2021 International Conference on Computer Communications and Networks (ICCCN) | 978-1-6654-1278-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCN52240.2021.9522256

MulTLoc: RF Hologram Tensor Filtering and Upscaling for Locating Multiple RFID Tags

Xiangyu Wang†, Jian Zhang‡, Shiwen Mao†, Senthilkumar CG Periaswamy‡, and Justin Patton‡
†Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
‡RFID Lab, Auburn University, Auburn, AL 36849, USA
Email: {xzw0042, jzz0043}@auburn.edu, smao@ieee.org, {szc0089, jbp0033}@auburn.edu

***Abstract*—In this paper, we present MulTLoc, a deep learning based indoor localization system for localizing multiple ultra-high frequency(UHF) passive RFID tags with RF hologram tensor filtering and upscaling. The proposed system leverages the RF hologram tensor as the input of the deep convolutional networks. The RF hologram tensor exhibits a strong relationship between the observation and the spatial location, which enhances the robustness of the system to the dynamic environment and equipment. To sanitize the RF hologram tensor, two architectures of deep networks are newly proposed. The hologram filter network suppresses the fake peaks resulting from the multipath and phase wrapping by leveraging the spatial relationship between tags. The tensor upscaling network recovers the high resolution hologram tensor from the output of the previous network, which enhances the localization accuracy of the system further. Comparing with the fingerprinting based localization systems using deep networks as the classifier, the networks in the MulTLoc system treat the localization problem as the regression problem, in which the ambiguity between fingerprints is reserved. To avoid the inherent errors in the fingerprinting based localization systems, the location estimation is given by intuitive peak finding algorithms using the recovered RF hologram tensor. We implement the proposed MulTLoc system with commodity RFID devices and verify its performance with extensive experiments.**

***Index Terms*—Radio-frequency identification (RFID), Ultra-high frequency (UHF) passive RFID tag, RF hologram tensor, Indoor localization, Deep learning (DL), Deep convolutional neural networks (DCNN)**

I. Introduction

Radio-frequency identification (RFID) is an automatic identification technology, which is capable of reading RFID tag data outside the line-of-sight. It has been widely implemented for many applications, such as supply chain management, inventory tracking, sports such as race timing systems, access control, toll collection system, and animal management. Recently, with the rapid development of the Internet of Things (IoT), the utilization of RFID technology has been extended to emerging areas including healthcare monitoring and environment sensing, due to its ubiquitous employment and low-cost tags. A growing number of functions and applications are attached to the existing RFID systems by leveraging the measurements in the RFID readings. RFID-based sensing systems, especially the systems for localization [1], gesture recognition [2], vital sign monitoring [3], [4], pose estimation [5], [6], temperature sensing [7], and material recognition [8], have attracted great interest from both industry and academia.

Among these existing and emerging applications, indoor localization has remained to be a hot research topic over the years, because it plays a fundamental role in solving position-related problems, such as gesture recognition and human pose estimation. The RFID-based localization system mainly relies on two measurements in the RFID readings, i.e., Received Signal Strength Indicator (RSSI) and phase angle. SpotOn leveraged RSSI along with a path loss model to perform trilateration for indoor localization [9]. LANDMARC collected RSSI readings from reference tags as fingerprints and estimated an unknown tag position by fingerprint matching [10]. However, due to the reflection and refraction of the RFID signals, the performance of the systems degrade significantly in the environment dominated by the multipath or non-line-of-sight paths.

The RFID phase angle also exhibits great sensitivity to environmental changes, including changes of the tag-antenna distance. Some recent applications have reached centimeter-level localization by estimating the direction of arrival (DoA) using the phase values in received RFID readings. SparseTag leveraged a spatial smoothing based method to utilize a novel sparse RFID tag array [1]. The median error of estimated angles of SparseTag is $1.831°$, while the corresponding median distance error is $5.012cm$. RF-Wear [11] achieved a mean error of 8-12° in tracking angles with a uniform linear array. Although the RFID tag array contributes to the high localization precision, it also restricts the real-time performance of the systems. Multiple rounds of interrogation have to be conducted to collect phase readings from all the tags in the array. Furthermore, RF-Kinect [12] introduced the body geometric model to the RF hologram to estimate the limb orientation and human joint position. With Kalman filter, the proposed system exhibited robust performance in tracking human gestures. However, the requirement of the reference tags also incurred extra computational cost.

Over recent years, deep neural networks have gained great interest and shown high promise in fields including computer vision and natural language processing. To utilize the outstanding classification performance of deep networks, researchers bring deep networks into indoor localization systems cooperating with the fingerprinting method. For example, deep autoencoders were leveraged to extract WiFi CSI features as fingerprints of the localization systems [13]–[16]. ResLoc improved the localization accuracy with a deep residual shar-

978-1-6654-1278-0/21/$31.00 ©2021 IEEE

ing learning model [17]. CiFi was the first work to employ a deep convolutional neural network (DCNN) for indoor localization [18]. The generated AoA image was used to train a 6-layer DCNN. Although the performance of such indoor localization system keeps improving with the iteration of deep networks, several inherent problems of fingerprinting based localization systems are still open. First, the minimum error of the fingerprinting based localization system relies on the distance between the stored fingerprints. To diminish the intrinsic error of the system, the number of fingerprints needs to be as large as possible. Apparently, it would be laborious or even impossible for some cases (e.g., a large area). Secondly, the fingerprinting method does not work well in dynamic environments. Any changes in the environment could force a fingerprint update. For the localization systems using DCNN, the entire network has to be trained from scratch. As a result, all of these inherent problems hinder the wide deployment of fingerprinting based localization systems.

