Cooperative ISAC for Localization and Velocity Estimation Using OFDM Waveforms in Cell-Free MIMO Systems

Zihuan Wang and Vincent W.S. Wong
Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
email: {zihuanwang, vincentw}@ece.ubc.ca

***Abstract*—In this paper, we present a cooperative integrated sensing and communication (ISAC) framework in cell-free multiple-input multiple-output (MIMO) systems, where multiple access points (APs), under the control of a central processing unit (CPU), collaboratively perform target sensing by using the reflected echo signals. Most of the existing works first estimate the sensing parameters (e.g., range, angle, relative velocity) observed by each AP and then use these estimated parameters for sensing tasks such as localization and velocity estimation. However, this approach may suffer from performance degradation due to errors in the estimated parameters. We propose a deep neural network (DNN)-based scheme to jointly process the echo signals received across the distributed APs and directly estimate the location and velocity of the targets. The proposed scheme bypasses the sensing parameter estimation stage and enhances the sensing performance. Simulation results show that our proposed scheme significantly reduces the localization and velocity estimation error when compared with a state-of-the-art approach.**

***Index Terms*—Cell-free MIMO, cooperative ISAC, localization and velocity estimation.**

I. INTRODUCTION

Future wireless networks are expected to support ubiquitous communications and pervasive sensing. Integrated sensing and communication (ISAC) has become a pivotal technology for achieving this goal [1], [2]. By enabling the integration of both functions using shared resources, ISAC has demonstrated great potential in various applications, including Internet of Things (IoT) and vehicular networks [3], [4]. Orthogonal frequency division multiplexing (OFDM) is a widely used waveform in ISAC systems, which can combat frequency-selective fading and provide high data rates [5]. OFDM waveform also exhibits Doppler tolerance and do not suffer from range-Doppler coupling, which makes it suitable for radar sensing [6], [7].

Unlike the conventional OFDM radar systems, where the transmitted signals do not carry useful information, OFDM signals used in ISAC systems contain communication data with specific modulations, which introduce different phase shifts in the transmitted signals. The impact of communication data should be taken into consideration when extracting sensing parameters from the reflected echo signals [8], [9]. In [8], an OFDM-ISAC system is proposed for range-velocity estimation, in which the communication data within the received echo signals is first compensated. A super-resolution method is proposed for range and velocity estimation using the compensated results. In [9], the sensing channel frequency response is first estimated, with the communication information being removed. A deep learning-based algorithm is applied to extract the range and relative velocity from the estimated frequency response. The studies in [8], [9] consider a single-input single-output (SISO) scenario, whereas in practice, multiple-input multiple-output (MIMO) architectures are widely employed at the base station to enhance spectral efficiency through beamforming. MIMO architectures provide additional degrees-of-freedom (DoFs) in the spatial domain, allowing the extraction of angle-range-velocity parameters of targets from the reflected echo signals [10], [11].

The aforementioned works consider a single base station for mono-static sensing, which may result in limited sensing performance due to restricted spatial diversity. To address this issue, cooperative multi-static sensing is desired, where multiple access points (APs) are deployed distributively in a coverage area to jointly collect multi-view sensing observations. The APs are connected to a central processing unit (CPU), which facilitates collaboration between the APs. In [12], cooperative ISAC in cell-free MIMO systems are considered. The sensing parameters (e.g., range, angle, relative velocity) observed by each AP are estimated using a compressive sensing algorithm. After that, the location and velocity of the target can be estimated based on the estimated sensing parameters [13]. However, this approach estimates the sensing parameters observed at each AP independently, without exploiting the benefits of joint processing across different APs. Moreover, this approach may suffer from performance loss because the estimation error of sensing parameters can degrade the accuracy of the subsequent sensing tasks.

In this paper, we propose a deep neural network (DNN)-based scheme, which processes the received echo signals across distributed APs jointly. The developed DNN-based scheme directly estimates the location and velocity of the target using the reflected echo signals and bypasses the intermediate step of sensing parameter estimation. Simulation results demonstrate that the proposed DNN-based scheme provides a better sensing performance when compared with a state-of-the-art approach from [12].

