# Deep Learning Based Multi-Node ISAC 4D Environmental Reconstruction with Uplink-Downlink Cooperation

Bohao Lu, Zhiqing Wei, Member, IEEE, Huici Wu, Member, IEEE, Xinrui Zeng, Lin Wang, Student Member, IEEE, Xi Lu, Dongyang Mei, and Zhiyong Feng, Senior Member, IEEE

Abstract—Utilizing widely distributed communication nodes to achieve environmental reconstruction is one of the significant scenarios for Integrated Sensing and Communication (ISAC) and a crucial technology for 6G. To achieve this crucial functionality, we propose a deep learning based multi-node ISAC 4D environment reconstruction method with Uplink-Downlink (UL-DL) cooperation, which employs virtual aperture technology, Constant False Alarm Rate (CFAR) detection, and Mutiple Signal Classification (MUSIC) algorithm to maximize the sensing capabilities of single sensing nodes. Simultaneously, it introduces a cooperative environmental reconstruction scheme involving multi-node cooperation and Uplink-Downlink (UL-DL) cooperation to overcome the limitations of single-node sensing caused by occlusion and limited viewpoints. Furthermore, the deep learning models Attention Gate Gridding Residual Neural Network (AGGRNN) and Multi-View Sensing Fusion Network (MVSFNet) to enhance the density of sparsely reconstructed point clouds are proposed, aiming to restore as many original environmental details as possible while preserving the spatial structure of the point cloud. Additionally, we propose a multi-level fusion strategy incorporating both datalevel and feature-level fusion to fully leverage the advantages of multi-node cooperation. Experimental results demonstrate that the environmental reconstruction performance of this method significantly outperforms other comparative method, enabling high-precision environmental reconstruction using ISAC system.

Index Terms—Integrated Sensing and Communication, Environmental Reconstruction, Virtual Aperture, Constant False-Alarm Rate, Mutiple Signal Classification, Multi-Node Cooperation, Uplink-Downlink Cooperation, Multi-Level Fusion, Deep Learning.

## I. INTRODUCTION

#### A. Background and Motivation

Integrated Sensing and Communication (ISAC) has been defined as a new feature of next-generation cellular networks by ITU-R as one of the six visions of IMT2030 (6G) [1]. The rapid development of 5th Generation (5G), Beyond 5G (B5G) and 6th Generation (6G) wireless communication has

This work was supported in part by the National Key Research and Development Program under Grant 2020YFA0711302, in part by the National Natural Science Foundation of China (NSFC) under Grant 62271081, 92267202, and U21B2014.

Bohao Lu, Zhiqing Wei, Huici Wu, Lin Wang, Xi Lu, Dongyang Mei and Zhiyong Feng are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (email: {bohaolu; weizhiqing; dailywu; wlwl; luxi; meidongyang; fengzy}@bupt.edu.cn). Xinrui Zeng is with Beijing University of Aeronautics and Astronautics, Beijing 100191, China (email: {xr\_zeng}@buaa.edu.cn)

Corresponding authors: Zhiqing Wei

conferred communication signals a larger bandwidth and higher frequency bands, which have significantly improved their sensing capabilities. In ISAC scenarios, cellular network devices not only serve as communication nodes to realize the interconnection of everything, but also can be used as sensing nodes to realize functions such as environmental imaging and reconstruction. The multi-node cooperation between the widely distributed communication infrastructure and User Equipment (UE) makes multi-view and omnidirectional environmental imaging and reconstruction possible. Meanwhile, the use of deep neural networks and multilevel fusion strategies can further improve the precision of environmental imaging and reconstruction and maximize the details of the sensing results.

## B. Related Works

Most of the current research on 3D and 4D environmental sensing based on wireless signals utilizes Frequency Modulated Continuous Wave (FMCW) waveform which primarily uses Fast Fourier Transform (FFT)-based algorithms for echo signal processing [2]. Santra et al. [3] and Sun et al. [4] proposed the high-resolution Multiple Input Multiple Output (MIMO) radar imaging system based on FMCW waveform. The above works use FFT-based algorithms to process the echoes, which are lacking in terms of environmental sensing resolution compared to MUSIC-based methods. And the FMCW waveform-based environmental sensing scheme is not suitable for ISAC scenarios [5], [6]. In recent years the research about single target [7]-[10] and multi-target [11] sensing based on Orthogonal Frequency Division Multiplexing (OFDM) signals in ISAC scenarios have achieved good performance. However, there is still a lack of research on OFDM radar-based 4D environmental reconstruction scenarios with massive targets. Guan et al. [12] built a single-node high-resolution 3D radar imaging system using 5G millimeter wave (mmWave) signals. However, due to the existence of occlusions, the system is unable to achieve a complete environmental reconstruction of the sensing area.

Utilizing multi-node cooperation to build a multi-view environmental reconstruction system can overcome the limitations of single-node sensing [13], [14]. Ren *et al.* [15] determined the weights of each sensing node by considering factors such as consistency, stability and correlation between parameters. Nuss *et al.* [16] performed multi-target sensing based on MIMO-OFDM radar through measurement matrix superposition algorithm. Gao *et al.* [17] proposed a gridbased probabilistic fusion framework to mitigate the effects of radar ghosting. The partly multi-radar sensing fusion based approach defines a maximum likelihood function for each position grid, which is estimated as the position corresponding to the maximum value of the function [18], [19]. Most of the above related works are cooperative schemes based on multi-base radar, which cannot utilize the inherent advantages of integrating communication node UL with DL and are not applicable to ISAC scenarios as compared to the cooperative schemes based on multiple communication nodes.

Since the sensing capability in ISAC scenario is limited by the frequency and bandwidth resources used by the communication nodes, the beam-pointing situation, and the electromagnetic environment, the point cloud density of the environmental reconstruction results from multi-node cooperation is similarly limited [20], [21]. In order to make the multinode cooperative environmental reconstruction result with higher point cloud density to recover the detail information of the original scenario more completely, it is necessary to introduce a point cloud complementation algorithm for lowdensity reconstruction point cloud density enhancement [22], [23]. More recently, several attempts [24], [25] have been made to incorporate Graph Convolutional Networks (GCN) [26] to build local graphs in the neighborhood of each point in the point cloud. However, constructing the graph relies on the K-Nearest Neighbor (KNN) algorithm, which is sensitive to the point cloud density [27]. In addition to the above, there are almost no open source datasets related to ISACbased 4D environmental reconstruction, resulting in a lack of work related to low-density point cloud enhancement networks based on ISAC environmental reconstruction scenarios.

# C. Contributions

To solve the above problems, we propose a deep learning based Multi-Node with Downlink-Uplink Cooperative (MN-DUC) ISAC 4D environmental reconstruction method. The main contributions and innovations of this paper are summarized as follows.

- A Base Station (BS) side ISAC 4D (velocity, 3D position information) environmental reconstruction algorithm based on a virtual receiver array with CFAR detection is proposed. The algorithm uses the standard 5G NR mmWave signal as the signal carrier, processes the echo signal using 2D-FFT with with 2D-MUSIC algorithm, and performs environment scatterers detection based on CFAR.
- A multi-node UL-DL cooperative 4D environmental reconstruction method is proposed. This method realizes active sensing at the BS side during the DL time slot and passive sensing at the BS side in cooperation with the UE during the UL time slot. And it achieves data-level ISAC 4D environmental reconstruction result fusion.
- A density-enhanced network Attention Gate Gridding Residual Neural Network (AGGRNN) for low-density reconstruction of point clouds for data-level fusion result is

proposed. We also apply distributed learning with Multimodal inputs for feature-level fusion on the foundation of AGGRNN, and build a Multi-View Sensing Fusion Network (MVSFNet) to further optimize the ISAC 4D environmental reconstruction result.

We use Chamfer Distance (CD) and F-Score@1% to measure the performance of the reconstruction results. The simulation results show that, our proposed ISAC 4D environmental reconstruction method has better performance compared to the comparison methods.

The rest of the paper is organized as follows. Section II presents our system and signal model. The processing scheme of DL and UL ISAC sensing signal is introduced in Section III. In Section IV, the schemes of sensing result fusion and enhancement are proposed. Section V discusses our simulation results. Finally, we conclude the work of this paper in Section VI.