To overcome these inherent problems of the fingerprinting method and take advantage of deep learning, we propose MulTLoc, a deep convolutional neural network based system for simultaneously localizing multiple Ultra-high frequency (UHF) passive RFID tags in a three-dimensional space. In the proposed system, radio frequency (RF) hologram tensors are generated with the phase readings from reader antenna pairs. Two deep network models, a hologram filter network and a tensor upscaling network, are proposed to filter and upscale the hologram tensors by eliminating the offsets resulted from the multipath and phase wrapping effects. Based on the processed RF hologram tensors, the location of multiple tags in a three-dimensional space will be inferred simultaneously with an intuitive peak detection algorithm. In the offline training phase, supervise training is leveraged for both deep networks. The ground truth tensor is generated with the ground truth coordinates measured by a computer vision sensor (i.e., a Kinect V2). In the online location estimation phase, the computer vision based sensor will not be needed in the system. The unknown locations are accurately estimated from the recovered RF hologram tensor only.

The main contributions made in this paper are summarized as follows.

- To the best of our knowledge, this is the first work to leverage RF hologram tensor to train deep networks for three-dimensional localization. The usage of the RF hologram tensor makes the deep networks independent to the changes in the environment, thus greatly enhancing the robustness and the transferability of the proposed system.
- We design two novel deep networks for sanitizing and adjusting the size of RF hologram tensors. In the hologram filter network, the spatial information between multiple tags is exploited to suppress the fake peaks existing in the original RF hologram tensors. In the tensor upscaling network, residual learning and pixel shuffle are leveraged to adjust the compressed hologram tensor to the desired size. Based on the recovered hologram tensor, location estimation becomes a simple peak detection problem, which can be easily accomplished.
- The proposed MulTLoc system is implemented with commodity RFID devices. The performance of the proposed system is evaluated by a multiple-joint localization experiment. The experimental results demonstrate that the MulTLoc system is effective on simultaneously localizing multiple tags in a three-dimensional space.

The remainder of this paper is organized as follows. Section II introduces the preliminaries and motivation our approach. We present the MulTLoc design in Section III and our experimental study in Section IV. Then, Section V concludes this paper.

II. Preliminaries and Motivation

A. RFID Phase Model

In order to locate RFID tags in real-time, sensitive and reliable measurements should be extracted from the original RFID readings. Compared with RSSI, the phase value has been widely used in many RFID based sensing applications [7], [19], [20]. As shown in (1), the phase reading $\theta_{i,m}$ is a periodic function with a period of 2π.

$$\theta_{i,m} = \bmod\left(\frac{4\pi\,|TA_m|}{\lambda_i} + \theta_{tag} + \theta_{equipment}, 2\pi\right), \quad (1)$$

where $|TA_m|$ denotes the distance between the tag T and the antenna A_m; λ_i is determined by the frequency of channel i; θ_{tag} and $\theta_{equipment}$ are the phase offsets caused by the RFID tag and RFID hardware such as antenna and reader, respectively. Usually $\theta_{equipment}$ is a constant for a given RFID system; hence it could be removed conveniently.

B. Hologram Tensor

The concept of RF hologram is firstly introduced in Tagoram [21]. The basic idea behind RF hologram is to calculate the similarities between the theoretical phase values and the measured phase values for each grid in the surveillance space. In our system, we leverage the phase difference as the observation to eliminate the tag-related phase offset, i.e., θ_{tag} in (1). The real phase difference obtained with the phases collected from an antenna pair (m, n) on channel i is denoted as

$$p_{i,m,n} = \bmod\left(\theta_{i,m} - \theta_{i,n}, 2\pi\right). \quad (2)$$

The theoretical phase difference between antenna pair (m, n) on channel i, can be computed when the positions of the two antennas are known. For the antenna pair (m, n), the theoretical phase difference at the the grid location, $G_{x,y,z}$, is depicted as

$$q^{x,y,z}_{i,m,n} = \bmod\left(\frac{4\pi\,|G_{x,y,z}A_m|}{\lambda_i} - \frac{4\pi\,|G_{x,y,z}A_n|}{\lambda_i}, 2\pi\right). \quad (3)$$

With the real and theoretical phase differences, their similarity, $S_{x,y,z}$, is estimated as follows.

$$S_{x,y,z} = \sum_{(M,N)} \sum_{I} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\delta^{x,y,z}_{i,m,n})^2}{2\sigma^2}\right)$$
$$\delta^{x,y,z}_{i,m,n} = \bmod\left(p_{i,m,n} - q^{x,y,z}_{i,m,n}, 2\pi\right), \quad (4)$$

where (M, N) represents the set consisting of all available antenna pairs, I denotes the set of all available channel indices. The hologram tensor, $\mathbf{S}$, is constructed as

$$\mathbf{S} = \begin{bmatrix} S_{1,1,z} & S_{1,2,z} & \cdots & S_{1,y,z} \\ S_{2,1,z} & S_{2,2,z} & \cdots & S_{2,y,z} \\ \vdots & \vdots & \ddots & \vdots \\ S_{x,1,z} & S_{x,2,z} & \cdots & S_{x,y,z} \end{bmatrix}, z = 1, 2, ..., Z, \quad (5)$$

where each element is scaled to have a value in $[0, 1]$ in the proposed system.

C. Motivation

To the best of our knowledge, MulTLoc is the first work using hologram tensors for training deep learning models for real-time three-dimensional localization. Although some indoor localization schemes, e.g., [18], [22], [23], utilized radio frequency signals to generate images or tensors for offline training, the generated data may not contain a strong relationship between the observation and the spatial location. These applications exploit the images and tensors as fingerprints and the deep networks as classifiers. The ambiguity between fingerprints may be lost when constructing the dataset, which also restricts the transferability of the localization model. The localization model has to be updated when the surveillance space or hardware are changed. Compared with images and tensors used in the previous works, the hologram tensor is interpretable. In our MulTLoc system, the hologram tensors represent how likely a tag is located at a grid position in the surveillance space. The similarity S in the hologram tensor is closely related to the distances between the tag and antennas, and more important, it is robust to environment changes.