II. COOPERATIVE ISAC IN CELL-FREE MIMO

Consider a cell-free MIMO system with N transmit APs and M receive APs. All APs are connected to a CPU via fronthaul links and they are fully synchronized. We consider there are K single-antenna communication users and Q point-like targets to be sensed in the area of interest. We consider the transmit and receive APs are equipped with uniform planar arrays (UPAs). For a half-wavelength-spaced UPA, the vertical and horizontal steering vectors are respectively given by

$$\mathbf{u}_{\mathrm{v}}(A_{\mathrm{v}},\theta)=\frac{1}{\sqrt{A_{\mathrm{v}}}}\left[1\ e^{-j\pi\cos\theta}\ \cdots\ e^{-j(A_{\mathrm{v}}-1)\pi\cos\theta}\right]^{\mathrm{T}}, \quad (1)$$

$$\mathbf{u}_{\mathrm{h}}(A_{\mathrm{h}},\phi,\theta) =\frac{1}{\sqrt{A_{\mathrm{h}}}}\left[1\ e^{-j\pi\sin\theta\cos\phi}\ \cdots\ e^{-j(A_{\mathrm{h}}-1)\pi\sin\theta\cos\phi}\right]^{\mathrm{T}}, \quad (2)$$

where ϕ and θ denote the azimuth and elevation angles, respectively. A_{v} and A_{h} are the number of vertical and horizontal antennas of the UPA, respectively. The beam steering vector of a UPA can be expressed as $\mathbf{a}(A,\phi,\theta)\in\mathbb{C}^{A}$:

$$\mathbf{a}(A,\phi,\theta)=\mathbf{u}_{\mathrm{h}}(A_{\mathrm{h}},\phi,\theta)\otimes\mathbf{u}_{\mathrm{v}}(A_{\mathrm{v}},\theta), \quad (3)$$

where $A=A_{\mathrm{v}}A_{\mathrm{h}}$ is the total number of antennas. $\otimes$ denotes the Kronecker product operator. In this work, we consider each transmit AP is equipped with $N_{\mathrm{t}}=N_{\mathrm{v}}N_{\mathrm{h}}$ antennas, where N_{v} and N_{h} are the number of antennas in the vertical and horizontal dimensions, respectively. Similarly, each receive AP has M_{v} vertical antennas and M_{h} horizontal antennas, and the total number of antennas $M_{\mathrm{r}}=M_{\mathrm{v}}M_{\mathrm{h}}$.

Let f_{c} denote the carrier frequency and $\lambda_{\mathrm{c}}=c/f_{\mathrm{c}}$ denote the wavelength, where c is the speed of light. Let N_{s} denote the number of subcarriers and B denote the total bandwidth. The subcarrier interval is equal to $\Delta f=B/N_{\mathrm{s}}$. The OFDM symbol duration is given by $\Delta T=1/\Delta f+T_{\mathrm{p}}$ where T_{p} is the period of the cyclic prefix. Let $\mathbf{s}_i[t]=[s_{i,1}[t]\cdots s_{i,K}[t]]^{\mathrm{T}}\in\mathbb{C}^{K}$ denote the t-th transmitted OFDM symbol for the K users on the i-th subcarrier, where $i=0,\ldots,N_{\mathrm{s}}-1$, $t=1,\ldots,T_{\mathrm{s}}$, and T_{s} denotes the number of symbols. We assume each element in $\mathbf{s}_i[t]$ has unit power and the transmitted symbols are statistically independent, i.e., $\mathbb{E}\{\mathbf{s}_i[t]\mathbf{s}_i[t]^{\mathrm{H}}\}=\mathbf{I}_K$. The transmitted OFDM symbols are then precoded by $\mathbf{W}_{n,i}\triangleq[\mathbf{w}_{n,i,1}\cdots\mathbf{w}_{n,i,K}]\in\mathbb{C}^{N_{\mathrm{t}}\times K}$, where each column, i.e., $\mathbf{w}_{n,i,k}\in\mathbb{C}^{N_{\mathrm{t}}}$, is the precoder assigned to the n-th transmit AP for the transmission on the i-th subcarrier for the k-th user, $n=1,\ldots,N$, $i=0,\ldots,N_{\mathrm{s}}-1$, $k=1,\ldots,K$. Let $\mathbf{x}_{n,i}[t]\in\mathbb{C}^{N_{\mathrm{t}}}$ denote the transmitted signal assigned to the n-th transmit AP on the i-th subcarrier during the t-th OFDM symbol duration. It can be expressed as

