data of the RFID tags extracted from received RFID phase data. Specifically, the recurrent encoder is to extract both long-term and short-term features from the RFID phase data sequence, which are then fed into the following recurrent decoder. With a given initial skeleton, the decoder layer will transfer the features of the RFID data sequence to a quaternion sequence. Finally, the Forward Kinematics module will construct the human pose sequence using the quaternion sequence, which is a widely used technique in robotics and 3D animation [42].

Rather than using RF signals to generate a confidence map for human skeleton reconstruction as in prior works [1], [6], our RFID-based human pose tracking system is designed to estimate human pose with the forward kinematic technique, which has been widely used in robotics and 3D animation [42]. This is because the information rate (or, the sampling rate) of the RFID system is too low to generate a useful confidence map with an acceptable resolution. However, the forward kinematic technique only requires the quaternions of the human skeleton joints, which indicates the 3D rotation angle of each human limb. Compared with AoA based localization techniques, the output human pose does not contain the global position of each human joint, but the location relative to the root joint (pelvis). The ambiguity in AoA based localization techniques is not an issue because continuous human pose estimation mainly focuses on monitoring the relative movement of human limbs. As a tradeoff, the additional constraint is that the initial human skeleton should be the input to the system, which contains the length of each human limb, so that the the precise relative joint location can be estimated based on rotation angles.

As in RFID-Pose [9] and Cycle-Pose [10], vision data collected by a Kinect 2.0 camera is used as labels for supervised training of the deep learning model. The model is trained with a loss function that computes the difference between the estimated pose and the labeled vision data sampled simultaneously when the RFID data is collected, so the well-trained network can effectively transform RFID data sequence to a sequence of 3D human poses [9].

IV. Challenges in Domain Adaptation

RF-based systems can better protect users' privacy and do not require sufficient lighting, compared to vision based approaches. However, they also bring about several unique challenges. Unlike vision data, the RF signal is usually sampled with unrelated noise from the system itself and the environment, which is hard to mitigate. The same test subject performing the same activity could generate very different RF data when being sampled in different environments, making it hard to reconstruct poses using a model well trained offline. To improve the adaptability of the system to different environments, generalization of the deep learning model is a big challenge needs to be addressed. To analyze the influence of the environment, we use the term *data domain* to denote a specific wireless propagation environment in this paper. Since the tags are attached to the human body, different data domains could be generated by the following ways. First, we fix the antennas and change the position of the subject. Second, we fix the subject position but change the antenna deployment. Finally, we could change the surrounding around the subject or the antenna. The wireless propagation environment depends on the characteristics of all the propagation paths, which could be significantly different in a different data domain.

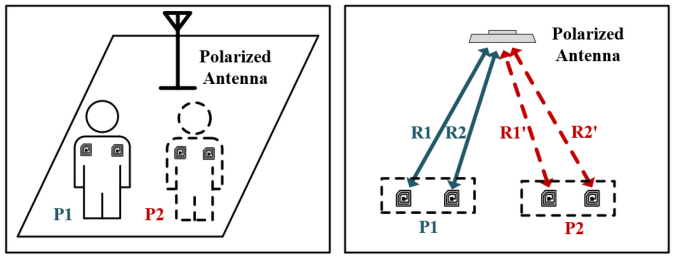


Fig. 3. Illustration of data domains in RFID sensing systems.

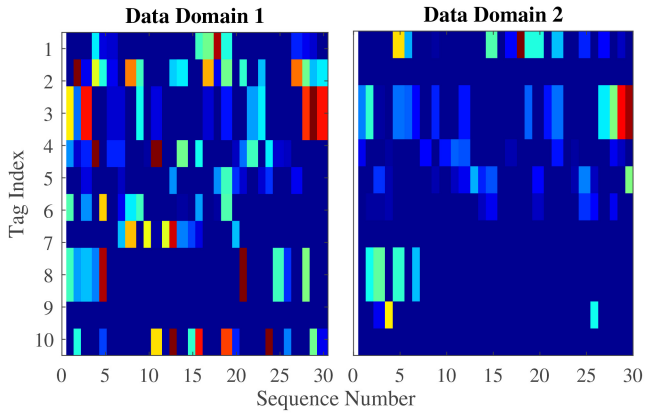


Fig. 4. Phase distortion in RFID data collected in two different data domains.

However, the passive tags are merely powered by the incident signal from the reader, while the reader has a threshold for received power needed for a successful tag interrogation. Following FCC regulations, the effective radiated power of RFID should be less than 1watt. Therefore, we usually need to ensure that all tags are within 5 meters from the antenna. Compared to other long-range systems, e.g., WiFi and Radar, the RFID system is considered as a near-field system, and the environmental interference is relatively weaker. The data domain of RFID sensing systems is mainly determined by the relative position between the subject (i.e., the tags) and the reader antenna. As shown in the left plot in Fig. 3, when the subject stands at different positions, i.e., P_1 or P_2, the sampled phase data will come from two different data domains denoted by D_{p1} and D_{p2}, respectively. The divergence of different data domains is mainly caused by: (i) the divergence in successful interrogation probability, and (ii) the distortion in RFID phase data, which are analyzed in the rest of this section.

A. Successful Interrogation Probability Divergence

The first cause of data divergence in different domains is the variation in Successful Interrogation Probability. When multiple tags are scanned by a reader, some tags are more likely to be detected, while some others may hardly be scanned. We define Successful Interrogation Probability as the probability for a tag to be successfully interrogated by the reader, which mainly depends on the received power strength from the tag. Following the Friis model, the received power S_r from a passive RFID tag can be written as [43]:

$$S_r = G_{An} G_{Tag} \gamma (\lambda_c/(4\pi R))^4 S_t, \tag{4}$$

where S_t is the reader's transmit power; G_{An} and G_{Tag} are the gains of the transmit antenna and the tag, respectively; γ represents the aggregated attenuation coefficient, accounting for the losses incurred in the antenna cable and polarization, etc. during the transmission process; λ_c is the wavelength of the current channel c; and R is the LOS path range as in (2). Eq. (4) shows that with the same antenna and tag, the received power strength is degraded by an increased LOS path distance R and the attenuation loss γ, as:

$$S_r = K_c \gamma R^{-4} S_t, \tag{5}$$

where K_c is the product of all other coefficients other than γ and R, which takes different values in different tag and antenna deployment scenarios.

For example, see Fig. 3. When the subject is in the P_1 position, the LOS path distances for Tag 1 and Tag 2 satisfy $R_1 > R_2$. Because of the limited scanning range of the polarized antenna, the polarization loss γ of Tag 1 is also higher than that of Tag 2. Referring to (5), on the same channel c, the received power from Tag 1, denoted by S_r^{Tag1}, should be smaller than that from Tag 2, denoted by S_r^{Tag2}. However, when the subject is sampled in position P_2, we will have $S_r^{Tag1} > S_r^{Tag2}$. In RFID systems, tags with a higher S_r are more likely to be successfully interrogated than tags with a lower S_r, especially when multiple tags are scanned by a single antenna. Consequently, when multiple tags are attached to the human body, the sensitivity of the tags could be very different in different data domains.