Fig. 1 presents a hologram matrix generated in a two-dimensional area. Fig. 1(a) plots the two-dimensional projection of the hologram matrix, while Fig. 1(b) is the corresponding three-dimensional mesh of the hologram matrix. The red pentagram denotes the actual location of the target tag. As we can see, there is a peak existing at the location of the ground truth. However, due to the multipath and phase wrapping effects, many fake peaks are produced and spread out in the hologram. Some of the fake peaks even have a higher similarity value. To avoid such problems, data prepossessing has been an essential part of many RFID-based sensing applications. Some approaches improve localization accuracy at the price of real-time performance. For example, channel selection [1] and phase sanitation [24] are leveraged to keep the systems away from those phase readings contaminated by the multipath effect. However, such approaches may be infeasible for real-time localization systems. This is due to the requirement for multiple-round interrogations, while the tag (or, target) would not keep stationary until the system makes a sufficiently large number of interrogations. Moreover, some applications rely on specific hardware and deployment, such as the synthetic-aperture array [25] and multi-resolution filtering [26], to mitigate the negative effect resulted from the ambiguity related to phase wrapping. Even though these

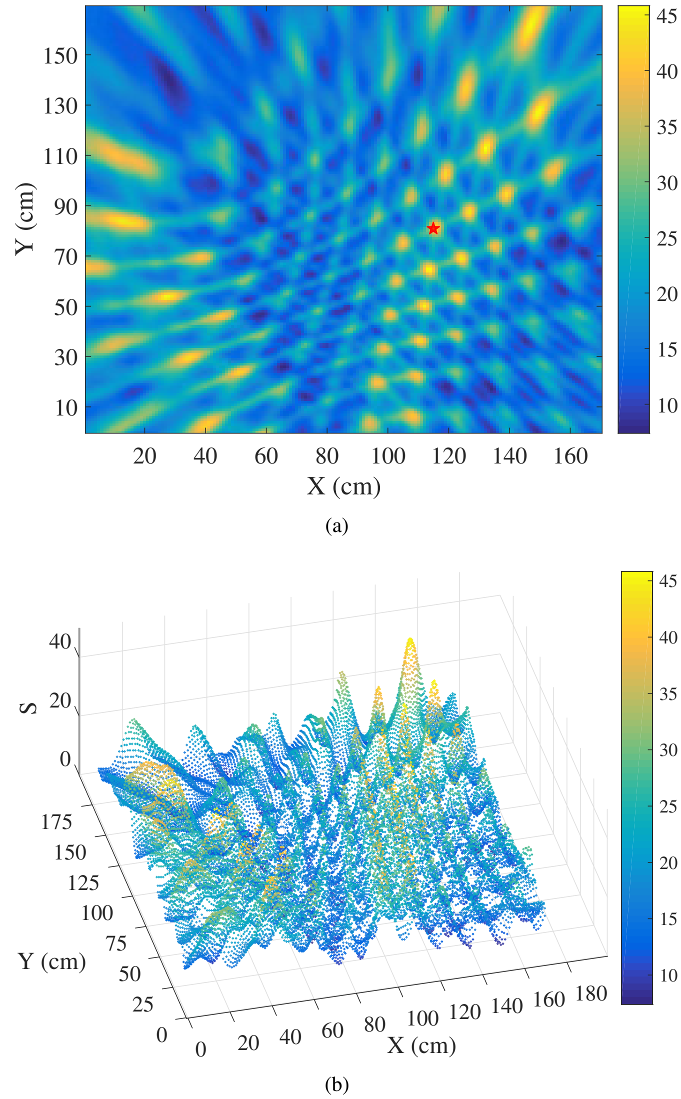


Fig. 1. Hologram in a two-dimensional scenario. The red pentagram denotes the ground truth. (a) The two-dimensional projection of the hologram matrix. (b) The three-dimensional mesh of the hologram matrix.

approaches achieve acceptable precision and real-time performance, the requirement for the special hardware incurs higher costs and limits the compatibility with COTS RFID systems. Furthermore, tag localization in a three-dimensional space is a more challenging problem than the two-dimensional case. To address such problems, two novel convolutional network architectures are proposed for the MulTLoc system.

III. Overview of the MulTLoc System

In this paper, we propose an RFID based localization system, termed MulTLoc, for estimating the location of multiple tags *simultaneously* using noisy hologram tensors. Even though MulTLoc, like most previous deep learning based localization systems, is trained with ground truths provided by sensors such as an RGB-D camera, the problem of localization is treated as *regression* in this work instead. Traditionally, the method of fingerprinting combined with deep learning based *classification* algorithms is utilized to predict the coordinates of unknown locations. The localization accuracy is constricted by the size of the fingerprint database, and the granularity of the fingerprints determines the inherent error of the system. In the MulTLoc system, location estimation would not be provided by the network instantly. Instead, noisy hologram tensors are regressed to single peak hologram tensors that exclude the fake peaks caused by the multipath and phase wrapping effects. Based on the filtered hologram tensor, location estimation could be accomplished intuitively.

A. MulTLoc System Architecture

The MulTLoc system architecture is illustrated in Fig. 2. To train the networks for location estimation, an RFID system cooperates with a vision-based sensor for generating the hologram tensors and the corresponding ground truth tensors. Since the hologram tensors and the ground truth coordinates provided by the vision-based sensor are usually in different coordinate systems, the Robot Operating System (ROS) is leveraged in our proposed system to synchronize and unify the data collected from different hardware. Data augmentation is also implemented to expand the training dataset to mitigate the overfitting problem.

The MulTLoc system consists of two deep networks, i.e., the hologram filter network and the tensor upscaling network, for recovering noisy-free, high-resolution hologram tensors. Both networks are supervised by the tensors generated with the ground truth coordinates from the vision-based sensor. Once the networks of the MulTLoc system are trained correctly, the vision-based sensor will not be needed for location estimation. To estimate the location of a tag, new hologram tensors from the RFID system are filtered and upscaled by the deep networks. Based on the output of the deep networks, the location estimation is accomplished with an intuitive peak detection algorithm.