#### II. SYSTEM AND SIGNAL MODEL

This section presents the scenario setup, ISAC channel models, and transmit-received signal models for a Multi-Node Downlink-Uplink Cooperative (MNDUC) ISAC imaging method.

A. Scenario Setup of MNDUC ISAC 4D Environmental Reconstruction Method



Fig. 1. MNDUC ISAC 4D environmental reconstruction scenario.

As shown in Fig. 1, we consider a MNDUC ISAC 4D environmental reconstruction scenario between BS, UE, and the environment. The BS is equipped with two spatially well-separated Uniform Planar Arrays (UPAs). The transmit and receive arrays of the BS have dimensions  $P_t \times Q_t$  and  $P_r \times Q_r$ , whose array element spacing and the arrangement will be described in detail below. The UE is equipped with antenna array size  $P_{rt} \times Q_{rt}$ .

The modes of sensing in the DL and UL periods are active and passive sensing, respectively. To obtain a complete

environmental reconstruction result, we consider a multi-level fusion strategy to fuse and match the sensing results of multiple sensing nodes in DL and UL periods.

#### B. UPAs Model and Virtual Aperture



Fig. 2. MIMO array design with the transmitting array, the receiving array and the corresponding virtual Rx (vRx) array.

We have designed a transmit-receive UPAs as shown in Fig. 2 to improve the sensing resolution. The receiving UPA is a  $2 \times 2$  array with an element spacing of 8d, the transmitting UPA is an  $8 \times 8$  array with a spacing of d, and the corresponding receive UPA is a  $16 \times 16$  array with a spacing of d.



Fig. 3. Schematic diagram of geometric relationships.

The geometric relationship between the kth scattering point and BS is shown in Fig. 3, with the azimuth and pitch angles noted as  $(\theta_k, \varphi_k)$ . The antenna array element in the virtual receiving array is denoted as  $A_{r_v}(p_{r_v}, q_{r_v})$ , and the reference antenna array element is  $A_{r_v}(1,1)$ , then the phase difference between  $A_{r_v}(p_{r_v}, q_{r_v})$  and  $A_{r_v}(1,1)$  caused by the kth scattering point can be expressed as [28]

$$\begin{aligned} \Delta \phi_{r_{v}} (p_{r_{v}}, q_{r_{v}})_{(\theta_{k}, \varphi_{k})} \\ &= e^{-\xi j 2\pi \frac{d[(p_{r_{v}}-1)\cos\theta_{k}+\psi(q_{r_{v}}-1)\sin\theta_{k}]\cos\varphi_{k}}{\lambda}} \\ \begin{cases} \xi = 1 \ \psi = 1, \theta_{k} < 90^{\circ}, \varphi_{k} < 90^{\circ} \\ \xi = -1 \ \psi = -1, \theta_{k} > 90^{\circ}, \varphi_{k} < 90^{\circ} \\ \xi = 1 \ \psi = -1, \theta_{k} < 90^{\circ}, \varphi_{k} > 90^{\circ} \\ \xi = -1 \ \psi = 1, \theta_{k} < 90^{\circ}, \varphi_{k} > 90^{\circ} \end{aligned}$$
(1)

As shown in (2), the vRx array phase difference matrix  $\mathbf{A}_{k}^{q}$  in DL sensing period corresponding to the *k*th target is obtained from (1).

$$\mathbf{A}_{k}^{D}\left(p_{r_{v}}, q_{r_{v}}\right) = \Delta \phi_{r_{v}}\left(p_{r_{v}}, q_{r_{v}}\right)_{\left(\theta_{k}, \varphi_{k}\right)}$$
(2)

As shown in (3), the Rx array phase difference matrix  $\mathbf{A}_{k}^{q}$  in UL sensing period corresponding to the *k*th target is obtained from (1).

$$\mathbf{A}_{k}^{U}\left(p_{r},q_{r}\right) = \Delta\phi_{r}\left(p_{r},q_{r}\right)_{\left(\theta_{k},\varphi_{k}\right)} \tag{3}$$

where q = D or U are for DL or UL ISAC signals, respectively.

# C. Transmitted ISAC Signals

To adapt to current mobile wireless communication network systems, OFDM signals are used for both DL and UL signals. The continuous time domain OFDM ISAC signal is defined as [29]

$$y^{q}(t) = \sum_{m=0}^{N_{sys}^{q}-1} \sum_{n=0}^{N_{c}^{q}-1} s_{\text{Tx}}^{q}(n,m) e^{j2\pi(f_{c}+n\Delta f^{q})t} \operatorname{rect}\left(\frac{t-mT_{\text{OFDM}}^{q}}{T_{\text{OFDM}}^{q}}\right)$$
(4)

where  $s_{\text{Tx}}^q(n,m)$  represents the modulated OFDM symbol in the *n*th subcarrier of *m*th OFDM symbol,  $N_{sys}^q - 1$  is the number of OFDM symbols and  $N_c^q - 1$  is the number of subcarriers,  $T_{\text{OFDM}}^q$  is the total duration of the OFDM symbol which satisfies  $T_{\text{OFDM}}^q = T^q + T_{\text{CP}}^q$ , where  $T^q$  is the effective OFDM symbol duration and  $T_{\text{CP}}^q$  is the cyclic prefix duration,  $f_c$  is the carrier frequency,  $\Delta f^q = 1/T^q$  is the frequency interval of subcarriers, rect  $\left(\frac{t}{T_{\text{OFDM}}^q}\right)$  is the rectangular function which is equal to 1 for  $0 \le t \le T_{\text{OFDM}}^q$  and 0 for otherwise.

# D. DL and UL ISAC Channel Models



Fig. 4. Channel models of DL and UL Channel.

We denote the UL communication channel as the UL ISAC channel; the DL communication channel is a transposition of the ISAC UL channel due to channel reciprocity; the DL sensing channel consists of the echo path from the scatterers as shown in Fig. 4. We only consider the echo reflected directly from the scatterers.

1) UL ISAC Channel Model: The UL ISAC channel response matrix on the *n*th subcarrier of the *m*th OFDM symbol is defined as [30]

$$\mathbf{H}_{C,n,m}^{U} = \sum_{l=0}^{L-1} \begin{bmatrix} b_{C,l} e^{j2\pi(f_{c,d,l})mT_{\text{OFDM}}^{U}e^{-j2\pi n\Delta f^{U}\left(\tau_{c,l}\right)}} \\ \times \mathbf{a}\left(\mathbf{p}_{Rx,l}^{U}\right) \mathbf{a}^{T}\left(\mathbf{p}_{Tx,l}^{U}\right) \end{bmatrix}$$
(5)

where  $\mathbf{H}_{C,n,m}^U \in \mathbb{C}^{P_tQ_t \times P_{rt}Q_{rt}}$ , l = 0 is for the channel response of the LoS (Line of Sight) path, and  $l \in \{1, ..., L-1\}$  is for the paths involved with the *l*th scatterer;  $\mathbf{a}\left(\mathbf{p}_{Rx,l}^U\right) \in \mathbb{C}^{P_tQ_t \times 1}$  and  $\mathbf{a}\left(\mathbf{p}_{Tx,l}^U\right) \in \mathbb{C}^{P_{rt}Q_{rt} \times 1}$  are the steering vectors for UL reception and transmission,  $\mathbf{p}_{Rx,l}^U$  and  $\mathbf{p}_{Tx,l}^U$  are the corresponding Angle of Arrival (AoA) and Angle of Departure (AoD), respectively;  $f_{c,d,0} = \frac{v_0}{\lambda}$  and  $\tau_{c,0} = \frac{r_{0,1}}{c}$  are the Doppler shift and time delay with  $v_0$  and  $r_{0,1}$  corresponding to the velocity and range between the UE and BS of the LoS path, respectively;  $f_{c,d,l} = f_{d,l,1} + f_{d,l,2}$ and  $\tau_{c,l} = \tau_{c,l,1} + \tau_{c,l,2}$  are the total Doppler shift and time delay of the *l*th NLoS (Non Line of Sight) path, respectively;  $f_{d,l,1} = \frac{v_{r,l,1}}{\lambda}$  and  $f_{d,l,2} = \frac{v_{r,l,2}}{\lambda}$  are the Doppler shifts between the UE and the *l*th scatterer, and between the *l*th scatterer and the BS, respectively, with  $v_{r,l,1}$  and  $v_{r,l,2}$  being the velocities;  $\tau_{c,l,1} = \frac{r_{l,1}}{c}$  and  $\tau_{c,l,2} = \frac{r_{l,2}}{c}$  are the time delays between the user and the *l*th scatterer, and between BS and the *l*th scatterer, respectively, with  $r_{l,1}$  and  $r_{l,2}$  being the ranges. The attenuation  $b_{C,l} = \sqrt{\frac{\lambda^2}{(4\pi r_{0,1})^2}}, l = 0$  for the LoS path and  $b_{C,l} = \beta_{C,l} \times \sqrt{\frac{\lambda^2}{(4\pi)^3 r_{l,1}^2 r_{l,2}^2}}, l \in \{1, ..., L-1\}$  for the NLoS paths, where  $\beta_{C,l}$  is the reflecting factor of the *l*th scatterer, following  $\mathcal{CN}\left(0, \sigma_{C\beta,l}^2\right)$ .