The influence of Success Interrogation Probability divergence in different data domains could be considerable for RFID based pose tracking using a deep learning model. Since the tags with a higher sensitivity are more likely to be sampled, the training dataset will be mostly composed of the data from such tags. Thus, the training variables in the deep learning model will be mostly trained by the data from the tags with higher sensitivity. When applying a trained deep learning model to a different data domain, the inference performance could be poor, since the Success Interrogation Probability in the new data domain could be very different from that where the model was originally trained.

B. Phase Distortion in Different Data Domains

The second cause of data domain divergence is the phase distortion caused by different antenna deployment scenarios. As (2) shows, the phase data of each tag is determined by the LOS propagation path distance R, which is the length of the space vector $\vec{R}$. For the tags attached to a moving human body, we can consider the overall space vector as the sum of two subspace vectors as: $\vec{R} = \vec{R}_s + \vec{R}_d$, where $\vec{R}_s$ is the *static vector* determined by the deployment scenario and $\vec{R}_d$ is the *dynamic vector* generated by the subject's movements.

Fig. 4 plots 30 sequentially received phase data from 10 RFID tags attached to the human body, where the phase value is represented by different colors. Two antennas are used to interrogate the tags simultaneously. We change the antenna

deployment positions to create two different data domains. From the figure, we can observe considerable divergence in the collected phase values from the two data domains.

According to (2), the sampled phase Θ is affected by both $\vec{R}_s$ and $\vec{R}_d$ as:

$$\Theta = 2\pi 2|\vec{R}_s + \vec{R}_d|f_c/v + \Theta_c,\ c = 1, 2, \ldots, 50. \quad (6)$$

Even if we have an identical $\vec{R}_d$ in the two data domains (i.e., the same subject and the same movement), the sampled phase could still be very different when the antennas are deployed differently (which leads to a different $\vec{R}_s$). Consequently, different antenna deployment scenarios will have an impact on the RFID phase distortion, causing considerable divergence between the datasets sampled from different environments.

Unlike Success Interrogation Probability divergence, a change in the operating environment usually causes considerable phase distortions in all sampled phase data. Thus, the model variables in the deep learning network should be trained and optimized to combat such phase distortion. Given all kinds of possible deployment environments, it is a big challenge to generate a well-optimized deep learning model, which is generalizable to all different environments.

V. Meta-Learning Based Solutions

A. Meta-Learning for Domain Adaptation

The adaptation problem to new data domains or new tasks has been investigated in prior works. On one hand, researchers try to optimize the model variables, so that the network can achieve good adaptation in different data domains. The most straightforward approach is to train the model using datasets from more and more data domains. However, to achieve a good generalization performance, the training data should cover numerous data domains, incurring an overly high cost on obtaining labeled training data. To address this issue, the adversarial learning approach has been proposed to improve network adaptability by training using a limited number of data domains with the generative adversarial network (GAN) model [11], [12]. A domain discriminator is leveraged to constrain the loss function of the neural network, in order to combat the unrelated features from different data domains. The advantage of this approach is that the network does not need to be trained again when applied to a new data domain, but the network variables are not well optimized when only considering the specific known data domains.

On the other hand, the network variables can be fine-tuned in the new data domain. Rather than addressing the data divergence issues in the well-trained network, this approach relies on additional training data for fine-tuning. The purpose is to let the network be further optimized in the specific new domain with a small amount of new training data sampled from the new domain. For typical pose tracking applications, e.g., video gaming or long-term pose monitoring, such light calibration is usually acceptable. Therefore, fine-tuning has been recognized as a promising way to improve generalization. With this approach, the network variables should first be well initialized in the pretraining stage, and then the fine-tuning process will be performed quickly with only a few additional data from the new data domain.

Algorithm 1: Reptile Based Initialization Algorithm

Input: Sampled data sets from the four known data domains (denoted by D_1, D_2, D_3, and D_4);
Output: Optimally initialized variables X_t for the pretrained network.
Randomly initialize the training variable as X;
for $i = 1{:}n$ **do**
  Generate d_i by randomly sampling from D_1, D_2, D_3, and D_4;
  Randomly sample k batches from d_i;
  Set the inner loop training variables: $X_{in} \leftarrow X$;
  for $j = 1{:}k$ **do**
    Update the variables in X_{in} with loss function L as: $X'_{in} = U^1_{d_i}(X_{in})$, $X_j = X'_{in} - X_{in}$, $X_{in} \leftarrow X'_{in}$;
  end
  Calculate the outer loop gradient as: $\hat{W}_i = \sum_{j=1}^{k} W_j$;
  Update the outer loop variables X as: $X \leftarrow X + \epsilon \hat{W}_i$;
end
Set $X_t \leftarrow X$;

Algorithm 2: MAML Based Initialization Algorithm

Input: Sampled data sets from the four data domains (denoted by D_1, D_2, D_3, and D_4);
Output: Optimally initialized variables X_t for the pretrained network.
Randomly initialize the training variables as X;
for $i = 1{:}n$ **do**
  Generate d_i by randomly sampling from D_1, D_2, D_3, and D_4;
  Randomly sample 2 batches B_1 and B_2 from d_i;
  Set the inner loop training variables: $X_{in} \leftarrow X$;
  Update the variables in X_{in} with loss function L and dataset B_1 as: $X'_{in} = U^1_{d_i}(X_{in})$;
  Update the variables in X'_{in} with loss function L and dataset B_2 as: $X''_{in} = U^1_{d_i}(X'_{in})$;
  Calculate the outloop gradient as: $\hat{W}_i = X''_{in} - X'_{in}$;
  Update variables X as: $X \leftarrow X + \epsilon \hat{W}_i$;
end
Set $X_t \leftarrow X$;

Meta-learning has been proved to be an effective technique for model pretraining so that a pretrained model can be quickly adapted for a new data domain [14]. In the case of the RFID based pose tracking, when data is sampled from an untrained domain, the performance of the previously trained model will usually degrade. New training data sampled from the new data domain is necessary to fine-tune the model for the new domain. The MAML algorithm is a representative meta-learning algorithm to pre-train the model for a satisfactory initialization before fine-tuning [15]. The Reptile algorithm [44] is another representative meta-learning algorithm for model pretraining, which has been shown to achieve a similar performance as MAML but at a lower computational complexity. We leverage these two algorithms in Meta-Pose to adapt the model to a new, unknown environment with few-shot fine-tuning using a few new training data. In the Meta-Pose system, we implement both meta-learning algorithms for network initialization, and

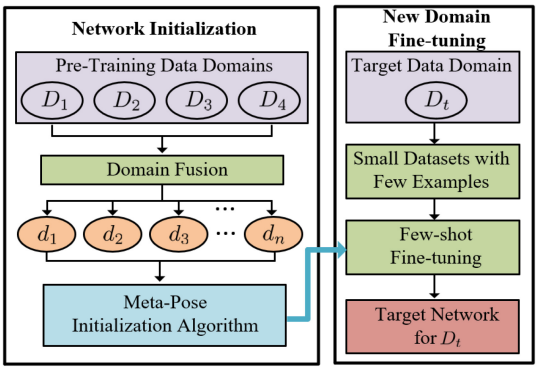


Fig. 5. Training framework of the proposed Meta-Pose system.

then fine-tune the pretrained network for a new data domain using only a few data examples. These three key components are presented in the remainder of this section.