B. Training Dataset Generation

To properly train the MulTLoc system, the hologram tensor has to be labeled with the corresponding ground truth tensor. However, the ground truth coordinates and the hologram tensors are collected by different sensors and have different coordinate systems. For most vision-based sensors, the reported coordinates are usually decided by the coordinate origin of the sensor space. For example, the center of the depth sensor is the origin of the coordinates for Kinect V2, while the coordinates of the antennas in MulTLoc are decided by the surveillance space. To label hologram tensors with correct ground truth tensors, ROS is utilized in MulTLoc to integrate the hologram tensors from the RFID system and the coordinates from the vision-based sensor. Considering the simplicity of the system, we transfer all coordinates from the vision-based sensor into the frames of the hologram tensors based on the pose and location of the sensor in the surveillance space. Meanwhile, timestamps are attached to both the hologram tensors and the ground truth coordinates for synchronization. An RF hologram

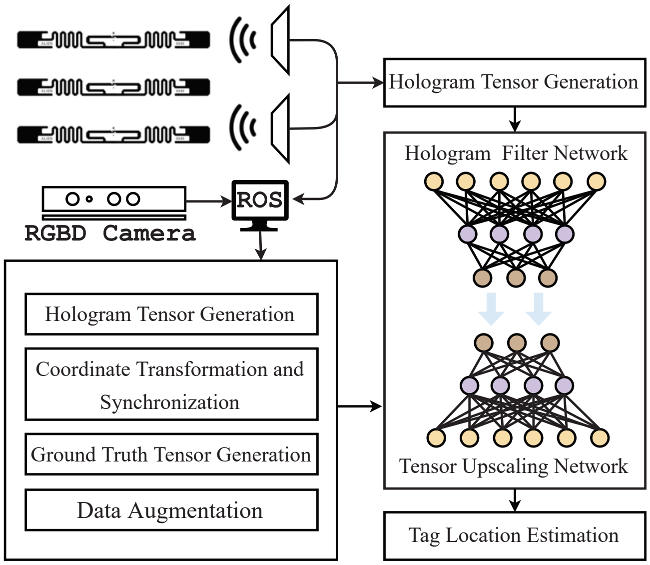


Fig. 2. The MulTLoc system architecture.

tensor will be paired with the coordinates that have the closest timestamp.

Based on the synchronized ground truth coordinates, the ground truth tensor, $\mathbf{K}$, is created with a gaussian kernel by measuring the Euclidean distance $|G_{x,y,z}H|$ between the gird location $G_{x,y,z}$ and the ground truth location H as

$$\mathbf{K} = \begin{bmatrix} K_{1,1,z} & K_{1,2,z} & \cdots & K_{1,y,z} \\ K_{2,1,z} & K_{2,2,z} & \cdots & K_{2,y,z} \\ \vdots & \vdots & \ddots & \vdots \\ K_{x,1,z} & K_{x,2,z} & \cdots & K_{x,y,z} \end{bmatrix}, z = 1, 2, ..., Z, \quad (6)$$

where each element of $\mathbf{K}$ is given by

$$K_{x,y,z} = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{|G_{x,y,z}H|^2}{2\epsilon^2}\right). \quad (7)$$

$\mathbf{K}$ supervises the training of the two deep networks in the MulTLoc system, including the hologram filter network and the tensor upscaling network. The ground truth tensor $\mathbf{K}$ is downsampled in the hologram filter network, while it keeps in the original size when the tensor upscaling network is trained. We apply data augmentation in the proposed MulTLoc system. Since the hologram tensor is interpretable spatially, our training dataset is augmented by the flipping and rotating operations.

C. Network Design for Filtering Hologram Tensors

We propose a hologram filter network, as shown in Fig. 3, for sanitizing the hologram tensors from the fake peaks. Unlike the dynamic environments that degrade the performance of fingerprinting-based localization systems, the positional relationship between tags is relatively stable, especially for the passive tags stuck on products. The hologram filter network is designed to learn the spatial relationship between tags to distinguish the actual peaks in the RF hologram tensors.

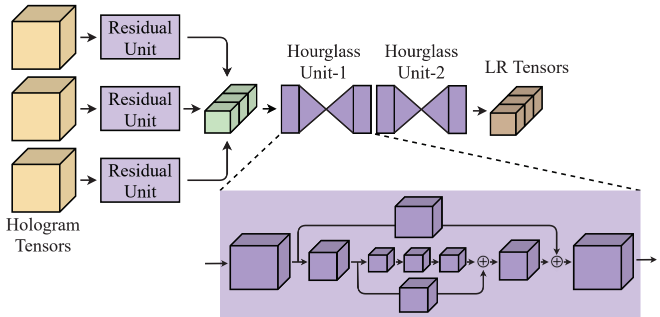


Fig. 3. Architecture of the hologram filter network.

Incorporating residual units, we downsample the hologram tensors from n tags and concatenate them into a n-channel tensor for compressing the number of weights in the proposed network and accelerating training. The newly generated n-channel tensor preserves the detailed information in the original hologram tensors and includes a coherent understanding among the tags. In our initial experiments, n is set to three for locating three tags simultaneously.

The residual unit in the hologram filter network includes three residual blocks. In each block, two three-dimensional convolutional layers are connected in a sequence [27]. As the backbone of the hologram filter network, the hourglass blocks are placed together end-to-end following the residual blocks to extract features in the n-channel tensor at different scales [28]. As is depicted in Fig. 3, the architecture of the hourglass unit is similar to an encoder-decoder network. The input tensor is first compressed and then upscaled in the unit. Each purple cube in the hourglass unit is composed of three residual blocks that include three three-dimensional convolutional layers. The skip connection is leveraged between the blocks with the same size to preserve spatial information at different resolutions. By stacking the hourglass units, the bottom-up, top-down inference is performed repeatedly.