2) *DL Communication Channel Model:* From the channel reciprocity, the transpose of the UL communication channel response is the DL communication channel response.

$$\mathbf{H}_{C,n,m}^{D} = \sum_{l=0}^{L-1} \begin{bmatrix} b_{C,l} e^{j2\pi(f_{c,d,l})mT_{\text{OFDM}}^{D}e^{-j2\pi n\Delta f^{D}\left(\tau_{c,l}\right)}} \\ \times \mathbf{a}\left(\mathbf{p}_{Rx,l}^{D}\right) \mathbf{a}^{T}\left(\mathbf{p}_{Tx,l}^{D}\right) \end{bmatrix}$$
(6)

where  $\mathbf{H}_{C,n,m}^{D} \in \mathbb{C}^{P_{t}Q_{t} \times P_{rt}Q_{rt}}$ ,  $\mathbf{a}\left(\mathbf{p}_{Rx,l}^{D}\right) \in \mathbb{C}^{P_{t}Q_{t} \times 1}$  and  $\mathbf{a}\left(\mathbf{p}_{Tx,l}^{D}\right) \in \mathbb{C}^{P_{rt}Q_{rt} \times 1}$  are the steering vectors for DL transmission,  $\mathbf{p}_{Rx,l}^{D} = \mathbf{p}_{Tx,l}^{U}$  and  $\mathbf{p}_{Tx,l}^{D} = \mathbf{p}_{Rx,l}^{U}$  are the DL communication AoA and AoD, respectively.

3) DL Sensing Channel Model: The DL sensing channel response matrix on the *n*th subcarrier of the *m*th OFDM symbol is defined as [30]

$$\mathbf{H}_{S,n,m}^{D} = \sum_{l=0}^{L-1} \begin{bmatrix} b_{S,l} e^{j2\pi(f_{s,l,1})mT_{\text{OFDM}}^{D}} e^{-j2\pi n\Delta f^{D}\left(\tau_{s,l}\right)} \\ \times \mathbf{a}\left(\mathbf{p}_{Rx,l}^{DS}\right) \mathbf{a}^{T}\left(\mathbf{p}_{Tx,l}^{D}\right) \end{bmatrix}$$
(7)

where  $\mathbf{a}\left(\mathbf{p}_{Rx,l}^{DS}\right) \in \mathbb{C}^{P_{r_v}Q_{r_v} \times 1}$  and  $\mathbf{a}^T\left(\mathbf{p}_{Tx,l}^D\right) \in \mathbb{C}^{P_{r_v}Q_{r_v} \times 1}$ are the steering vectors for DL echo reception and DL transmission, respectively;  $\mathbf{p}_{Rx,l}^{DS}$  and  $\mathbf{p}_{Tx,l}^D$  are the AoD and AoA of the ISAC sensing receiver and transmitter array, respectively. Since mmWave arrays are typically small and equivalent to virtual receiver arrays during the DL sensing period as shown in Fig. 2,  $\mathbf{p}_{Rx,l}^{DS} = \mathbf{p}_{Tx,l}^D$ . Where  $f_{s,0,1} = \frac{2v_0}{\lambda}$  and  $f_{s,l,1} = \frac{2v_{r,l,2}}{\lambda}$  are the total Doppler shift of the *l*th echo path with  $v_0$  and  $v_{r,l,2}$  being the velocities;  $\tau_{s,0} = \frac{2r_{0,1}}{c}$  and  $\tau_{s,l} = \frac{2r_{l,2}}{c}$  are the time delays of the *l*th echo path with  $r_{0,1}$  and  $r_{l,2}$  being the ranges.  $b_{S,l} = \beta_{S,l} \times \sqrt{\frac{\lambda^2}{(4\pi)^3 d_{l,2}^4}}, l \in \{0, ..., L-1\}$  is the attenuation for the echo path, where  $\beta_{S,l}$  is the reflecting factor of the *l*th scatterer, following  $\mathcal{CN}\left(0, \sigma_{S\beta,l}^2\right)$ .

## E. Received ISAC Signals

1) Received Communication Signals: The received communication signal on the *n*th subcarrier of the *m*th OFDM symbol is defined as [28], [30]

$$\begin{aligned} \mathbf{y}_{C,n,m}^{q} &= s_{Tx}^{q}\left(n,m\right) \mathbf{H}_{C,n,m}^{q} \mathbf{w}_{Tx}^{q} + \mathbf{n}_{t,n,m}^{q} \\ &= s_{Tx}^{q}\left(n,m\right) \sum_{l=0}^{L-1} \begin{bmatrix} b_{C,l} e^{j2\pi \left(f_{c,d,l}\right) m T_{\text{OFDM}}^{q}} e^{-j2\pi n \Delta f^{q}\left(\tau_{c,l}\right)} \\ &\times \mathbf{a} \left(\mathbf{p}_{Rx,l}^{q}\right) \chi_{Tx,l}^{q} \end{bmatrix} + \mathbf{n}_{t,n,m}^{q} \end{aligned} \end{aligned}$$
(8)

where q = D or U are for DL or UL ISAC signals, respectively;  $\mathbf{n}_{t,n,m}^q$  is the noise vector and each element of  $\mathbf{n}_{t,n,m}^q$  follow  $\mathcal{CN}(0, \sigma_N^2)$ ;  $\mathbf{y}_{C,n,m}^q, \mathbf{n}_{t,n,m}^q \in \mathbb{C}^{P_t Q_t \times 1}$ and  $\mathbf{y}_{C,n,m}^q, \mathbf{n}_{t,n,m}^q \in \mathbb{C}^{P_{r_t} Q_{r_t} \times 1}$  when q = U and q = D, respectively;  $s_{Tx}^q(n,m)$  represents the modulated OFDM symbol;  $\mathbf{w}_{Tx}^q$  is the transmit Beamforming (BF) vector, and  $\chi_{Tx,l}^q = \mathbf{a}^T \left( \mathbf{p}_{Tx,l}^q \right) \mathbf{w}_{Tx}^q$  is the transmit BF gain. 2) *Received Sensing Signals:* The received signal when the

2) Received Sensing Signals: The received signal when the OFDM signal shown in (4) reflected by the *k*th scatterer is defined as [28]

$$y_{k}^{q}(t) = G_{k} \sum_{m=0}^{N_{sym}^{q}-1} e^{j2\pi f_{d}(k)t} \sum_{n=0}^{N_{c}^{q}-1} s_{Rx}^{q}(n,m,k) \times e^{j2\pi (f_{c}+n\Delta f^{q})\left(t-\frac{2R_{k}}{c}\right)} \times \operatorname{rect}\left(\frac{t-mT_{\text{OFDM}}^{q}-\frac{2R_{k}}{c}}{T_{\text{OFDM}}^{q}}\right)$$
(9)

where  $R_k$  and  $f_d(k)$  is the range and Doppler shift of kth scatter, respectively.  $G_k$  represents the attenuation factor associated with the path loss, radar cross section (RCS) of the kth scatter. The relationship between the received modulation symbols  $s_{Rx}^q(n,m,k)$  and the transmitted modulation symbol  $s_{Tx}q(n,m)$  can be defined as

$$s_{Rx}^{q}(n,m,k) = G_{k}s_{Tx}^{q}(n,m) e^{-j2\pi f_{n}^{q}\frac{2R_{k}}{c}} e^{j2\pi f_{d}(k)mT_{\text{OFDM}}^{q}}$$
(10)

where  $f_n^q = f_c + n\Delta f^q$ .