B. Meta-Learning Framework With Domain Fusion

The objective of meta-learning is to determine the satisfactory initial model variables through network initialization, which can then be adapted to a new data domain with a few training examples. With appropriately trained initial variable x, the network loss for data domain D should be minimized after few steps of fine-tuning. Thus, the optimization problem for network initialization can be formulated as:

$$\min_{X} \ \mathbb{E}_D\Big[L\Big(U_D^k(X)\Big)\Big], \tag{7}$$

where $L(\cdot)$ denotes the loss function of the network, and $U_D^k(X)$ denotes the gradient descent operation that updates variables X for k times using the data sampled from D, which is the Adam algorithm.

Equation (7) shows that, the meta-learning algorithm considers the gradient descent process as optimization target. Thus, rather than the normal training process based on the gradient of the loss function $\Delta L(X)$, meta-learning calculates the gradient of the gradient descent $\Delta L(U_D^k(X))$ in each training step. From the equation we can see that the performance of meta-learning can be determined by the amount of training data domain in D. However, it is highly costly to directly sample a large amount of human pose data from numerous data domains. Therefore we develop a domain fusion based meta-learning algorithm for model pretraining. The domain fusion algorithm randomly selects samples from the four known domains to form new domains, in order to increase the number of known domains for pretraining.

Figure 5 represents the brief structure of the training procedure of the proposed Meta-Pose system, which consists of network initialization and fine-tuning in a new domain. As shown in the figure, the deep learning model is first pretrained using datasets from a few (e.g., four) known data domains, which are sampled when the subject is standing at four different positions. The network is pretrained with two different meta-learning algorithms. Since the second-order gradient $\Delta L(U_D^k(X))$ is hard to calculate in practice, we leverage the first-order approximation instead to update the training variables. Based on the divergence in the first-order approximation, we develop two different initialization approaches based on Reptile and MAML algorithms, respectively. When transferring the learning task to a target data domain D_t, we only need to collect very few examples in the target domain to fine-tune the generalized network.

C. Reptile-Based Network Initialization

In the Reptile-based algorithm, we first fuse the four data domains (i.e., D_1, D_2, D_3, and D_4) into a larger number of fused data domains (i.e., $d_1, d_2, \ldots, d_n$). Specifically, each d_i contains 40 batches of data randomly sampled from D_1, D_2, D_3, and D_4. To solve the optimization problem (7), we need to find the gradient of any fused data domain $\Delta L[U_{d_i}^k(X)]$, so the gradient descent algorithm can be applied to find X by recursive updating. With the Reptile learning algorithm [44], we first calculate $\Delta L[U_{d_i}^1(X)]$ for each iteration in the inner loop as:

$$\Delta L\Big[U_{d_i}^1(X_{in})\Big] = U_{d_i}^1(X_{in}) - X_{in} = X'_{in} - X_{in}, \tag{8}$$

where X_{in} is the set of variables used in the inner loop. In the algorithm, denote the one-step gradient $\Delta L[U_{d_i}^1(X_{in})]$ as W_j. The overall gradient after k iterations is calculated as:

$$\Delta L\Big[U_{d_i}^k(X)\Big] = \sum_{j=1}^{m} W_j. \tag{9}$$

$\Delta L[U_{d_i}^k(X)]$ is denoted as $\hat{W}_i$ for each data domain d_i. In the algorithm, we set $k = 8$ for effective training in each data domain. With gradient $\hat{W}_i$, we solve problem (7) by recursively training variable X in the outer loop iterations as:

$$X \leftarrow X + \epsilon \hat{W}_i, \tag{10}$$

where ϵ is the learning rate, which is set to 0.1 in the system. We repeat the updating process for 5,000 times (i.e., setting $n = 5{,}000$), so the final training result X_t could satisfy the initialization requirement of problem (7).

D. MAML-Based Network Initialization

With MAML based initialization, we leverage the same method to generate fused data domains d_i for each iteration in the outer loop updates. Unlike the Reptile algorithm, $\Delta L(U_D^k(X))$ is approximated by two-step training [15]. Thus, for each iteration, we firstly sample two batches of data B_1 and B_2 from the fused data domain d_i and update the variables X_i with one-step gradient descent using batch data B_1 to obtain X'_{in}. In the MAML-based learning algorithm, we set $k = 1$ to reduce the training complexity. So the outer loop gradient can be approximated by:

$$\Delta L\Big[U_{d_i}^1(X)\Big] = \Delta L\big[X'_{in}\big] = U_{d_i}^1\big(X'_{in}\big) - X'_{in}. \tag{11}$$

We next update X'_{in} by one more step of gradient descent using another batch data B_2 and generate $X''_{in} = U_{d_i}^1(X'_{in})$. Accordingly, the outer loop gradient is estimated as the gradient of the second step training, which is calculated by $X''_{in} - X'_{in}$. With the outer loop gradient found by the MAML based algorithm, the training variables X is initialized by recursively

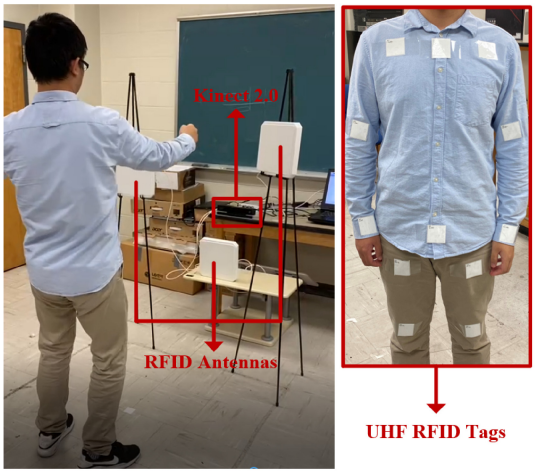


Fig. 6. Experiment configuration of the Meta-Pose system.

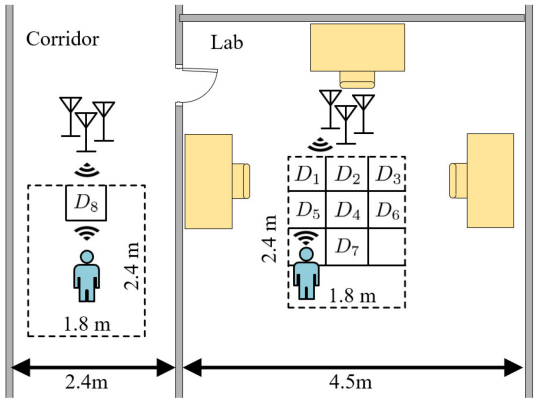


Fig. 7. Illustration of the data domains used in the Meta-Pose experiments.

updating it in the outer loop following (10), where the learning rate ϵ is also set to 0.1. The update is also iterated for 5,000 times, so a large number of fake data domains will be used in model pretraining. After initialization training, the network will be able to be quickly fine-tuned using a few shots of data sampled from a new data domain.