In the hologram filter network, intermediate supervision is implemented at each hourglass unit for accelerated training. The output of the hologram filter network is a low resolution, n-channel tensor (i.e., the LR Tensor), which will be separated into n low resolution hologram tensors to be processed by the tensor upscaling network.

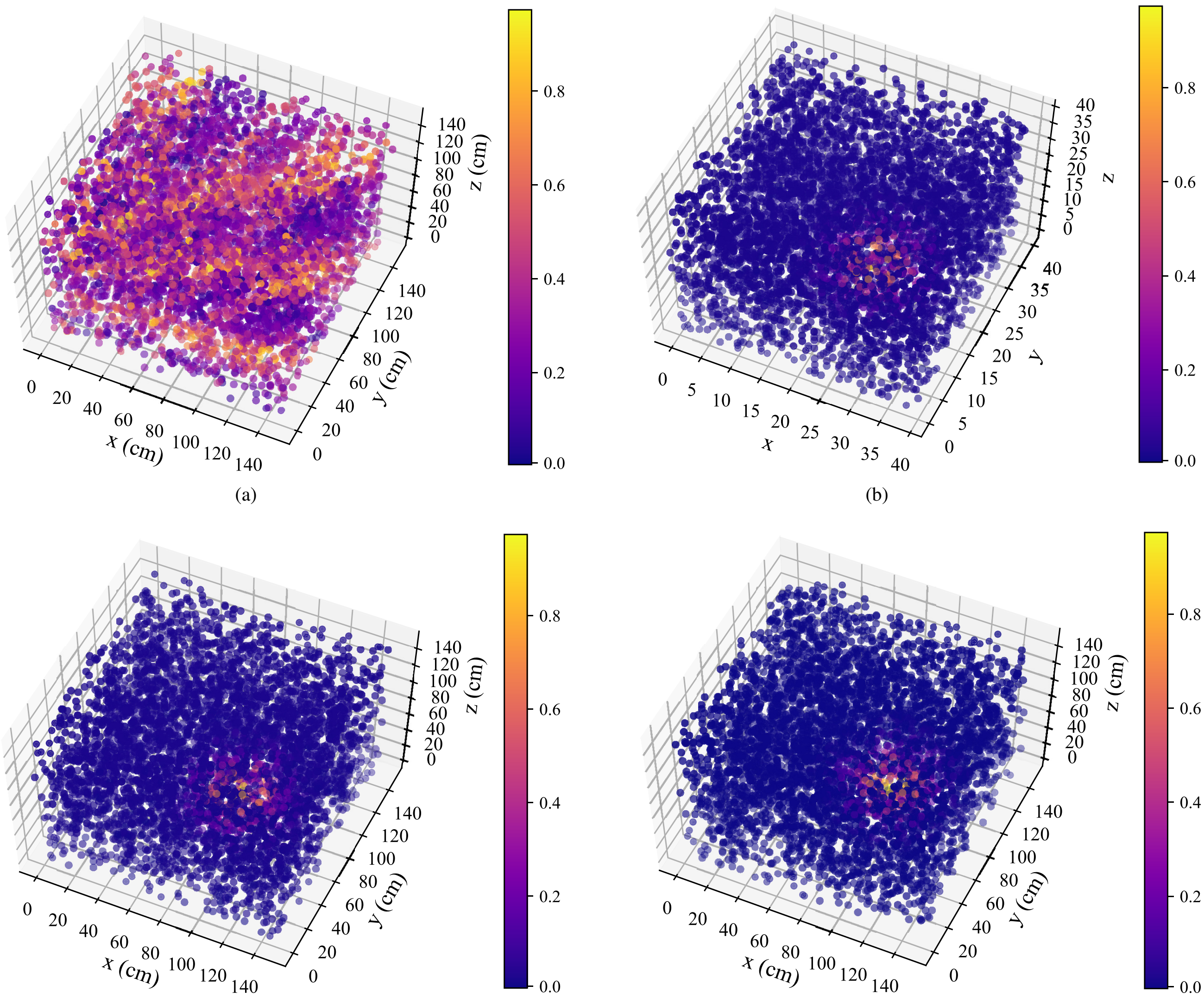
Fig. 4(a) and Fig. 4(b) illustrate the input and output of the hologram filter network, respectively. The bluish pixels in the figures denote the lower similarity values. As shown in Fig. 4(a), the RF hologram tensor is generated with the phases collected from our testbed, which covers a space of dimension $1.5m \times 1.5m \times 1.5m$ (see Section IV-A for details). Similar to the hologram matrix in the two-dimensional scenario, the fake peaks spread out in the hologram tensor. Although there are four bands composed of higher similarity values in the space, there is no obvious peak that can be identified. By combining the holograms from three tags, the sanitized hologram tensor, depicted in Fig. 4(b), is generated by the hologram filter network. Most of the fake peaks in the input tensor are now suppressed. The only bright area locates at the middle part of the filtered low resolution hologram tensor. The location information hidden below fake peaks is now recovered by the spatial relationship among the tags. To adjust the size of the filtered low resolution hologram tensor, the tensor upscaling network is adopted in MulTLoc, which will be presented in the following section.

D. Network Design for Upscaling Hologram Tensors

Considering that the size of the hologram tensor is compressed by the tensor filter network, we propose a tensor upscaling network for adjusting the size of the filtered low resolution hologram tensor and improving the precision of the system. Unlike the hologram filter network, the task of the tensor upscaling network is independent of the spatial relationship among tags. To train the tensor upscaling network, the n-channel LR tensors from the hologram filter network are break down into single-channel LR tensors, which are labeled with the corresponding full-size ground truth tensors.

The tensor upscaling network in the MulTLoc system is largely inspired by the deep networks for single image super-resolution (SISR) in computer vision [29]–[31], exploring how to extract useful information from low-resolution data and recover the tensor into the original size. Compared with the images of everyday life, the elements in the LR tensor are stable and more predictable. Thus, the architecture of the tensor upscaling network is simplified significantly compared to the models used in computer vision.