The CSI(Channel State Information) matrix information matrix  $\mathbf{s}_{a}^{q}$  is defined as

$$\mathbf{s}_{g}^{q} = \frac{\mathbf{s}_{Rx}^{q}}{\mathbf{s}_{Tx}^{q}} = G_{k} \left( \mathbf{k}_{r}^{q} \otimes \mathbf{k}_{d}^{q} \right)$$
(11)

where  $\mathbf{k}_r^q = \left(0, e^{-j2\pi\Delta f \frac{2R_k}{c}}, \cdots, e^{-j2\pi(N_c^q-1)\Delta f \frac{2R_k}{c}}\right)$  and  $\mathbf{k}_d^q = \left(0, e^{j2\pi T_{\text{OFDM}}^q f_d(k)}, \cdots, e^{j2\pi(N_{sym}^q-1)T_{\text{OFDM}}^q f_d(k)}\right)$  are the two vectors carrying the Doppler and range information.  $\otimes$  refers to a dyadic product [29].

The Direction of Arrival (DoA) information matrix  $\mathbf{A}_{S}^{q}$  of all array elements is defined as

$$\mathbf{A}_{S}^{q} = \sum_{k} \mathbf{s}_{g}^{q} \mathbf{A}_{k}^{q} \tag{12}$$

It is worth noting that during the DL sensing period, the DoA satisfies the Far-field assumption that the AoAs are in the same direction and thus the vRx arrays can be used, while in the UL sensing period, the above assumption is not satisfied and thus only real aperture Rx arrays could be used [31].

The  $\mathbf{A}_{S}^{q}$  in our simulation work exists in the form of 4D arrays which contain the dimensions of fast time dimension, slow time dimension, horizontal antenna array dimension and vertical antenna array dimension. In actual ISAC 4D environmental reconstruction system,  $\mathbf{A}_{S}^{q}$  represents the environmental scatterer reflection signal received by UPA of BS.

# III. METHODOLOGY OF ISAC SENSING SIGNAL PROCESSING

The DL and UL ISAC sensing signal processing scheme is shown in Fig. 5. The multi-level fusion strategy and enhancement method of the sensing results will be shown in



Fig. 5. The illustration of DL and UL ISAC signal processing and multi-level fusion strategy.

Section IV. In this section, we first introduce the DL sensing signal processing algorithm, then we introduce the UL sensing signal processing algorithm. We abstract the description of the scenario in Fig. 1 as shown in Fig. 6.



Fig. 6. Simulation scenario modeling.

#### A. DL ISAC Sensing Signal Processing

In this section, we first introduce the Doppler and range estimation based on 2D-FFT and OSCA (Ordered Statistics & Cell Averaging)-CFAR algorithms, then introduce the DoA estimation based on 2D-MUSIC and CA (Cell Averaging)-CFAR algorithms.

1) Range and Doppler Estimation:  $\mathbf{k}_r^q, q = D$  is the linear phase shift vector along the fast time axis caused by range R. We initially process the received signal of each vRX antenna array element by using (11) and then performing

the Inverse Discrete Fourier Transform (IDFT) along the subcarrier dimension [29].

$$r^{D}(\alpha) = \text{IDFT}\left[k_{r}^{D}(n)\right] = \frac{1}{N_{c}^{D}} \sum_{n=0}^{N_{c}^{D}-1} k_{r}^{D}(n) e^{jn\alpha \frac{2\pi}{N_{c}^{D}}}$$
$$= \frac{1}{N_{c}^{D}} \sum_{n=0}^{N_{c}^{D}-1} e^{-j2\pi n\Delta f^{D} \frac{2R}{c}} e^{jn\alpha \frac{2\pi}{N_{c}^{D}}}, \alpha = 0, \dots, N_{c}^{D} - 1$$

where  $r^{D}(\alpha)$  obtains peak value when  $\alpha = \left\lfloor \frac{2R\Delta f^{\nu}N_{c}^{\nu}}{c} \right\rfloor$ .

 $\mathbf{k}_{d}^{q}, q = D$  is the linear phase shift vector along the slow time axis caused by velocity  $v_{k}$ . Similarly  $v_{k}$  can be estimated by performing a Discrete Fourier Transform (DFT) along the OFDM symbol dimension [29].

$$v^{D}(\beta) = \text{DFT}\left[k_{d}^{D}(m)\right] = \sum_{m=0}^{N_{sym}^{D}-1} k_{d}^{D}(m) e^{-jm\beta \frac{2\pi}{N_{sym}^{D}}}$$
$$= \sum_{m=0}^{N_{sym}^{D}-1} e^{j2\pi m T_{\text{OFDM}}^{D} \frac{2v_{k} f_{c}^{D}}{c}} e^{-jm\beta \frac{2\pi}{N_{sym}^{D}}}, \beta = 0, \dots, N_{sym}^{D} - 1$$
(14)

where  $v^D(\beta)$  obtains peak value when  $\beta$  in (14) satisfies  $\beta = \left| \frac{2v_k f_c^D T_{\text{OFDM}}^D N_{sym}^D}{c} \right|.$ 



Fig. 7. (a) is the sliding reference window design for 2D OSCA-CFAR, (b) and (c) are the Radar Doppler Map (RDM) and its detection threshold, respectively.

The RDM can be obtained by processing  $\mathbf{s}_g^D$  in (11) using the 2D-FFT algorithm, denoted as  $\mathbf{s}_g^{D,R,v}$ . As shown in Fig. 7 (a), we use the 2D OSCA-CFAR detection algorithm to estimate the range and velocity information of the scatterers by detecting the RDM. We select a reference window of size  $9 \times 9$  within the RDM to show the detection process.

- We apply 1D OS-CFAR to the sliding reference window by column and the  $\gamma$ th value of each colum is selected as the estimated noise intensity of this sliding window. We take  $\gamma = \lfloor 0.75N \rfloor$ , where N is the number of reference cells durning the sliding reference window.
- As shown in (15), the estimated values of noise intensity *μ*<sub>(γ)</sub> are averaged by row as the noise threshold of the Center Detection Unit (CUT) in reference window.

$$\bar{\mu}_{(\gamma)} = \frac{1}{N} \sum_{n=1}^{N} X_{(\gamma),n}$$
(15)

where  $X_{(\gamma),n}$  represents the noise reference value selected by OS-CFAR, since the size of the sliding reference window we select is  $9 \times 9$ , N = 9. • The detection threshold factor  $T_f$  of the selected CUT is only related to the number of reference units N with the specified false alarm probability  $P_{fa}$ , which is given by (16).

$$T_f = (P_{fa})^{-\frac{1}{N}} - 1 \tag{16}$$

We can calculate the detection threshold for the selected CUT T by  $T = T_f \times \bar{\mu}_{(\gamma)}$ , and the 2D adaptive detection threshold can be obtained by applying the above operations to all CUTs in the RDM as shown in Fig. 7.



Fig. 8. (a) is the sliding reference window design for 2D CA-CFAR, (b) and (c) are the MUSIC pseudo-spectrum and its detection threshold, respectively.

2) DoAs Estimation: The peak of the RDM indicates the presence of targets with velocity v and range R but  $\varphi$ ,  $\theta$  of the DoAs are unknown. The linear phase shift of the corresponding  $\mathbf{s}_{g}^{D,R,v}$  value due to R and v has been removed [29], which carries only the linear phase shift caused by the wave range difference of the receiving antenna array elements. As shown in (17), we combine the RDM peaks detected by 2D OSCA-CFAR of all receiving antenna elements into  $k_m$  ( $k_m < k$ ) manifolds for MUSIC-based DoA estimation.

$$\mathbf{A}_{k_{m}}^{D}\left(p_{r_{v}}, q_{r_{v}}\right) = \sum_{t} G_{t} \mathbf{A}_{t}^{D} = \sum_{t} G_{t} \Delta \phi_{r_{v}}(p_{r_{v}}, q_{r_{v}})_{\left(\theta_{t}, \varphi_{t}\right)} \quad (17)$$



Fig. 9. Using the pseudo-spectrum multiplication of two 1D-MUSIC searches instead of 2D-MUSIC searches to reduce complexity.