E. Few-Shot Fine-Tuning

After an appropriate initialization of X, the fine-tuning process only requires a very small dataset from the new data domain. Since the training data are all in the form of data sequences, including RFID phase data and Kinect vision data [10], the data shots are defined specifically in the Meta-Pose system. We divide the data sequence into small segments during the training process, each consisting of 30 consecutive data samples sampled within a window of 6s. We consider one such data batch as a shot in Meta-Pose, and less than 5 shots of data from the new data domain will be leveraged for fine-tuning. We also find that the type of activities also affects the fine-tuning performance and will discuss this further in Section VI.

VI. Implementation and Evaluation

A. System Implementation

To evaluate the performance of Meta-Pose, we develop a prototype system using an off-the-shelf Impinj R420 reader, which is equipped with three S9028PCR polarized antennas, as shown in Fig. 6. ALN-9634 (HIGG-3) RFID tags are used in Meta-Pose operating in the Ultra High Frequency (UHF) band. The vision data, used for training supervision as well as ground truth for evaluating the precision of inference, is collected using an Xbox Kinect 2.0 device. As shown in the figure, we attach 12 RFID tags to the 12 joints of the subject, including neck, pelvis, left hip, left knee, right hip, right knee, left shoulder, left elbow, left wrist, right shoulder, right elbow, and right wrist. With the three reader antennas placed at different positions with different heights, every RFID tag can be interrogated by at least one of the antennas.

TABLE I
Performance Evaluation for Different Subjects

Subject Index	*Estimation Error*
Subject 1	3.62cm $\pm$ 0.24cm
Subject 2	4.35cm $\pm$ 0.34cm
Subject 3	3.78cm $\pm$ 0.22cm
Subject 4	5.12cm $\pm$ 0.37cm
Subject 5	4.17cm $\pm$ 0.31cm

Environment adaption is validated using RFID data collected from eight different data domains, which are generated by specific deployments of the subject and antennas as illustrated in Fig. 7. Seven data domains are sampled in a computer lab, and the eighth data domain is sampled in an empty corridor. Each domain is a 0.6×0.6 m^2 square area, where the subject shall stand inside performing certain activities during data collection. With the 900 MHz frequency and 0.33 m wavelength used in the proposed RFID system, a 0.6 m interval is sufficient to generate considerable divergence to create different data domains. Among these domains, D_1 to D_4 are used for model pretraining, where 70% of the sampled data from each domain is used for training, and the rest 30% is used for testing. D_5 to D_8 are considered as new data domains for evaluating the generalization performance, where 50% of the data from each of these domains is used for fine-tuning, and the rest 50% is used for testing.

RFID phase data is collected when the subject stands in front of the antennas and repeatedly performs specific activities. Different types of activities are sampled in all the data domains, such as walking, body twisting, deep squatting, and moving a single limb. Five subjects participate in the experiments for sufficient data diversity, including one female and four males. The sampling rate of the antenna is 110 Hz. However, due to the collision avoidance protocol, when the reader is interrogating multiple tags, only one randomly chosen tag could respond to the reader at a time. The sampling rate for each tag of a multi-tag system is not even nor constant, depending on the relative location of each tag, interference, and the mutual coupling effect [31]. To deal with the low sampling rate and sparse RFID data, we firstly construct a tensor with the sampled raw data, and then leverage tensor completion to interpolate the missing data. Finally, the calibrated data is downsampled to 5 Hz for processing in realtime.

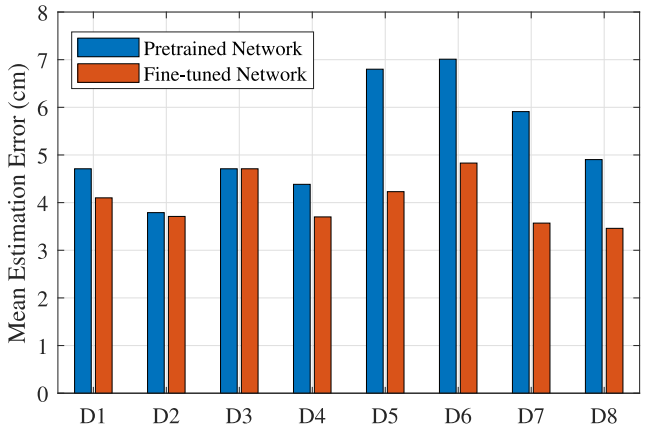


Fig. 8. Overall performance in terms of mean estimation error in the eight different data domains.

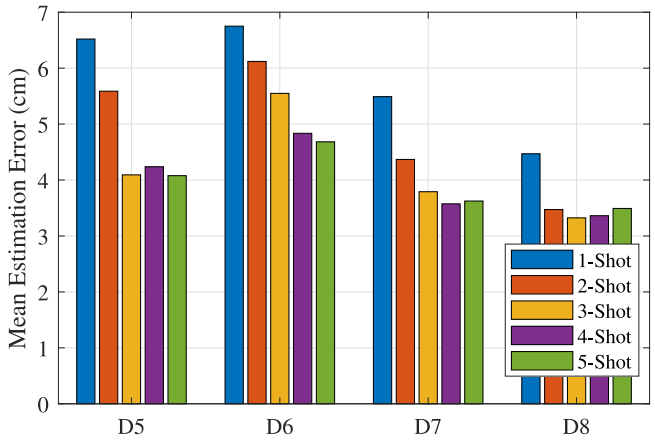


Fig. 9. Fine-tuning performance of Reptile based initialization using different shots of new data.

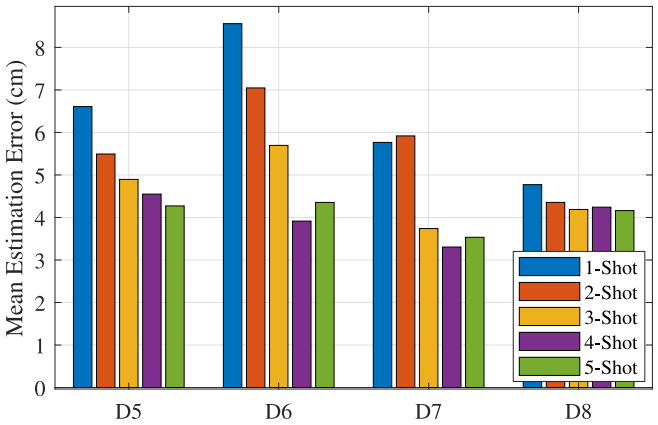


Fig. 10. Fine-tuning performance of MAML based initialization using different shots of new data.

B. Overall Performance Evaluation

To demonstrate the overall system performance, we use the 3D human pose data collected by Kinect 2.0 as ground truth. For each video frame, we calculate the mean error Ψ_{all} of all the 12 joints as:

$$\Psi_{all} = \frac{1}{12}\sum_{n=1}^{12}\left\|\hat{T}_n - \dot{T}_n\right\|, \tag{12}$$

where $\hat{T}_n$ represents the estimated 3D position of joint n, $\dot{T}_n$ is the ground truth, and $\|\hat{T}_n - \dot{T}_n\|$ is the Euclidean distance between the two 3D positions.