A lightweight network will contribute to fast forward propagation, which is a key factor for real-time localization. As depicted in Fig. 5, residual-learning is utilized in the tensor upscaling network. The trilinear interpolate copy of the LR tensor is passed through all layers until it reaches the output layer. For the main branch in the tensor upscaling network, we implement two Pixel Shuffle Units (PSUnit) to extract and recover the high-resolution residual components from the LR tensor. The subpixel convolution layer is the core of the unit. In the subpixel convolution layer, extra feature maps are generated with the regular convolution kernels, and then the upscaled output is obtained by resizing the feature maps. For example, the regular convolution kernel expands a tensor with the size of $C \times D \times H \times W$ into the one of $r^3 \cdot C \times D \times H \times W$, which is rearranged into an upscaled tensor with the size of $C \times rD \times rH \times rW$. Compared with traditional upscaling methods, such as unpooling and upsampling, the subpixel convolution is flexible and learnable. The values in the upscaled tensor inherit the information from the upper level and are not generated by duplicating and interpolating. On the other hand, the subpixel convolution will not introduce meaningless padding values, which are common in the transposed convolution. Such values have to be updated later and require extra computation. Before the output layer, the residual component from the main branch merges with the trilinear interpolate copy of the input tensor to recover the High-Resolution hologram Tensor (HR Tensor).



(a) (b) (c) (d)

Fig. 4. The hologram tensors: (a) The original hologram tensors for training the hologram filter network; (b) The filtered low resolution hologram tensor produced by the hologram filter network; (c) The recovered high resolution hologram tensor by the tensor upscaling network; (d) The full-size ground truth tensor.

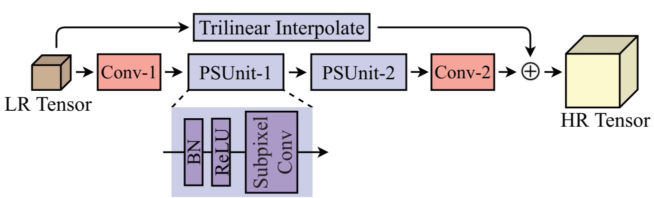


Fig. 5. Architecture of the tensor upscaling network.

In Fig. 4(c), the high-resolution hologram tensor recovered by the tensor upscaling network is plotted. Comparing with the tensor shown in Fig. 4(b), we can see that the peak area in Fig. 4(c) is not stretched linearly in the tensor upscaling network, even though the recovered hologram tensor is built with the trilinear interpolation. More high-similarity values now cluster in a small region around the ground truth coordinates. The network we proposed upgrades the performance of the trilinear interpolation by introducing the residual components, which are achieved by subpixel convolution. Moreover, it is obvious that the peak in Fig. 4(c) almost overlaps with the peak in Fig. 4(d), which demonstrates that the tensor upscaling network is capable of recovering full-size high resolution hologram tensors from the low resolution hologram tensors. We will examine the performance of the proposed system in the next section, and the performance analysis would be presented in Section IV-B.

IV. Experimental Study

A. Testbed Configuration

To evaluate the performance of the proposed system, we prototype the MulTLoc system with a Zebra FX9600 reader

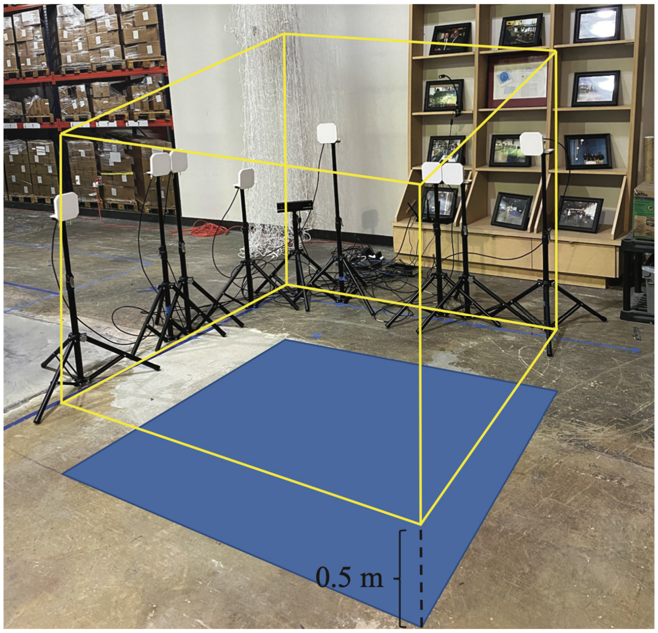


Fig. 6. The MulTLoc testbed setup.

equipped with eight Zebra AN720 antennas. Three UPM Raflatac Frog 3D tags are used as targets to be localized. In the experiment, we evaluate the performance of the system by simultaneously localizing the tags attached to the human body. The ground truth coordinates for supervised learning are provided by a Kinect V2 device cooperating with a three-dimensional human pose estimation algorithm [32], which achieves higher accuracy than the Kinect SDK in human keypoint localization. The target tags are attached to the shoulders and neck for dataset generation and tag position estimation. We use the ROS Kinetic Kame to synchronize and unify the coordinates and tensors from the Kinect V2 and the RFID reader. To guarantee real-time performance of the MulTLoc system, we adjust the requirement for hologram tensor generation. The phases from seven channels will be utilized to generate the hologram tensors when five antenna pairs are available. As shown in Fig 6, the yellow lines delineate the surveillance space of the MulTLoc system, which covers a space of dimension $1.5m \times 1.5m \times 1.5m$ at $0.5m$ above the ground. The grid size in the space is set to $1cm$. Furthermore, the similarities at each grid location in the surveillance space are calculated in parallel using CUDA GPU programming to accelerate the generation of hologram tensors and the ground truth tensors.