We take  $(\mathbf{A}_{k_m}^D)_{1,:}$  and  $(\mathbf{A}_{k_m}^D)_{:,1}$  of  $\mathbf{A}_{k_m}^D$  to construct the searching manifolds, which are related to the 2D angle  $\mathbf{p} = (\theta, \varphi)$  of the targets simultaneously, and eventually just multiply the two search results to get the 2D angle information of t targets in  $\mathbf{A}_{k_m}^D$ . The visualized image of the MUSIC search results of the above two sets of 1D line arrays and their multiplication is shown in Fig. 9.

The algorithmic flow of the MUSIC DoA search through  $(\mathbf{A}_{k_m}^D)_{:,1}$  is as follows, and the processing of  $(\mathbf{A}_{k_m}^D)_{1,:}$  can be obtained in the same way.

• We need to ensure the availability of the MUSIC algorithm by spatial smoothing algorithm to decoherence the individual scatterer echoes. As shown in Fig. 10, we can define the forward spatial smoothing matrix  $\mathbf{R}_f = \frac{1}{L} \sum_{l=1}^{L} \mathbf{R}_l^f$ , and similarly the backward spatial smoothing matrix  $\mathbf{R}_b = \frac{1}{L} \sum_{l=1}^{L} \mathbf{R}_l^b$ , where  $\mathbf{R}_l^f$  and  $\mathbf{R}_l^b$  are the covariance matrices of the subarray and L is the number of subarray elements. Then use the average of them  $\mathbf{R}_{\mathbf{X}} = \frac{\mathbf{R}_f + \mathbf{R}_b}{2}$  to replace the original covariance matrix.



Fig. 10. Schematic of the forward and backward space smoothing algorithm.

• We can get (18) by applying eigenvalue decomposition to  $\mathbf{R}_{\mathbf{X}}$ .

$$[\mathbf{U}_x, \mathbf{\Sigma}_x] = \operatorname{eig}(\mathbf{R}_{\mathbf{X}}) \tag{18}$$

where  $\Sigma_x$  is the real-valued eigenvalue diagonal matrix and  $U_x$  is the orthogonal eigenmatrix. We obtain the number of targets  $k_s$  by performing 1D CA-CFAR detection on  $\Sigma_x$ .

• Construct  $U_N = U_x (:, 1 : N_s - N_x)$  as the noise subspace basis. Then we can obtain the spatial angular spectrum function as [32]

$$f_{a}(\mathbf{p};\mathbf{U}_{N}) = \mathbf{a}^{\mathrm{H}}(\mathbf{p}) \mathbf{U}_{N}(\mathbf{U}_{N})^{\mathrm{H}}\mathbf{a}(\mathbf{p})$$
(19)

where  $\mathbf{p} = (\theta, \varphi)$  is the 2D angle,  $\mathbf{a} (\mathbf{p})$  is given in  $\mathbf{A}_k^D$  as shown in (2). The spatial pseudo-spectrum is represented as [32]

$$\mathbf{S}_{a}^{row}\left(\mathbf{p};\mathbf{U}_{N}\right)=\left[\mathbf{a}^{\mathsf{H}}\left(\mathbf{p}\right)\mathbf{U}_{N}\left(\mathbf{U}_{N}\right)^{\mathsf{H}}\mathbf{a}\left(\mathbf{p}\right)\right]^{-1}$$
(20)

Similarly we can obtain  $\mathbf{S}_{a}^{col}(\mathbf{p}; \mathbf{U}_{N})$  by taking the above operation for  $(\mathbf{A}_{k_{m}}^{D})_{1}$ .

We obtained the estimated DoAs by performing 2D CA-CFAR detection on  $\mathbf{S}_{a}(\mathbf{p}; \mathbf{U}_{N}) = \mathbf{S}_{a}^{col}(\mathbf{p}; \mathbf{U}_{N}) \odot \mathbf{S}_{a}^{row}(\mathbf{p}; \mathbf{U}_{N})$ ,  $\odot$  represents Hadamard product and  $\mathbf{S}_{a}(\mathbf{p}; \mathbf{U}_{N})$  is shown in Fig. 9.

As shown in Fig. 8(a), the noise of the 2D CA-CFAR is determined by averaging the  $N_R$  reference cells excluding the protection cell and the CUT.

$$\bar{\mu} = \frac{1}{N_R} \sum_{n=1}^{N_R} X_n$$
 (21)

where  $X_n$  represents the reference unit value. The detection threshold of this CUT can be calculated from  $T = T_f \times \bar{\mu}$ where  $T_f$  is given by (16), and the 2D adaptive detection threshold can be obtained by applying the above operations to all CUTs in  $\mathbf{S}_a(\mathbf{p}; \mathbf{U}_N)$  as shown in Fig. 8 (b)-(c).



Fig. 11. The geometric relationship between the UE, BS, and scatterer.

#### B. UL ISAC Sensing Signal Processing

The positional relationship of the scatterers and the UE Tx array does not satisfy the far-field assumption [31]. Therefore we can't receive the echoes of UL signal with the vRx array during the UL sensing period, but only with the BS-side real Rx array.

The UL sensing based on UL signal echoes from l NLos paths. The methodology for estimating the scatterers' 4D information is essentially the same as shown in III-A. As shown in Fig. 11, the DoAs of the UL signal echoes can be correctly estimated, but due to the time delay  $\tau$  of the lth path sourced with the two distances  $r_{l,1}$  and  $r_{l,2}$ , the scatterer's range information from the BS cannot be correctly estimated. The position of the scatterer of the lth path estimated by the BS in this case is the mirror position of the UE concerning the reflection plane, denoted as Virtual UE (vUE). We determine the real position of the scatterer by the geometrical relationship between the vUE, UE, and BS.

We denote the positions of the real scatterer, vUE, UE, and BS as  $\mathbf{x}_l = [x_l, y_l, z_l]$ ,  $\mathbf{x}_{vl} = [x_{vl}, y_{vl}, z_{vl}]$ ,  $\mathbf{x}_1 = [x_1, y_1, z_1]$ , and  $\mathbf{x}_0 = [x_0, y_0, z_0]$ , respectively. Then the midpoint of UE and vUE is  $\mathbf{M}_{1,vl} = \left[\frac{x_1+x_{vl}}{2}, \frac{y_1+y_{vl}}{2}, \frac{z_1+z_{vl}}{2}\right]$ . The normal vector  $\vec{u}_{1,vl}$  of a line segment in space consisting of UE and vUE is defined as

$$\vec{u}_{1,vl} = \frac{|x_{vl} - x_1, y_{vl} - y_1, z_{vl} - z_1|}{|[x_{vl} - x_1, y_{vl} - y_1, z_{vl} - z_1]|}$$
(22)

Substituting  $M_{1,vl}$  into (22) gives the equation of the perpendicular plane of the above line segment

$$\vec{u}_{1,vl} \cdot (\vec{r} - \mathbf{M}_{1,vl}) = 0 \tag{23}$$

The coordinate of the intersection of the line where BS and vUE are located and the above plane is the true position of the scatterer. The normal vector of the line between BS and vUE is given by

$$\vec{u}_{0,vl} = \frac{[x_{vl} - x_0, y_{vl} - y_0, z_{vl} - z_0]}{|[x_{vl} - x_0, y_{vl} - y_0, z_{vl} - z_0]|}$$
(24)

The parametric equation of the above line with unknown parameter t after substituting  $\mathbf{x}_0$  is defined as

$$\vec{Q}_{0,l} = \mathbf{x}_0 + t\vec{u}_{0,vl} \tag{25}$$

Joining (22) and (25) gives

$$\vec{u}_{0,vl} \cdot \left( (\mathbf{x}_0 + t\vec{u}_{0,vl}) - \mathbf{M}_{1,vl} \right) = 0 \tag{26}$$

# IV. METHODOLOGY OF SENSING RESULT FUSION AND ENHANCEMENT

In this section, we demonstrate the fusion method of DL-UL ISAC 4D environmental reconstruction results. Firstly, we show the data-level fusion strategy based on the simple solution of the theory of Sphere Packings; then we show the AGGRNN network model based on deep learning for improving the point cloud density of the environmental reconstruction results after data-level fusion; and finally, we present the feature-level fusion reconstruction network MVSFNet based on the multi-view feature-level fusion strategy.