The overall performance (i.e., mean error) of the fine-tuned network for all the eight data domains is presented in Fig. 8. Recall that only the first four data domains are used for model pretraining, while the other four domains are used for testing. In addition, we also present the accuracy of the pretrained network in the figure (i.e., without fine-tuning using additional data from the new data domain). As shown in the figure, the maximum error of the fine-tuned network is 4.83 cm achieved in D_6, while the minimum error is 3.46 cm achieved in D_8. The minimum pretraining error for the new data domain (i.e., D_5 to D_8) is 4.91 cm in D_8, which is higher than that of all the pretrained domains (i.e., D_1 to D_4). The higher pretrained errors imply the large divergence between the known and new data domains. However, with few-shot fine-tuning, the mean error for all the four new data domains is reduced to 3.98cm, which is very similar to that of the known data domains. The considerable error reduction in D_5, D_6, D_7, and D_8 is due to the Meta-Pose initialization algorithm. With the well optimized training variables, the deep learning model can be effectively fine-tuned for new data domains. Compared to the height of the subject and range of motions, the 3D human pose estimation errors are all small and negligible. These results demonstrate the high adaptability of the Meta-Pose system.

C. Fine-Tuning for the Two Pretrain Algorithms

For most effective fine-tuning, we conduct experiments to investigate the impact of the number of shots and the type of activities on different initialization algorithms. Fig. 9 illustrates the accuracy of human pose tracking in the four new data domains with Reptile initialization, which are fine-tuned with different numbers of data shots ranging from 1 to 5. Fig. 10 shows similar fine-tuning results but with MAML based initialization. As defined earlier, one-shot of data in Meta-Pose is defined as a consecutive data sequence within a time window of 6 s. It can be seen that, after 5-shot fine-tuning after Reptile initialization, the minimum error 3.49 cm is achieved in D_8, while the error in D_6 is the highest (i.e., 4.68 cm). For MAML based initialization, the minimum error 3.53 cm is achieved in D_7, and the max error is 4.34cm achieved in D_6. From the performance of different data domains shown in Figs. 9 and 10, it can be seen that both Reptile and MAML are able to compute satisfactory initial learning variables. Both models can be adapted to different new data domains within five shots of fine-tuning.

In addition, although the final estimation accuracy is different in the four data domains, the performance of fine-tuning is generally improved as more data shots are used. However, as the figure shows, the improvement becomes not obvious beyond four shots of data for both algorithms. Thus, four-shot fine-tuning will be sufficient when the Meta-Pose system is transferred to a new environment.

We also examine the impact of different types of activities based on the accuracy of tracking different types activities. In

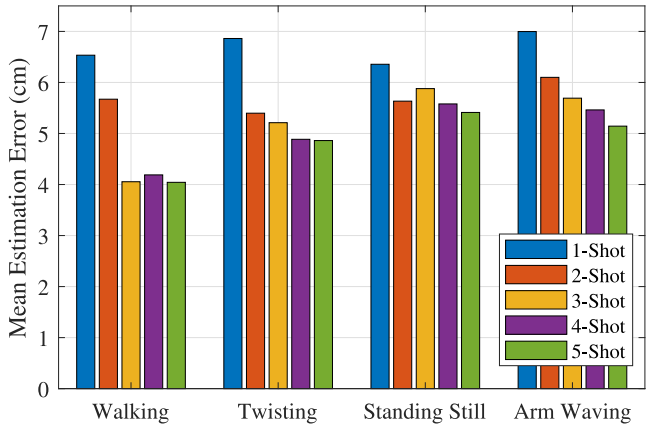


Fig. 11. Fine-tuning performance of Reptile based initialization for different activities in new data domain D_5.

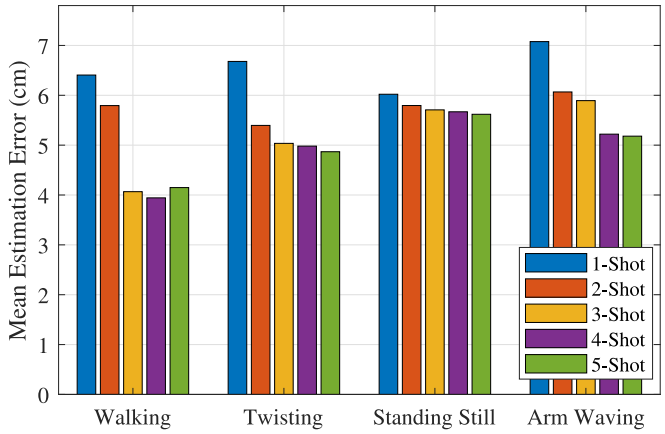


Fig. 12. Fine-tuning performance of MAML based initialization for different activities in new data domain D_5.

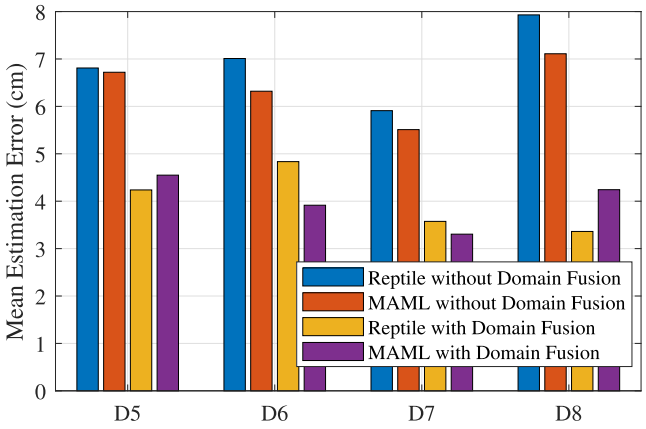


Fig. 13. Fine-tuning performance of the domain fusion algorithm and typical meta-learning algorithm.

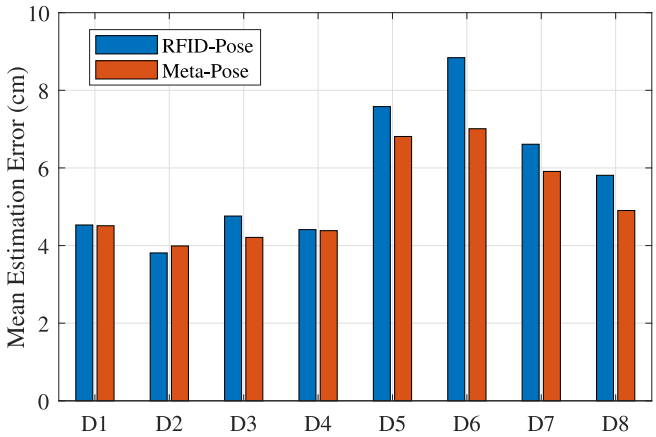


Fig. 14. Pretraining comparison with the baseline method RFID-Pose [9] without fine-tuning.

Fig. 11, we present the n-shot fine-tuning results of Reptile based initialization in the specific data domain D_5 with different types of activities, including walking, body twisting, standing still, and arm waving. We also provide the n-shot fine-tuning result of MAML initialization in Fig. 12 for comparison purpose.