$$\mathbf{x}_{n,i}[t]=\sum_{k=1}^{K}\mathbf{w}_{n,i,k}s_{i,k}[t]=\mathbf{W}_{n,i}\mathbf{s}_i[t]. \quad (4)$$

We define the transmit power of each transmit AP as $P=\sum_{i=0}^{N_{\mathrm{s}}-1}\|\mathbf{W}_{n,i}\|_F^2$. Conventional MIMO precoding techniques, such as maximum ratio transmission (MRT), zero-forcing, and minimum mean squared error (MMSE) precoders, can be employed for the design of $\mathbf{W}_{n,i}$. In this work, we assume the MMSE precoder is employed which can effectively mitigate the multi-user interference and guarantee a high achievable sum-rate. The precoded signals in (4) are then transformed into the time domain signals via inverse discrete Fourier transform (IDFT) and a cyclic prefix of period T_{p} is inserted to mitigate inter-symbol interference. The time domain signals are assigned to the corresponding APs. After digital-to-analog conversion and radio frequency (RF) conversion, the RF signals with carrier frequency f_{c} are emitted through the transmit AP antennas.

The transmitted signals will be reflected by the targets in the area of interest and the reflected echo signals are received at the receive APs. We assume a line-of-sight (LoS) path exists between each AP and each target[1] [10]–[12]. After sampling and discrete Fourier transform (DFT) processing, the received echo signal for the t-th OFDM symbol on the i-th subcarrier at the m-th receive AP is given by (5) shown at the bottom of this page, where $t=1,\ldots,T_{\mathrm{s}}$, $i=0,\ldots,N_{\mathrm{s}}-1$, and $m=1,\ldots,M$. In (5), $\beta_{n,m,q}$ is a Gaussian-distributed complex coefficient, with zero mean and variance of χ^2. It includes the effects due to small-scale pathloss between the n-th transmit AP and m-th receive AP, and radar cross section of the q-th target. $\mathrm{PL}(d_{n,m,q})=\alpha_0(d_{n,m,q}/d_0)^{-\zeta}$ is the large-scale LoS pathloss coefficient between the n-th transmit AP and m-th receive AP through the q-th target, where α_0 is the pathloss at the reference distance d_0 and ζ is the pathloss exponent. $d_{n,m,q}$ is the bi-static range measured from the transmit AP n, via the q-th target, to the receive AP m. $\phi_{n,q}$ and $\theta_{n,q}$ correspond to the angles of departure (AoDs) from the transmit AP n to the q-th target. $\varphi_{m,q}$ and $\vartheta_{m,q}$ denote the angles of arrival (AoAs) of the q-th target observed from the m-th receive AP, respectively. $\mathbf{z}_{i,m}[t]\in\mathbb{C}^{M_{\mathrm{r}}}$ is the observed noise at the receive AP m on the i-th subcarrier, which follows the complex Gaussian distribution with zero mean and variance of $\xi_{\mathrm{z}}^2\mathbf{I}_{M_{\mathrm{r}}}$. $\tau_{n,m,q}=d_{n,m,q}/c$ and $f_{\mathrm{D},n,m,q}=f_{\mathrm{c}}/c(v_{n,q}+v_{m,q})$ are the bi-static delay and Doppler frequency shift associated with the n-th transmit AP and m-th receive AP through the q-th target, respectively. $v_{n,q}$ and $v_{m,q}$ are the relative velocities of the q-th target with respect to the n-th transmit AP and the m-th receive AP, respectively.