#### A. Data Level Fusion Strategy



Fig. 12. (a) and (b) are the result of the original ISAC 4D environmental reconstruction and the data-level fusion result, respectively.

1) Cartesian Coordinate System Rotation and Unification: In order to realize the environmental reconstruction, it is necessary to restore their coordinates to the world coordinate system from the BS coordinate system by using the Eulerian rotation operators  $\mathbf{R}_{ZYX}$ .

$$\mathbf{R}_{ZYX} = \mathbf{R}_{Z}(\gamma)\mathbf{R}_{Y}(\beta)\mathbf{R}_{X}(\alpha)$$

$$= \begin{bmatrix} \cos\gamma & -\sin\gamma & 0\\ \sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta\\ 0 & 1 & 0\\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0\\ 0 & \cos\alpha & -\sin\alpha\\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$
(27)

The coordinates of a certain scatterer in the world coordinate system is given by

$$\mathbf{P}_{\text{World}} = \mathbf{P}_{\text{BS}} \times \mathbf{R}_{ZYX} + \mathbf{P}_{\text{Offset}}$$
(28)

where  $P_{Offset} = [x_o, y_o, z_o]$  is the positional offset of the BS with respect to the reference point of the world coordinate system,  $P_{BS} = [x, y, z]$  is the scatterer coordinates of the BS coordinate system, and  $P_{World} = [x', y', z']$  is the scatterer coordinates of the transformed world coordinate system; and  $\alpha$ ,  $\beta$ , and  $\gamma$  are the angles of rotation of the BS coordinate system around the x, y, and z axes, respectively.

2) Offset Scatterers Fusion: The ISAC 4D environmental reconstruction results after directly stitching the sensing results from multiple nodes are shown in Fig. 12(a). It can be seen that the coordinates of the same scatterer are slightly shifted. The key to data-level fusion is to select the appropriate fusion radius  $R_c$ .



Fig. 13. Schematic of the relationship between the fusion radius of the covered sphere and the growth rate of the number of points in the sphere.

We randomly select 10 scatterers and create expanding covering spheres with radius  $R_c$ , and we define the growth rate  $h_i$  of the number of scatterers in the covering sphere as a monitoring metric as

$$h_i = \frac{C_n - C_b}{C_b} \tag{29}$$

where  $C_n$  is the number of scatterers in the current coverage sphere and  $C_n$  is the number of scatterers in the coverage sphere at the previous differential radius. We calculate the  $h_i$ variation curve when  $R_c \in [0, 10]$  as shown in Fig. 13.

Fig. 13 shows that this interval is an selectable interval for the fusion radius  $R_c$  when the radius of all covered spheres increases but the growth rate of the number of scatterers in the sphere satisfied  $h_i \rightarrow 0$ .

# B. AGGRNN: Low-density Reconstruction Point Cloud Enhancing Method

As shown in Fig. 14, AGGRNN contains four components are Gridding module (GM), Gridding Feature Extraction Module (GFEM), Detail Recovery Module (DRM) and Gridding Reverse Module (GRM). GM first transforms the input lowdensity point cloud into a 3D grid  $\xi = \langle V, K \rangle$ , where V and K are the vertex set and value set of  $\xi$ , respectively. Then GFEM extracts  $\xi$ 's high dimensional spatial features output as K' through AGUNet. The DRM then works with the GRM to generate the spatial features  $F^c$  of the coarse point cloud  $P^c$  and uses the Multi-Layer Perceptron (MLP) to generate the adjustment vector  $W^c$  of  $P^c$ . Finally, GRM superimposes the adjustment vector  $W^c$  on the generated coarse point cloud  $P^c$  to output the high-density point cloud  $P^f$  [33]. The Loss function for our supervised model training is GriddingLoss.

1) Gridding Module (GM): In GM we introduce 3D grids as intermediate representations of regularized point clouds, and further introduce a differentiable Gridding layer that converts a disordered, irregular point cloud  $P = \{p_i\}_{i=1}^n$  into an ordered, regular point cloud  $\xi = \langle V, K \rangle$  while preserving the spatial layout of the point cloud, where  $p_i \in \mathbb{R}^3$ ,  $V = \{v_i\}_{i=1}^{N^3}$ ,  $K = \{k_i\}_{i=1}^{N^3}$ ,  $v_i \in \{(-\frac{N}{2}, -\frac{N}{2}, -\frac{N}{2}), ..., (-\frac{N}{2} - 1, -\frac{N}{2} - 1, -\frac{N}{2} - 1)\}$ ,  $k_i \in \mathbb{R}$ , n is the number of points in P, and N is the resolution of the 3D grid  $\xi$ . As shown in Fig. 15(a), the 3D Grid contains

a number of Cells, and each Cell contains 8 vertices, for each vertex  $v_i = (x_i^v, y_i^v, z_i^v)$  in the Cell, if the point p = (x, y, z) in the low-density point cloud satisfies 30,then p falls in the neighborhood  $\mathcal{N}$  of vertex  $v_i$ , denoted as  $p \in \mathcal{N}(v_i)$ .

$$\begin{cases} x_i^v - 1 < x < x_i^v + 1 \\ y_i^v - 1 < y < y_i^v + 1 \\ z_i^v - 1 < z < z_i^v + 1 \end{cases}$$
(30)

As shown in Fig. 15(a), for a given vertex  $v_i$  with  $\mathcal{N}$ , the corresponding weight  $k_i$  of the vertex is defined as

$$k_{i} = \sum_{p \in \mathcal{N}(v_{i})} \frac{k(v_{i}, p)}{|\mathcal{N}(v_{i})|}$$
(31)

where  $k_i(v_i, p) = (1 - |x_i^v - x|) (1 - |y_i^v - y|) (1 - |z_i^v - z|)$ , in particular, for $|\mathcal{N}(v_i)| = 0, k_i = 0$ .

2) Gridding Feature Extraction Module (GFEM): GFEM aims to extract the depth features of low-density point clouds in terms of spatial arrangement and spatial distribution. Its basic structure is AGUNet, which is a 3D encoder-decoder structure introducing AG modules with U-net connections [34]. The details of the network parameters and network structure are shown in Fig. 14. For a given set K of values of a 3D Grid, the GFEM process can be given formally as K' = GFEM(K), where  $K' = \{k'_i\}_{i=1}^{N^3}, k' \in \mathbb{R}$ .

3) Detail Recovery Module(DRM): DRM is composed of Cubic Feature Sampling (CFS) layer and MLP. We introduce a CFS layer [33] to extract features at different depths of the point cloud from the up-sampled 3D feature maps and preserve the contextual information.

As shown in Fig. 16, the feature set  $\mathcal{F} = \{f_1^v, f_2^v, ..., f_{t^3}^v | f_i^v \in \mathbb{R}^c\}$  of a certain feature map of AGUNet and a certain point  $p_i^c$  in the coarse point cloud  $P^c$  are given, and the feature  $f_i^c$  corresponding to  $p_i^c$  can be denoted as  $f_i^c = \left[f_{\theta_1^i}^v, f_{\theta_2^i}^v, ..., f_{\theta_8^i}^v\right]$ , where [·] is the concatenation operation,  $\{f_{\theta_j^i}^v\}_{j=1}^8$  denotes the set of features of the eight vertices of the Cell where  $p_i^c$  is located, the coarse point cloud  $P^c$  is obtained from the GRM in Section IV-B4. We randomly sample 2048 points from the coarse point cloud  $P^c$ , and the CFS layer extracts features from the three up-sampled feature maps of AGUNet and generates the feature  $F^c$  of the above 2048 points, as shown in Fig. 16.