Figure 11 shows that, after 5-shot fine-tuning, the minimum error 4.04 cm of Reptile based initialization is achieved when the system is fine-tuned for the walking activity, while standing has the maximum error of 5.41 cm. For MAML based initialization, the minimum error 3.94 cm is also achieved by the walking activity. We also find that fine-tuning is not as effective for arm-waving and standing for both initialization algorithms. This is because simple activities, such as standing and arm waving, contain less information than the more complicated activities, such as walking. Generally, fine-tuning will be more effective when more information is carried in the new data shots. Thus, we conclude that fine-tuning is more effective for more complicated activities, no matter which algorithm is used for network pretraining.

D. Effect of the Domain Fusion Algorithm

The superiority of the domain fusion algorithm used in meta-learning based pretraining is demonstrated by the next experiment. As shown in Fig. 5, we randomly sample training data from the four known data domains to generate more virtual data domains, i.e., the d_i's, to enhance the performance of the meta-learning algorithms. Fig. 13 illustrates the fine-tuning performance of the domain fusion algorithm and the two representative meta-learning algorithms. The figure presents the four-shot fine-tuning results with different initialization algorithms for all the four untrained data domains. As the figure shows, without the domain fusion algorithm, the minimum estimation error is 5.51 cm, and the maximum estimate error is 7.93 cm, which are quite high for 3D human pose tracking. In contrast, with the domain fusion algorithm, the minimum error is reduced to 3.36 cm while the maximum error is only 4.83 cm now. Thus, the greatly reduced errors prove that, the domain fusion algorithm could effectively enhance the model pretraining and reduce the cost of obtaining training data.

E. Comparison With a Baseline Scheme

Finally, we conduct a comparison study using our recent RFID based pose tracking system RFID-Pose as a baseline scheme [9]. As in Meta-Pose, we leverage the same training dataset collected from D_1 to D_4 to pretrain the RFID-Pose model. The estimation error for all the domains are presented in Fig. 14 without fine-tuning. The figure validates that both

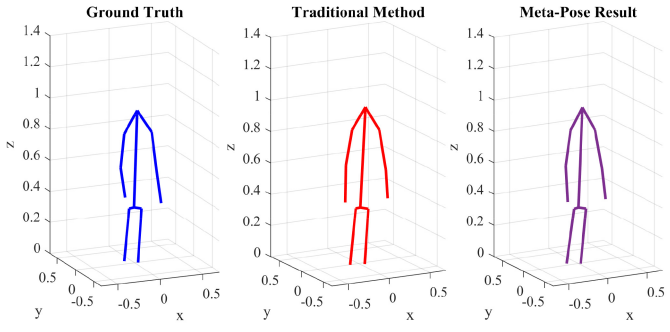


Fig. 15. Comparison results for a pretrained data domain D_4.

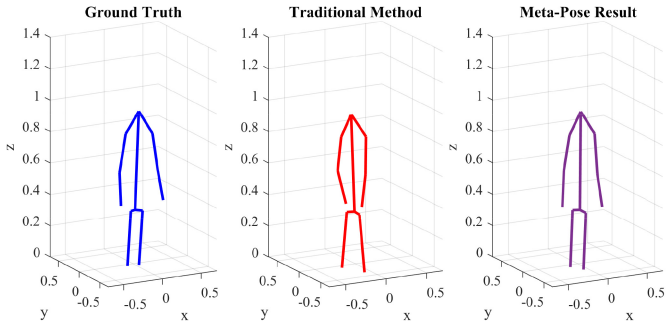


Fig. 16. Comparison results after 4-shot fine-tuning for new data domain D_5.

systems achieve a good, comparable performance for the known domains (i.e., D_1 to D_4). However, RFID-Pose has relative larger errors when applied to the four unknown domains (i.e., D_5 to D_8). The maximum error of RFID-Pose is 8.84 cm and the mean error of all the new data domains is 7.21 cm. In contrast, the mean error of Meta-Pose for the unknown domains is 6.12 cm without fine-tuning. These results indicate that the Meta-Pose initialization algorithm finds better initial model variables for the new data domains than RFID-Pose.

The superiority of the Meta-Pose initialization algorithm is further demonstrated with the fine-tuned results. Fig. 15 illustrates examples of estimated poses obtained by Meta-Pose and RFID-Pose for a pretrained data domain D_4. The ground truth shown on the left is generated by Kinect. The middle and the right poses are estimated by RFID-Pose and Meta-Pose after pretraining, respectively. Fig. 16 depicts the estimated poses for an unknown data domain D_5 following a four-shot fine-tuning. As demonstrated in these two examples, for a pretrained data domain, the predicted human poses by RFID-Pose and Meta-Pose are both similar to the ground truth. However, for the new data domain, the Meta-Pose generated pose is still close to the ground truth, while the traditional method generated pose looks obviously different from the ground truth.

Fig. 17 illustrates the performance of RFID-Pose for the untrained data domains ($D_5 \sim D_8$), while different numbers of data shots are used for fine-tuning. The figure shows that the traditional system requires considerably more data for adaptation to new environments. For example, at least 300 data shots are needed for fine-tuning when adapting RFID-Pose to new untrained data domain D_5, while D_6 and D_7 require 250 and 200 data shots, respectively. However, as illustrated in Fig. 9

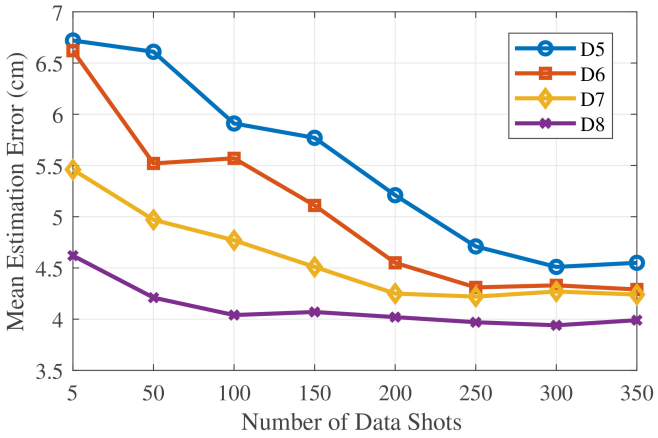


Fig. 17. Fine-tuning performance of the baseline method RFID-Pose in different data domains.