III. DNN for Cooperative ISAC-Assisted Localization and Velocity Estimation

We note that the received echo signal (5) contains information on angles, ranges (related to delays), and relative

[1]There may also exist multipath components. We assume the contribution of the multipath components is small and can be ignored for simplicity.

$$\mathbf{y}_{i,m}[t]=\sum_{n=1}^{N}\sum_{q=1}^{Q}\beta_{n,m,q}\sqrt{\mathrm{PL}(d_{n,m,q})}e^{-j2\pi(i\tau_{n,m,q}\Delta f-tf_{\mathrm{D},n,m,q}\Delta T)}\mathbf{a}(M_{\mathrm{r}},\varphi_{m,q},\vartheta_{m,q})\mathbf{a}^{\mathrm{H}}(N_{\mathrm{t}},\phi_{n,q},\theta_{n,q})\mathbf{x}_{n,i}[t]+\mathbf{z}_{i,m}[t]. \quad (5)$$

velocities (associated with Doppler frequency shifts). Considering a three-dimensional (3D) $(\mathrm{x}, \mathrm{y}, \mathrm{z})$ coordinate system, the location and velocity of the q-th target, $q = 1, \ldots, Q$, can be denoted as $\mathbf{l}_q = (l_q^{\mathrm{x}}, l_q^{\mathrm{y}}, l_q^{\mathrm{z}})$ and $\mathbf{v}_q = (v_q^{\mathrm{x}}, v_q^{\mathrm{y}}, v_q^{\mathrm{z}})$, respectively. Our goal is to estimate the location and velocity of the target by leveraging the echo signals received from multiple APs. Unlike conventional approaches [10]–[12] which extract the sensing parameters (e.g., angle, range, relative velocity) from each receive AP before estimating the location $\mathbf{l}_q$ and velocity $\mathbf{v}_q$ of the q-th target, we propose a DNN-based scheme to enable joint processing of the echo signals received across multiple APs and direct estimation of the location and velocity of the targets. The proposed DNN-based scheme avoids the estimation errors associated with sensing parameters.

In particular, each receive AP transmits the reflected echo signals to the CPU via the fronthaul link. The CPU first pre-processes the transmitted OFDM signals in (4) and the received echo signals in (5) as follows. Given the reflected echo signal in the t-th OFDM symbol duration on the i-th subcarrier at the m-th receive AP, the concatenated echo signals across all the receive APs can be expressed by $\mathbf{y}_i[t] = [(\mathbf{y}_{i,1}[t])^{\mathrm{T}} \ \cdots \ (\mathbf{y}_{i,M}[t])^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{MM_{\mathrm{r}}}$. Then, the aggregated echo signals on all the N_{s} subcarriers can be written as $\mathbf{Y}[t] = [\mathbf{y}_0[t] \ \cdots \ \mathbf{y}_{N_{\mathrm{s}}-1}[t]] \in \mathbb{C}^{MM_{\mathrm{r}} \times N_{\mathrm{s}}}$. We further aggregate the echo signals during all the T_{s} OFDM symbols duration and denote it as a 3D tensor $\mathbf{Y} = [\mathbf{Y}[1] \ \cdots \ \mathbf{Y}[T_{\mathrm{s}}]] \in \mathbb{C}^{MM_{\mathrm{r}} \times N_{\mathrm{s}} \times T_{\mathrm{s}}}$. We extract the real and imaginary parts of $\mathbf{Y}$, which are given by $\mathrm{Re}\{\mathbf{Y}\}$ and $\mathrm{Im}\{\mathbf{Y}\}$, respectively. Similarly, we construct a 3D tensor $\mathbf{X} \in \mathbb{C}^{NN_{\mathrm{t}} \times N_{\mathrm{s}} \times T_{\mathrm{s}}}$, which aggregates the transmitted OFDM signals from (4) across all the N transmit APs on all the N_{s} subcarriers during all the T_{s} OFDM symbols duration. The real and imaginary parts of $\mathbf{X}$ are denoted as $\mathrm{Re}\{\mathbf{X}\}$ and $\mathrm{Im}\{\mathbf{X}\}$, respectively. Then, $\mathrm{Re}\{\mathbf{X}\}$, $\mathrm{Im}\{\mathbf{X}\}$, $\mathrm{Re}\{\mathbf{Y}\}$, and $\mathrm{Im}\{\mathbf{Y}\}$ are normalized and are used as the input to the developed DNN for target sensing.