The MLP estimates r offsets for each point  $p_i^c$  to generate the final point cloud. Specifically, given a coarse point cloud  $P^c$  with its corresponding feature  $F^c$ , the MLP and GM eventually collaborate to generate a high-density point cloud  $P^f = \left\{ p_i^f \right\}_{i=1}^n$ .

$$P^{f} = \mathrm{MLP}\left(F^{c}\right) + \mathrm{GM}\left(P^{c}, r\right)$$
(32)

where  $p_i^f \in \mathbb{R}^3$ , *n* is the number of points in the final high-density point cloud  $P^f$ . GM creates a new tensor by replicating  $P^c r$  times. As shown in Fig. 14, we take r = 8 and the output of MLP is reshaped to  $r \times m \times 3 = 16384 \times 3$ , which corresponds to the offsets of the coordinates of 16384 points.



Fig. 14. Overall framework of AGGRNN.



Fig. 15. (a) and (b) are Gridding and Gridding Reverse, respectively.



Fig. 16. Overall framework of CFS layer.

4) Gridding Reverse Module (GRM): GM consists of GriddingReverse with tail processing of AGGRNN. GriddingReverse is the inverse of Gridding which converts the 3D Grid back to a point cloud. GriddingReverse takes the 3D Grid  $\xi' = \langle V, K' \rangle$  generated by AGUNet as input to generate a coarse point cloud  $P^c = \{p_i^c\}_{i=1}^m$ , where  $p_i^c \in \mathbb{R}^3$  and m is the number of points in the coarse point cloud  $P^c$ . As shown in Fig. 15(b), GriddingReverse generates a point  $\xi'$ , whose coordinates can be obtained by weighting the coordinates of the 8 vertices of the current Cell with the weights.

$$p_i^c = \frac{\sum_{\theta \in \Theta^i} w'_{\theta} v_{\theta}}{\sum_{\theta \in \Theta^i} w'_{\theta}}$$
(33)

where  $\Theta^{i} = \{\theta_{j}^{i}\}_{j=1}^{8}$  represents the set of subscript indexes of the vertices of this Cell,  $\{v_{\theta} | \theta \in \Theta^{i}\}$  and  $\{w'_{\theta} | \theta \in \Theta^{i}\}$ 

represent the set of vertices and values of the Cell, respectively. In particular, if  $\sum_{\theta \in \Theta^i} w'_{\theta} = 0$ , then *GriddingReverse* will not generate a point  $p_i^c$  for the Cell. The tail processing section is described in Section IV-B3 and will not be repeated here.



Fig. 17. Gridding Loss.

5) Gridding Loss: We introduce a new loss function GriddingLoss based on GM, defined as the L1 distance between two 3D Grids value sets  $\xi_{Pred} = \langle V^{Pred}, K^{Pred} \rangle$  and  $\xi_{GT} = \langle V^{GT}, K^{GT} \rangle$ , respectively [33].

$$\mathcal{L}_{Gridding}\left(K^{Pred}, K^{GT}\right) = \frac{1}{N_G^3} \sum \left\|K^{Pred} - K^{GT}\right\|$$
(34)

Where  $K^{Pred} \in \mathbb{R}^{N_G^3}, K^{GT} \in \mathbb{R}^{N_G^3}, N_G$  is the resolution of the two 3D Grids.

#### C. MVSFNet: Feature Level Fusion Strategy

As shown in Fig. 18, MVSFNet consists of two parts, Multi-Viewpoint Feature Extraction Module (MVFEM) and Multi-Viewpoint Feature Fusion Module (MVFFM).

1) Multi-Viewpoint Feature Extraction Module (MVFEM): The coarse point cloud set  $\mathcal{P} = \{P_g^c, P_1^c, ..., P_{31}^c\}$  of reconstruction results from different viewpoints can be obtained after GM, GFEM, and GriddingReverse, where  $P_g^c$  is the coarse point cloud generated from the global low-density point cloud after data-level fusion,  $P_{\eta}^c$  ( $\eta \in [1, 31]$ ) is the coarse



Fig. 18. Overall framework of MVSFNet.

point cloud generated from a multi-view low-density point cloud.  $F_{full}^c = [F_g^c, F_1^c, ..., F_{31}^c]$ , where [·] is the concatenation operation,  $F_g^c$  corresponds to the feature map of global low-density point cloud  $P_g^c$  after data-level fusion,  $F_\eta^c$  ( $\eta \in [1, 31]$ ) corresponds to the feature map of multi-view low-density point cloud.

We specify the maximum number of channels of  $F_{full}^c$  to be 32, so the maximum number of viewpoints supported by MVSFNet is 31 after removing the stationary-occupied global feature map  $F_g^c$ . If the number of viewpoints is less than 31, the corresponding  $F_\eta^c$  ( $\eta \in [1, 31]$ ) adopts an all-zero Tensor.

2) Multi-Viewpoint Feature Fusion Module (MVFFM): We use a two-dimensional encoder-decoder structure for fusing multi-view feature maps in MVFEM, replacing the stationary feature map weights by trainable network parameters. The specific network parameters and network structure of MVFFM are shown in Fig. 18. It is worth noting that the point cloud that is complemented in the tail processing is still the coarse point cloud  $P_a^c$  from the global low-density point cloud.

# V. SIMULATION RESULTS AND ANALYSIS

This section first introduces the evaluation metrics of MN-DUC ISAC 4D environmental reconstruction method; then introduces the methods of generating the training and evaluation data; then introduces the parameter settings of the simulation; and finally, the qualitative and quantitative evaluations of the MNDUC ISAC 4D environmental reconstruction method proposed in this research are conducted.

## A. Dataset Generation

As shown in Fig. 19(a), we first randomly generate stationary buildings and moving vehicles of different sizes but not overlapping each other in a fixed-size 3D space, and then randomly place sensing nodes such as BSs, UEs, UAVs, and so on, in the areas that are not covered by the buildings and vehicles. As shown in Fig. 19(b), we construct a set



Fig. 19. (a)-(d) show the generation process of our dataset.

of scatterers in the scenario space by extracting scattering points on the surfaces of buildings and vehicles. As shown in Fig. 20(a)(c)(e), we calculate the scatterers locate on the LoS paths and single reflection NLoS paths in DL and UL period of different sensing nodes by Raytracing and occlusion detection algorithms. The environmental reconstruction results of the corresponding selected sensing nodes are shown in Fig. 20(b)(d)(f), and their direct fusion results are shown in Fig. 19(c). As shown in Fig. 19(d), we apply the data-level fusion strategy in Section IV-A to process the direct fusion results to obtain a low-density reconstructed point cloud as the training data for AGGRNN. For the Multi-Modal Model MVSFNet, one of its inputs is the low-density point cloud training data of the AGGRNN described above, and the other input is the feature maps extracted from the multi-view sensing result, and the extraction method is described in Section IV-C1.

The specifications of this dataset are 10,000 data for the training set, 100 data for the validation set, and 100 data for the test set. And it is labeled in ShapeNet format.

## B. Simulation Parameters and Environment Setting

TABLE I SIMULATION PARAMETERS OF SYSTEM FOR DL AND UL PERIODS.

Parameter names	DL	UL
MIMO size	$\begin{vmatrix} 8\times8 \text{ Tx}, \ 2\times2 \text{ Rx} \\ 16\times16 \text{ vRx} \end{vmatrix}$	$1 \times 1$ Tx, $8 \times 8$ Rx
Carrier frequency	70 GHz	70 GHz
Bandwidth	491.52 MHz	209.76 MHz
Subcarrier spacing	240 kHz	240 kHz
Subcarrier count	2048	1024
OFDM symbol count	224	224
Slot count	16	16
Spatially smoothed subarray size	8	4

The signal and array parameters in the DL period and the UL period are set as shown in Table I.

The experiments are all implemented on an  $Intel^{(0)} Core^{TM}$ i9-13900KF CPU (4.3GHz, 64GB RAM) and an NVIDIA GeForce RTX 4090 GPU (CUDA version 11.6) with Python 3.8.18 (PyTorch 2.1.2) in 64 Bit Ubuntu 22.04.1 Long Term Support operating system. We select the Adaptive Moment Estimation (Adam) optimizer with  $\beta_1 = 0.9, \beta_2 = 0.99$  for finding the optimal parameters, the initial learning rate is 0.00001 and the *Batchsize* is set to 32.