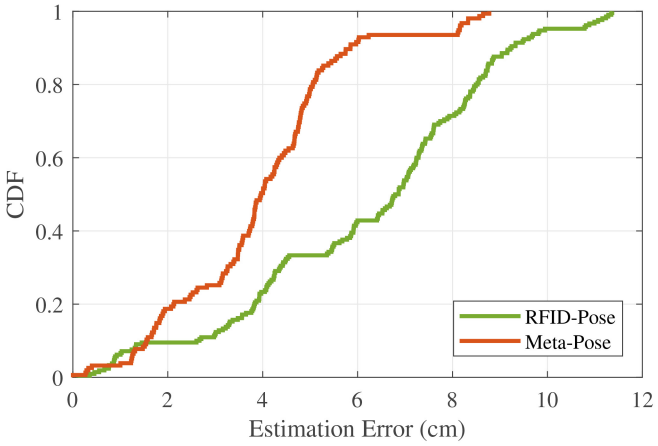


Fig. 18. The CDF curves of the four-shot fine-tuning results of RFID-Pose and Meta-Pose.

and Fig. 10, four data shots are sufficient for Meta-Pose. As defined before, one-shot data consists of 6s of consecutive data samples, and so four data shots mean the system needs to collect 24 seconds of training data for a new data domain. However, with 200 training data shots for D_7, the traditional system requires at least 20 minutes of new training data for domain adaptation, while D_5 and D_6 need 30 and 25 minutes of new training data, respectively. The large difference in the amount of training data show that Meta-Pose effectively reduces the expense of new environmental adaption.

The Cumulative distribution functions (CDF) of the estimation errors of the two systems are plotted in Fig. 18. The figure presents the estimation error after four-shot fine-tuning for all untrained data domains. The figure shows that the median estimation error of RFID-Pose is 6.87cm, whereas the median error of Meta-Pose is 3.94cm. Furthermore, we observe that the overall estimation error of RFID-Pose is considerably higher than the Meta-Pose system.

Table II presents the mean estimation error for each untrained data domain. As the table shows, the mean error of RFID-Pose for all the new data domains is 6.27 cm, while the mean errors of Meta-Pose with Reptile and MAML based pretraining are 3.97 cm and 4.03 cm, respectively. We find

TABLE II
PERFORMANCE COMPARISON AFTER FINE-TUNING

Domain	*RFID-Pose*	*Meta-Pose (Reptile)*	*Meta-Pose (MAML)*
D_5	6.72cm	4.23cm	4.55cm
D_6	7.62cm	4.83cm	3.91cm
D_7	5.46cm	3.57cm	3.30cm
D_8	4.62cm	3.36cm	4.24fcm
D_{all}	6.27cm	3.97cm	4.03cm

that the RFID-Pose error is also reduced by fine-tuning, but its estimation error for new data domains is still quite high.

The experiments show that larger datasets sampled in the new environments are needed for RFID-Pose to achieve a satisfactory fine-tuning performance, which considerably increases the training data collection effort and cost. In contrast, the error of Meta-Pose can be effectively reduced by few-shot fine-tuning, because the meta-learning-based algorithms have suitably initialized the model variables based on the known data domains. Meta-Pose is able to quickly optimize its training variables for untrained data domains with a few data examples. Through these experiments, we demonstrate that Meta-Pose can better adapt to unknown environments compared with the baseline scheme. Thus it can be easily deployed in practice in different application environments.

VII. CONCLUSION

In this paper, we proposed an RFID based realtime 3D pose tracking system, termed Meta-Pose, that is environment-adaptive. A novel Meta-Pose initialization algorithm was proposed to pretrain the network with several known data domains, and few-shot fine-tuning was then utilized to adapt to unknown data domains. The Meta-Pose system was developed with two different meta-learning algorithms, i.e., Reptile and MAML. The Meta-Pose system was implemented using off-the-shelf RFID reader and tags. Extensive experiments were conducted with ground truth provided by Kinect 2.0 vision data. Meta-Pose's high accuracy and adaptability to new environments were demonstrated by our experimental results and a comparison study with a state-of-the-art baseline scheme.

REFERENCES

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 7291–7299.

[2] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," in *Proc. IEEE CVPR*, San Francisco, CA, USA, Jun. 2010, pp. 623–630.

[3] Tom's Guide. "Millions of Wireless Security Cameras are At Risk of Being Hacked: What To Do," 2020. (Accessed: Aug. 28, 2020). [Online]. Available: https://www.tomsguide.com/news/hackable-security-cameras

[4] M. Zhao *et al.*, "Through-wall human pose estimation using radio signals," in *Proc. IEEE CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 7356–7365.

[5] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10032–10044, Sep. 2020.

[6] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proc. IEEE ICCV*, Seoul, South Korea, Oct. 2019, pp. 5451–5460.

[7] W. Jiang *et al.*, "Towards 3D human pose construction using WiFi," in *Proc. ACM MobiCom*, London, U.K., Sep. 2020, pp. 1–14.

[8] J. Zhang, S. Periaswamy, S. Mao, and J. Patton, "Standards for passive UHF RFID," *GetMobile Mobile Comput. Commun.*, vol. 23, no. 3, pp. 10–15, Sep. 2019.

[9] C. Yang, X. Wang, and S. Mao, "RFID-Pose: Vision-aided three-dimensional human pose estimation with RFID," *IEEE Trans. Rel.*, vol. 70, no. 3, pp. 1218–1231, Sep. 2021.

[10] C. Yang, X. Wang, and S. Mao, "Subject-adaptive Skeleton tracking with RFID," in *Proc. IEEE MSN*, Tokyo, Japan, Dec. 2020, pp. 599–606.

[11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE ICCV*, Venice, Italy, Oct. 2017, pp. 2242–2251.

[12] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. ACM MobiCom*, New Delhi, India, Sep. 2018, pp. 289–304.

[13] L. Wang, S. Mao, B. Wilamowski, and R. Nelms, "Pre-trained models for non-intrusive appliance load monitoring," *IEEE Trans. Green Commun. Netw.*, early access, Jun. 9, 2021, doi: 10.1109/TGCN.2021.3087702.

[14] J. Vanschoren, "Meta-learning: A survey," Oct. 2018, *arXiv:1810.03548*.

[15] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. ICML*, Sydney, NSW, Australia, Aug. 2017, pp. 1126–1135.

[16] C. Yang, L. Wang, X. Wang, and S. Mao, "Meta-Pose: Environment-adaptive human skeleton tracking with RFID," in *Proc. IEEE GLOBECOM*, Madrid, Spain, Dec. 2021, pp. 1–6.

[17] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Comput. Vis. Image Understand.*, vol. 192, Mar. 2020, Art. no. 102897.

[18] R. Mitra, N. B. Gundavarapu, A. Sharma, and A. Jain, "Multiview-consistent semi-supervised learning for 3D human pose estimation," in *Proc. IEEE CVPR*, Seattle, WA, USA, Jun. 2020, pp. 6907–6916.

[19] X. Fan, K. Zheng, Y. Lin, and S. Wang, "Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 1347–1355.

[20] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[21] L. Sigal, A. O. Balan, and M. J. Black, "HUMANEVA: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, pp. 1–24, Jul. 2010.

[22] M. Zhao *et al.*, "RF-based 3D skeletons," in *Proc. ACM SIGCOM*, Budapest, Hungary, Aug. 2018, pp. 267–281.

[23] Y. Hou, Y. Wang, and Y. Zheng, "TagBreathe: Monitor breathing with commodity RFID systems," in *Proc. IEEE ICDCS*, Atlanta, GA, USA, Jun. 2017, pp. 404–413.

[24] R. Zhao, D. Wang, Q. Zhang, H. Chen, and A. Huang, "CRH: A contactless respiration and heartbeat monitoring system with COTS RFID tags," in *Proc. IEEE SECON*, Hong Kong, Jun. 2018, pp. 325–333.