The developed DNN extracts useful features from the pre-processed signals and estimates the locations and velocities of the targets within the area of interest. We note that the transmitted signals and reflected echo signals contain independent information across the spatial, frequency, and time domains, which is critical for effective feature extraction. We employ 3D convolutional neural networks (CNNs) to extract the spatial-frequency-time domain features from the transmitted signals $\mathbf{X}$ and the reflected echo signals $\mathbf{Y}$. The real and imaginary parts of the transmitted signal, i.e., $\mathrm{Re}\{\mathbf{X}\}$ and $\mathrm{Im}\{\mathbf{X}\}$, are regarded as two input channels for 3D convolution. We employ a 3D convolutional filter with parameters $\mathbf{\Phi}_{\mathrm{x}}^{\mathrm{CNN}}$ to process the input tensors $\mathrm{Re}\{\mathbf{X}\}$ and $\mathrm{Im}\{\mathbf{X}\}$. The number of output channels is denoted by H_{x}. After 3D convolution, the processed transmitted signal is given by $\bar{\mathbf{X}} \in \mathbb{R}^{H_{\mathrm{x}} \times D_{\mathrm{x}}^1 \times D_{\mathrm{x}}^2 \times D_{\mathrm{x}}^3}$, where D_{x}^1, D_{x}^2, and D_{x}^3 are the dimensions after 3D convolution operation. Similarly, we employ another 3D convolutional filter with parameters $\mathbf{\Phi}_{\mathrm{y}}^{\mathrm{CNN}}$ to extract the features from the real and imaginary parts of the reflected echo signal, i.e., $\mathrm{Re}\{\mathbf{Y}\}$ and $\mathrm{Im}\{\mathbf{Y}\}$. Then, we obtain the processed echo signal $\bar{\mathbf{Y}} \in \mathbb{R}^{H_{\mathrm{y}} \times D_{\mathrm{y}}^1 \times D_{\mathrm{y}}^2 \times D_{\mathrm{y}}^3}$, where H_{y} is the number of output channels of the convolutional filter. D_{y}^1, D_{y}^2, and D_{y}^3 are the dimensions after the 3D convolution operation. We use the rectified linear unit (ReLU) as the activation function for both CNNs. After pooling and flattening, we obtain $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, which contain the encoded transmit and sensing information, respectively. We then apply linear projectors with weight matrices $\mathbf{W}_{\bar{\mathrm{x}}}$ and $\mathbf{W}_{\bar{\mathrm{y}}}$ on the flattened vectors $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, respectively, to extract the combined and high-level features. The outputs are concatenated and fed into a fully connected layer with weight matrix $\mathbf{W}_{\mathrm{FC}}$. Finally, we employ another fully connected layer with weight matrix $\mathbf{W}_{\mathrm{OUT}}$ to generate the estimated locations $\hat{\mathbf{l}} \in \mathbb{R}^{3Q}$ and velocities $\hat{\mathbf{v}} \in \mathbb{C}^{3Q}$ for all the targets. For the q-th target, the estimated location and velocity are given by $\hat{\mathbf{l}}_q = \hat{\mathbf{l}}[3(q-1)+1 : 3q]$ and $\hat{\mathbf{v}}_q = \hat{\mathbf{v}}[3(q-1)+1 : 3q]$, respectively. We stack the CNN and fully connected layers together. The overall architecture of the proposed DNN is shown in Fig. 1. The set of network parameters is denoted as $\mathbf{\Phi} = \{\mathbf{\Phi}_{\mathrm{x}}^{\mathrm{CNN}}, \mathbf{\Phi}_{\mathrm{y}}^{\mathrm{CNN}}, \mathbf{W}_{\bar{\mathrm{x}}}, \mathbf{W}_{\bar{\mathrm{y}}}, \mathbf{W}_{\mathrm{FC}}, \mathbf{W}_{\mathrm{OUT}}\}$.