To train the deep networks in the MulTLoc system, we collect tensors and the corresponding coordinates from multiple volunteers who are attached with three tag and move and pose randomly in the surveillance space. No tensor labeled by duplicate coordinates is included in the collected data. Totally, we obtained three hundred groups of data. Each group includes two tensors from the shoulder tags and one tensor from the neck tag. The collected data are separated randomly for training, validating, and testing. Among them, 80% of the groups are leveraged for training the deep networks. To avoid overfitting and enhance the generalization ability of the networks, the training dataset is augmented by flipping, horizontal rotation, and vertical rotation. In total, seven hundred and twenty RF hologram tensors are included in the training dataset, while the rest sixty tensors are evenly partitioned into the validation dataset and the testing dataset. By computing the mean squared error (MSE) loss between the ground truth tensors and the output tensors, the deep networks in the MulTLoc system are optimized with the Adam algorithm. Furthermore, early stopping is adopted in both networks to avoid overfitting. The estimation location $\widehat{G}$ is computed as follows.

$$\widehat{G} = \{G | f(\mathbf{S}_R, G) = \max(\mathbf{S}_R)\}, \tag{8}$$

where $f(\cdot)$ extracts the similarity value at the grid location G from the high-resolution hologram tensor $\mathbf{S}_R$. We leverage an Nvidia RTX3090 GPU to accelerate the computation of the two deep networks.

B. Experiment Results and Discussions

Fig. 7 presents the mean distance errors of the three tags and the overall average error. We compare two scenarios that affect the location estimation: with and without the tensor upscaling network. As we can see, a better location estimation is achieved when the tensor upscaling network is implemented in the system. The lowest mean distance error for the neck tag reaches $5.11cm$, which outperforms the error obtained with the trilinear interpolation. The same situation happens to the errors of shoulder tags. The tensor upscaling network benefits the location estimation for both shoulders. Although the location estimation of the neck tag is more accurate than those of the shoulder tags, the distance errors remain at the same level overall. On the whole, the mean distance error obtained with the tensor upscaling network is $6.29cm$, which is lower than the one from the hologram tensor recovered by the trilinear interpolation technique.

Fig. 8 presents the cumulative distribution function (CDF) of distance errors across all of the test tensors with and without the tensor upscaling network. The maximum distance errors for both scenarios are about $15cm$, which are acceptable for a space of $1.5m \times 1.5m \times 1.5m$. This result demonstrates that the hologram filter network has suppressed most of the offsets resulted from the multipath and phase wrapping effects. In addition, the median distance error obtained when the tensor upscaling network is in place is $6.32cm$, while a median distance error of $6.63cm$ is reached with the trilinear interpolation. This comparison shows that the tensor recovered by the tensor upscaling network contains more precise information. It would satisfy the applications' requirement of high accuracy localization. However, the choice of the upscaling method provides a trade-off. The applications that rely on real-time performance and use equipment with limited computational capability may benefit from the trilinear interpolation technique. Overall, the proposed MulTLoc system successfully

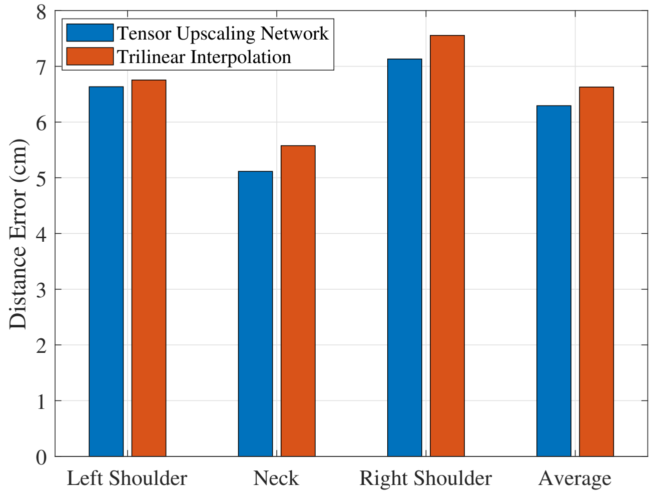


Fig. 7. Location estimation errors of the three tags with and without the tensor upscaling network.

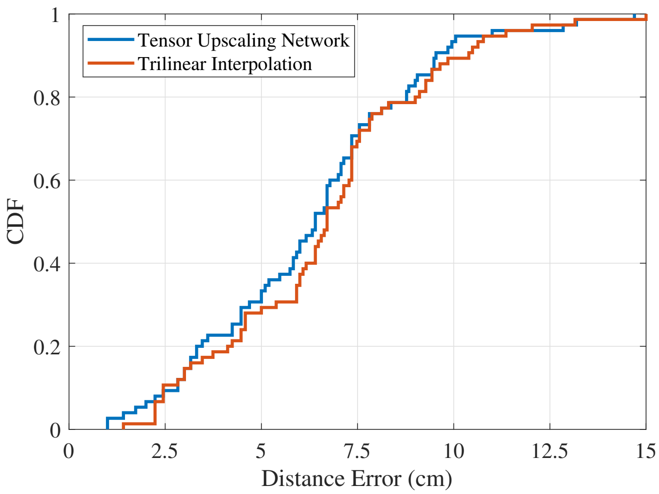


Fig. 8. CDFs of location estimation errors of the three tags with and without the tensor upscaling network.

achieves the goal of simultaneously localizing multiple tags in a three-dimensional space.

V. Conclusions

In this paper, we presented MulTLoc, a multiple-tag simultaneous localization system that utilizes deep networks for RF hologram tensor filtering and upscaling. To the best of our knowledge, this is the first work to leverage hologram tensor to train deep networks for three-dimensional RFID tag based indoor localization. We proposed two deep learning models to be incorporated in the MulTLoc system. The hologram filter network was designed to eliminate the effect of multipath propagation and phase wrapping by referring the spatial relationship among the tags. Next, the resolution of the hologram tensor was adjusted by a hologram upscaling network, which improved the precision of the system. We evaluated the proposed system with a multiple-joint location estimation application. The result demonstrated the superior performance of the proposed MulTloc system.