[25] C. Wang, L. Xie, W. Wang, Y. Chen, Y. Bu, and S. Lu, "RF-ECG: Heart rate variability assessment based on COTS RFID tag array," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–26, Jun. 2018.

[26] C. Yang, X. Wang, and S. Mao, "Respiration monitoring with RFID in driving environments," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 500–512, Feb. 2021.

[27] P. Li, Z. An, L. Yang, and P. Yang, "Towards physical-layer vibration sensing with RFIDs," in *Proc. IEEE INFOCOM*, Paris, France, Jun. 2019, pp. 892–900.

[28] C. Zhao *et al.*, "RF-Mehndi: A fingertip profiled RF identifier," in *Proc. IEEE INFOCOM*, Paris, France, Jun. 2019, pp. 1513–1521.

[29] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, "TagScan: Simultaneous target imaging and material identification with commodity RFID devices," in *Proc. ACM MobiCom*, Oct. 2017, pp. 288–300.

[30] X. Wang, J. Zhang, Z. Yu, S. Mao, S. C. G. Periaswamy, and J. Patton, "On remote temperature sensing using commercial UHF RFID tags," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10715–10727, Dec. 2019.

[31] C. Yang, X. Wang, and S. Mao, "SparseTag: High-precision backscatter indoor localization with sparse RFID tag arrays," in *Proc. IEEE SECON*, Boston, MA, USA, Jun. 2019, pp. 1–9.

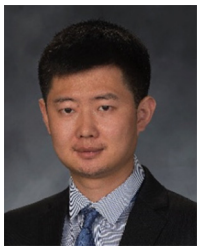
[32] J. Wang and D. Katabi, "Dude, where's my card? RFID positioning that works with multipath and non-line of sight," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 51–62, Oct. 2013.

[33] L. Shangguan and K. Jamieson, "The design and implementation of a mobile RFID tag sorting robot," in *Proc. ACM MobiSys*, Singapore, Jun. 2016, pp. 31–42.

[34] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultra-wideband localization for deployed RFID tags," in *Proc. ACM MobiCom*, Oct. 2017, pp. 248–260.
[35] J. Zhang *et al.*, "RFHUI: An intuitive and easy-to-operate human-uav interaction system for controlling a UAV in a 3D space," in *Proc. EAI MobiQuitous*, New York, NY, USA, Nov. 2018, pp. 69–76.
[36] J. Zhang *et al.*, "RFHUI: An RFID based human-unmanned aerial vehicle interaction system in an indoor environment," *Digit. Commun. Netw. J.*, vol. 6, no. 1, pp. 14–22, Feb. 2020.
[37] J. Zhang *et al.*, "Robust RFID based 6-DoF localization for unmanned aerial vehicles," *IEEE Access J.*, vol. 7, no. 1, pp. 77348–77361, Jun. 2019.
[38] H. Jin, Z. Yang, S. Kumar, and J. I. Hong, "Towards wearable everyday body-frame tracking using passive RFIDs," *Proc. ACM Interact Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–23, Dec. 2017.
[39] C. Wang, J. Liu, Y. Chen, L. Xie, H. B. Liu, and S. Lu, "RF-Kinect: A wearable RFID-based approach towards 3D body movement tracking," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–28, Mar. 2018.
[40] F. Wang, J. Liu, and W. Gong, "Multi-adversarial in-car activity recognition using RFIDs," *IEEE Trans. Mobile Comput.*, vol. 20, no. 6, pp. 2224–2237, Jun. 2021.
[41] M. Lenehan, "Application Note–Low Level User Data Support," Feb. 2019. [Online]. Available: https://support.impinj.com/hc/en-us/articles/202755318-Application-Note-Low-Level-User-Data-Support
[42] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proc. IEEE CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 8639–8648.
[43] L. Ukkonen and L. Sydanheimo, "Threshold power-based radiation pattern measurement of passive UHF RFID tags," *PIERS Online*, vol. 6, no. 6, pp. 523–526, 2010.
[44] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," Oct. 2018, *arXiv:1803.02999*.

Chao Yang (Student Member, IEEE) received the B.S. degree in electrical engineering from Yanshan University, He'bei, China, in 2015, and the M.S. degree in electrical and computer engineering (ECE) from Auburn University, Auburn, AL, USA, in 2017, where he is currently pursuing the Ph.D. degree in ECE. His current research interests include health sensing, indoor localization, Internet of Things, and wireless networks. He is a co-recipient of the Best Paper Award of IEEE GLOBECOM 2019.

Lingxiao Wang received the B.E. degree in electrical engineering and automation from the Nanjing University of Information Science and Technology, Nanjing, China, in 2012, and the M.S. and Ph.D. degrees in electrical and computer engineering from Auburn University, Auburn, AL, USA, in 2016 and 2021, respectively. His research interests include deep learning, neural network optimization, and time-series prediction.



Xuyu Wang (Member, IEEE) received the B.S. degree in electronic information engineering and the M.S. degree in signal and information processing from Xidian University, Xi'an, China, in 2009 and 2012, respectively, and the Ph.D. degree in electrical and computer engineering from Auburn University, Auburn, AL, USA, in August 2018. He is an Assistant Professor with the Department of Computer Science, California State University, Sacramento, CA, USA. His research interests include indoor localization, deep learning, and big data. He is a co-recipient of the Second Prize of Natural Scientific Award of Ministry of Education, China, in 2013, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the Best Paper Award of IEEE GLOBECOM 2019, the Best Journal Paper Award of IEEE Communications Society Multimedia Communications Technical Committee in 2019, the Best Demo Award of IEEE SECON 2017, and the Best Student Paper Award of IEEE PIMRC 2017.

Shiwen Mao (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. After joining Auburn University, Auburn, AL, USA, in 2006, he held the McWane Endowed Professorship from 2012 to 2015 and the Samuel Ginn Endowed Professorship from 2015 to 2020 with the Department of Electrical and Computer Engineering. He is currently a Professor and the Earle C. Williams Eminent Scholar Chair, and the Director of the Wireless Engineering Research and Education Center, Auburn University. His research interests include wireless networks, multimedia communications, and smart grid. He received the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award in 2019 and the NSF CAREER Award in 2010. He is a co-recipient of the 2021 IEEE INTERNET OF THINGS JOURNAL Best Paper Award, the 2021 IEEE Communications Society Outstanding Paper Award, the 2021 Best Paper Award of Elsevier/KeAi Digital Communications and Networks Journal, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the IEEE ComSoc MMTC 2018 Best Journal Award and 2017 Best Conference Paper Award, the Best Demo Award of IEEE SECON 2017, the Best Paper Awards from IEEE GLOBECOM 2019, 2016, and 2015, IEEE WCNC 2015, and IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is an Associate Editor-in-Chief of IEEE/CIC CHINA COMMUNICATIONS, an Area Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, and ACM GetMobile, and an Associate Editor of IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE NETWORK, IEEE MULTIMEDIA, and IEEE NETWORKING LETTERS. He is a Distinguished Lecturer of the IEEE Communications Society and the IEEE Council of RFID.