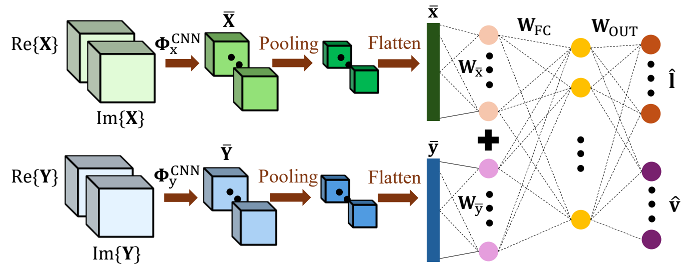
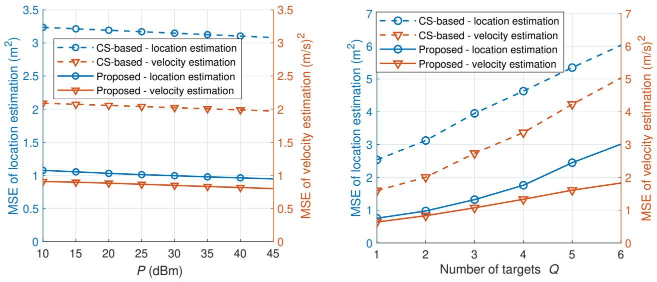


Fig. 1. The proposed DNN architecture for cooperative ISAC-assisted localization and velocity estimation.

During offline training, we construct a training dataset, which contains the transmitted OFDM signals $\mathbf{X}$ and the reflected echo signals $\mathbf{Y}$ as input, and the true location $\mathbf{l_q}$ and velocity $\mathbf{v}_q$ of the q-th target as labels, where $q = 1, \ldots, Q$. The labels are normalized to be between zero and one by using max-min normalization during the training stage. The normalized values are denoted as $\bar{\mathbf{l}}_q$ and $\bar{\mathbf{v}}_q$ for the q-th target. We use the mean square error (MSE) loss function for DNN training, which is defined as follows:

$$\mathcal{L}(\mathbf{\Phi}) = \sum_{q=1}^{Q} \left(\|\bar{\mathbf{l}}_q - \hat{\mathbf{l}}_q\|^2 + \|\bar{\mathbf{v}}_q - \hat{\mathbf{v}}_q\|^2 \right). \tag{6}$$

The proposed DNN is trained in a supervised manner to minimize the MSE loss between the normalized ground truth and estimated results as in (6) using the Adam optimizer [14]. After training, we can obtain the trained DNN with its optimized parameters $\mathbf{\Phi}^{\star}$. During online execution, given the transmitted OFDM signal in (4) and the reflected echo signal in (5), we first pre-process the signals. The pre-processed signals are then fed into the trained DNN model, which outputs the normalized target locations and velocities within the area of interest. Finally, these normalized results are rescaled to their nominal values to provide estimations.



(a) (b)

Fig. 2. MSE of location and velocity estimation versus (a) the transmit power P and (b) number of targets Q.