Acknowledgments

This work was supported in part by the NSF under Grant ECCS-1923163, the RFID Lab, and the Wireless Engineering Research and Education Center (WEREC) at Auburn University, Auburn, AL, USA.

References

[1] C. Yang, X. Wang, and S. Mao, “SparseTag: High-precision backscatter indoor localization with sparse RFID tag arrays,” in *Proc. IEEE SECON’19*, Boston, MA, June 2019, pp. 1–9.

[2] S. Pradhan *et al.*, “RIO: A pervasive RFID-based touch gesture interface,” in *Proc. ACM MobiCom’17*, New York, NY, Oct. 2017, pp. 261–274.

[3] C. Yang, X. Wang, and S. Mao, “Respiration monitoring with RFID in driving environments,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 500–512, Feb. 2021.

[4] C. Yang, X. Wang, and S. Mao, “Unsupervised detection of apnea using commodity RFID tags with a recurrent variational autoencoder,” *IEEE Access Journal*, vol. 7, no. 1, pp. 67 526–67 538, June 2019.

[5] C. Yang, X. Wang, and S. Mao, “RFID-Pose: Vision-aided 3D human pose estimation with RFID,” *IEEE Transactions on Reliability*, to appear. DOI: 10.1109/TR.2020.3030952.

[6] C. Yang, X. Wang, and S. Mao, “Subject-adaptive skeleton tracking with RFID,” in *Proc. The 16th IEEE International Conference on Mobility, Sensing and Networking (MSN 2020)*, Tokyo, Japan, Dec. 2020.

[7] X. Wang, J. Zhang, Z. Yu, S. Mao, S. Periaswamy, and J. Patton, “On remote temperature sensing using commercial UHF RFID tags,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 715–10 727, Dec. 2019.

[8] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, “Simultaneous material identification and target imaging with commodity RFID devices,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 739–753, Feb. 2021.

[9] J. Hightower, R. Want, and G. Borriello, “SpotON: An indoor 3D location sensing technology based on RF signal strength,” UW CSE Technical Report, 2000.

[10] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, “LANDMARC: Indoor location sensing using active RFID,” in *Proc.IEEE PerCom’03.*, Dallas, TX, Mar. 2003, pp. 407–415.

[11] H. Jin, Z. Yang, S. Kumar, and J. I. J. I Hong, “Towards wearable everyday body-frame tracking using passive RFIDs,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. Article No.: 145, Jan. 2018.

[12] C. Wang, J. Liu, Y. Chen, L. Xie, H. Liu, and S. Lu, “RF-kinect: A wearable RFID-based approach towards 3D body movement tracking,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. Article No: 41, Mar. 2018.

[13] X. Wang, L. Gao, S. Mao, and S. Pandey, “DeepFi: Deep learning for indoor fingerprinting using channel state information,” in *Proc. IEEE WCNC’15*, New Orleans, LA, Mar. 2015, pp. 1666–1671.

[14] X. Wang, L. Gao, S. Mao, and S. Pandey, “CSI-based fingerprinting for indoor localization: A deep learning approach,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[15] X. Wang, L. Gao, and S. Mao, “PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach,” in *Proc. GLOBECOM’15*, San Diego, CA, Dec. 2015, pp. 1–6.

[16] X. Wang, L. Gao, and S. Mao, “CSI phase fingerprinting for indoor localization with a deep learning approach,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.

[17] X. Wang, X. Wang, and S. Mao, “Indoor fingerprinting with bimodal CSI tensors: A deep residual sharing learning approach,” *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4498–4513, Mar. 2021.

[18] W. Wang, X. Wang, and S. Mao, “Deep convolutional neural networks for indoor localization with CSI images,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 316–327, Jan./Mar. 2020.

[19] X. Wang, C. Yang, and S. Mao, “TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi,” *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 1, pp. 8:1–8:27, Sept. 2017.

[20] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S. Periaswamy, J. Patton, and X. Wang, "RFHUI: An RFID based human-unmanned aerial vehicle interaction system in an indoor environment," *Elsevier Digital Communications and Networks Journal*, vol. 6, no. 1, pp. 14–22, Feb. 2020.
[21] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proc. ACM Mobicom'14*, Maui, HI, Sept. 2014, pp. 237–248.
[22] X. Wang, X. Wang, and S. Mao, "ResLoc: Deep residual sharing learning for indoor localization with CSI," in *Proc. IEEE PIMRC'17*, Montreal, Canada, Oct. 2017, pp. 1–6.
[23] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, no. 1, pp. 18 066–180 747, Sept. 2017.
[24] G. Wang, C. Qian, K. Cui, X. Shi, H. Ding, W. Xi, J. Zhao, and J. Han, "A universal method to combat multipaths for RFID sensing," in *Proc. IEEE INFOCOM'20*, Toronto, Canada, July 2020, pp. 277–286.
[25] L. Shangguan and K. Jamieson, "The design and implementation of a mobile RFID tag sorting robot," in *Proc. ACM MobiSys'16*, Singapore, June 2016, pp. 31–42.
[26] J. Wang, D. Vasisht, and D. Katabi, "Rf-idraw: Virtual touch screen in the air using rf signals," *ACM SIGCOMM'14*, pp. 235–246, Aug. 2014.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR'16*, Las Vegas, NV, June 2016, pp. 770–778.
[28] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *ACM ECCV'16*, Amsterdam, The Netherlands, Oct. 2016, pp. 483–499.
[29] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE CVPR'16*, Las Vegas, NV, June 2016, pp. 1646–1654.
[30] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE CVPR'16*, Las Vegas, NV, June 2016, pp. 1637–1645.
[31] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE CVPR'16*, Las Vegas, NV, June 2016, pp. 1874–1883.
[32] C. Zimmermann, T. Welschehold, C. Dornhege, W. Burgard, and T. Brox, "3D human pose estimation in RGBD images for robotic task learning," in *Proc. IEEE ICRA'18*, Brisbane, Australia, May 2018, pp. 1986–1992.