IV. Performance Evaluation

In this section, we evaluate the sensing sensing of the proposed DNN-based scheme through simulations. We consider a cell-free MIMO system with $N = 2$ transmit APs and $M = 2$ receive APs within a coverage area of 100×100 m^2. The transmit APs are located at $(0, 50, 20)$ and $(100, 50, 20)$. The receive APs are placed at $(50, 0, 20)$ and $(50, 100, 20)$ in 3D coordinates. Each AP has 8 horizontal antennas and 2 vertical antennas. The carrier frequency is 30 GHz and the subcarrier spacing is 240 kHz. We consider the setting of $N_{\rm s} = 512$ subcarriers and $T_{\rm s} = 256$ OFDM symbols during downlink transmission. There are $K = 4$ users and $Q = 2$ targets. The users and targets are randomly located in the 3D environment, with x and y coordinates ranging from 0 to 100, and the z coordinate ranging from 0 to 30. The velocity of each target is between -30 m/s and 30 m/s. For the sensing channel, we set $\alpha_0 = 0$ dBm, $d_0 = 1$ m, $\zeta = 2.6$, and $\chi^2 = 0.1$. The AoDs and AoAs of the targets are determined by their corresponding locations. We generate $10,000$ data samples, where $8,000$ of them are used for offline training and the remaining $2,000$ data samples are used for online testing.

In Fig. 2, we illustrate the MSE of location and velocity estimation under different system settings. The results are obtained during online testing, where the estimated results are rescaled back to their normal values. We compare our proposed scheme with a state-of-the-art approach from [12], which uses compressive sensing (CS). The CS-based scheme estimates the sensing parameters from each AP independently followed by target sensing. Fig. 2(a) shows the MSE for location and velocity estimation versus the transmit power P at each AP. It can be observed from Fig. 2(a) that our proposed scheme results in a significantly lower MSE for both localization and velocity estimation compared with the CS-based approach. This is mainly due to the fact that our proposed DNN-based scheme can benefit from jointly extracting features from the echo signals collected by the distributed APs. Moreover, the proposed scheme performs the localization and velocity estimation directly from the echo signals. Thus, it can avoid any errors associated with the intermediate parameter estimation. Fig. 2(b) illustrates the MSE of location and velocity estimation versus the number of targets Q, where the transmit power at each transmit AP is set to $P = 30$ dBm. The results indicate that the MSE increases as the number of targets increases. This is because a higher number of targets makes it more challenging to distinguish them in the spatial-frequency-time domain, leading to higher estimation errors. In addition, our proposed DNN-based scheme consistently achieves better sensing performance than the CS-based scheme. In Fig. 3, we evaluate the convergence performance of the proposed DNN-based scheme, where the training loss function in (6) versus the training epoches is illustrated. It can be observed from the figure that the proposed DNN can converge quickly within 300 training epochs. Finally, in Fig. 4, we visualize the results of target localization and velocity estimation and compare the estimated values with the ground truth. The results in Fig. 4 demonstrate that the proposed scheme achieves a high localization accuracy, with the estimation errors to be within 1 m in the considered cell-free MIMO system. Similarly, the estimated velocity of the target is also shown to be close to the ground truth, as can be observed from the figure.

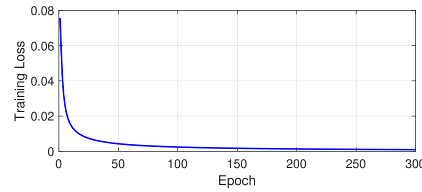


Fig. 3. Training loss versus the training epoches.

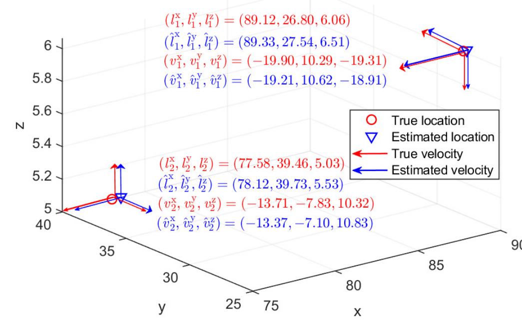


Fig. 4. Comparison between the ground truth and the estimated location and velocity.

V. Conclusion

In this paper, we proposed a DNN-based scheme for cooperative ISAC-assisted target localization and velocity estimation. In our proposed scheme, the received echo signals are jointly processed across geographically distributed APs. Different from the conventional approaches which estimate sensing parameters prior to target sensing, our proposed scheme directly performs the sensing tasks using the reflected echo signals. The proposed scheme bypasses the intermediate sensing parameter estimation stage and improves the sensing performance. Simulation results demonstrated that our proposed scheme can provide more accurate localization and velocity estimation when compared with a CS-based approach.

References

[1] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

[2] N. González-Prelcic, M. F. Keskin, O. Kaltiokallio, M. Valkama, D. Dardari, X. Shen, Y. Shen, M. Bayraktar, and H. Wymeersch, "The integrated sensing and communication revolution for 6G: Vision, techniques, and applications," *Proc. of the IEEE*, vol. 112, no. 7, pp. 676–723, Jul. 2024.

[3] A. Kaushik, R. Singh, M. Li, H. Luo, S. Dayarathna, R. Senanayake, X. An, R. A. Stirling-Gallacher, W. Shin, and M. D. Renzo, "Integrated sensing and communications for IoT: Synergies with key 6G technology enablers," *IEEE Internet Things Mag.*, vol. 7, no. 5, pp. 136–143, Sept. 2024.

[4] Z. Wang, V. W.S. Wong, and R. Schober, "Integrated sensing and communications for end-to-end predictive beamforming design in vehicle-to-infrastructure networks," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 5, pp. 933–949, Jul. 2024.

[5] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath Jr., Z. Feng, L. Zheng, and Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *IEEE J. Sel. Topics in Signal Process.*, vol. 15, no. 6, pp. 1295–1315, Nov. 2021.

[6] S. H. Dokhanchi, M. R. B. Shankar, T. Stifter, and B. Ottersten, "OFDM-based automotive joint radar-communication system," in *Proc. of IEEE Radar Conf.*, Oklahoma City, OK, Apr. 2018.

[7] J. B. Sanson, P. M. Tomé, D. Castanheira, A. Gameiro, and P. P. Monteiro, "High-resolution delay-Doppler estimation using received communication signals for OFDM radar-communication system," *IEEE Trans. Veh. Techn.*, vol. 69, no. 11, pp. 13 112–13 123, Nov. 2020.

[8] Y. Liu, G. Liao, Y. Chen, J. Xu, and Y. Yin, "Super-resolution range and velocity estimations with OFDM integrated radar and communications waveform," *IEEE Trans. Veh. Techn.*, vol. 69, no. 10, pp. 11 659–11 672, Oct. 2020.

[9] Y. Wu, F. Lemic, C. Han, and Z. Chen, "Sensing integrated DFT-spread OFDM waveform and deep learning-powered receiver design for terahertz integrated sensing and communication systems," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 595–610, Jan. 2023.

[10] Z. Xu and A. Petropulu, "A bandwidth efficient dual-function radar communication system based on a MIMO radar using OFDM waveforms," *IEEE Trans. Signal Process.*, vol. 71, pp. 401–416, Feb. 2023.

[11] Z. Xiao, R. Liu, M. Li, and Q. Liu, "A novel joint angle-range-velocity estimation method for MIMO-OFDM ISAC systems," *IEEE Trans. Signal Process.*, vol. 72, pp. 3805–3818, Aug. 2024.

[12] M. L. Rahman, J. A. Zhang, X. Huang, Y. J. Guo, and R. W. Heath, "Framework for a perceptive mobile network using joint communication and radar sensing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 6, pp. 1926–1941, Jun. 2020.

[13] Z. Zhang, H. Ren, C. Pan, S. Hong, D. Wang, J. Wang, and X. You, "Target localization in cooperative ISAC systems: A scheme based on 5G NR OFDM signals," *IEEE Trans. Commun.*, 2024 (early access).

[14] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, San Diego, CA, May 2015.