# C. Evaluation metrics

In our experiments, we use both Chamfer Distance (CD) and F-Score as quantitative evaluation metrics.

1) Chamfer Distance: The distance between  $\mathcal{T}$  and  $\mathcal{R}$  are defined as (35) [35], [36]

$$\mathcal{D}_{\rm CD} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \min_{r \in \mathcal{R}} \|t - r\|_2^2 + \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \min_{t \in \mathcal{T}} \|t - r\|_2^2$$
(35)

where  $\mathcal{T} = \{(x_i, y_i, z_i)\}_{i=1}^{n\tau}$  is the Ground Truth(GT) and  $\mathcal{R} = \{(x_i, y_i, z_i)\}_{i=1}^{n_{\mathcal{R}}}$  is the reconstructed point set being evaluated,  $n_{\mathcal{T}}$  and  $n_{\mathcal{R}}$  are the numbers of points in  $\mathcal{T}$  and  $\mathcal{R}$ , respectively.

2) F-Score: We introduce F-Score as an extended metric to evaluate the performance of low-density point cloud enhancement networks, which can be defined as [33]

$$F-Score(d) = \frac{2P(d) R(d)}{P(d) + R(d)}$$
(36)

where  $P(d) = \frac{1}{n_{\mathcal{R}}} \sum_{r \in \mathcal{R}} \left[ \min_{t \in \mathcal{T}} ||t - r|| < d \right]$  and  $R(d) = \frac{1}{n_{\mathcal{T}}} \sum_{t \in \mathcal{T}} \left[ \min_{r \in \mathcal{R}} ||t - r|| < d \right]$  are the precision and recall for a

distance threshold d, respectively.

# D. Qualitative Analysis of Simulation Results

1) Comparison of Single Node and Multi-Node Cooperative ISAC 4D Environmental Reconstruction Results: The environmental reconstruction results of some sensing nodes



(e) UL, Original, UAV1 to BS2. (f) UL, Result, UAV1 to BS2.

Fig. 20. Original points cloud and reconstruction result of a few sensing nodes in the DL and UL periods.

are shown in Fig. 20, and the multi-node cooperative sensing environmental reconstruction results are shown in Fig. 12.

By comparing the single-node and multi-node sensing results, we can find that the multi-node cooperative ISAC 4D environmental reconstruction scheme can well overcome the problem of reconstructed point cloud disability caused by the single-node case due to the singularity of viewpoints.

2) Comparison of Low-Density Point Cloud Enhancement Results: The enhancement performance for the selected randomly generated scenarios of the proposed low-density reconstruction point cloud enhancement network AGGRNN with MVSFNet is shown in Fig. 21. The simulation results show that,

- · Firstly, the results of the data-level fusion multi-node UL-DL collaborative ISAC 4D environmental reconstruction will lose a large amount of surface details, although it can basically depict the contours of the environment's scatterers.
- Second, the data-level fusion reconstructed point cloud • enhanced by AGGRNN is significantly higher in density and can recover partial environmental surface details.
- Finally, MVSFNet with multi-view feature-level fusion can recover more surface details compared with AG-GRNN, and reconstruction deviation is smaller compared with the original scenario.

# E. Quantitative Analysis of Simulation Results

1) Comparative Experiments on High-Density Point Cloud Generation: The quantitative results in Table II show that



Fig. 21. Selected images of original 3D scenario, low-density data-level fusion point cloud, AGGRNN reconstruction results, and MVSFNet reconstruction results.

TABLE II LOW-DENSITY RECONSTRUCTION POINT CLOUD ENHANCEMENT RESULTS ON THE GENERATED DATASET COMPARED USING F-SCORE @1% and CD.

Methods	F-Score@1%	CD
AtlasNet [37]	0.566	0.268
PCN [38]	0.589	0.314
FoldingNet [39]	0.237	0.499
TopNet [35]	0.414	0.336
MSN [40]	0.601	0.287
GRNet [33]	0.613	0.352
Data-level Fusion(Ours)	-	0.517
AGGRNN(Ours)	0.652	0.270
MVSFNet(Ours)	0.751	0.191

our proposed Multi-Modal model MVSFNet outperforms the other compared schemes in both Chmafer Distance and F-Score@1% metrics, and outperforms our proposed AGGRNN without supplemental local features.

2) Ablation Study: In order to demonstrate the superiority of our proposed model more completely, we will perform ablation study for transfer learning and fusion&enhancement strategies.

• We perform an ablation study for transfer learning of MVSFNet, where the GFEM of MVSFNet loads a pretrained weight file based on the ShapeNet dataset [41] when using transfer learning, and conversely without using that pre-trained weights. The quantitative results in Table III show that transfer learning can optimize the performance of MVSFNet with other experimental conditions being the same.

TABLE III Ablation Study (Transfer Learning)

Transfer Learning	F-Score@1%	CD
	0.716	0.238
$\checkmark$	0.751	0.191

• We perform ablation study for Fusion&Enhancement strategies, where MVSFNet cannot perform ablation study alone since the feature-level fusion strategy MVSFNet is based on the enhancement strategy AG-GRNN. Quantitative results show in Table IV that the ISAC environmental reconstruction performance is best when the multi-level fusion strategy is applied together with the low-density point cloud enhancement strategy, which proves the superiority and rationality of our proposed method. It is worth noting that direct low-density point cloud enhancement without applying a data-level fusion strategy can lead to deterioration of the environmental reconstruction performance, since the unfused point cloud accumulates weights to the corresponding 3D Grids, which in turn leads to disruption of the overall spatial structure.

TABLE IV Ablation Study (Fusion & Enhancement)

Fusion&Enhancement		F-Score@1%	CD	
Data-level Fusion	Enhancement (AGGRNN)	Feature-level Fusion(MVSFNet)		
~			-	0.517
$\checkmark$	$\checkmark$		0.652	0.270
	$\checkmark$		0.545	0.391
	$\checkmark$	$\checkmark$	0.674	0.332
$\checkmark$	$\checkmark$	$\checkmark$	0.751	0.191

# VI. CONCLUSION

We propose a deep learning based MNDUC ISAC 4D environmental reconstruction method in this research, which is based on the MUSIC algorithm in the echo processing part, and introduces multi-node UL and DL cooperation as well as multi-level fusion strategy, and optimizes the results of the environmental reconstruction by the deep learning models AGGRNN and MVSFNet. We have performed a large number of comparison and ablation studies for the proposed method, and the experimental results show that

- The proposed multi-node cooperative sensing scheme combining the active and passive sensing can make up for the shortcomings of the single-node scheme's single viewpoint, and realize the comprehensive, multilevel, multifaceted environmental reconstruction results;
- The proposed reconstruction points cloud density enhancement networks AGGRNN and MVSFNet can sig-

nificantly reduce the gap between the multi-node cooperative ISAC 4D reconstruction results and the real scenario, and reduce the dependence of ISAC-based 4D environmental reconstruction methods on frequency bands and bandwidth resources.

#### REFERENCES

- [1] Z. Wei, H. Qu, Y. Wang, X. Yuan, H. Wu, Y. Du, K. Han, N. Zhang, and Z. Feng, "Integrated sensing and communication signals toward 5ga and 6g: A survey," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11068–11092, 2023.
- [2] S. Sun, A. P. Petropulu, and H. V. Poor, "Mimo radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98– 117, 2020.
- [3] A. Santra, A. R. Ganis, J. Mietzner, and V. Ziegler, "Ambiguity function and imaging performance of coded fmcw waveforms with fast 4d receiver processing in mimo radar," *Digital Signal Processing*, vol. 97, p. 102618, 2020.
- [4] S. Sun and Y. D. Zhang, "4d automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE Journal of Selected Topics* in Signal Processing, vol. 15, no. 4, pp. 879–891, 2021.
- [5] Y. Niu, Z. Wei, L. Wang, H. Wu, and Z. Feng, "Interference management for integrated sensing and communication systems: A survey," *arXiv* preprint arXiv:2403.16189, 2024.
- [6] C. Meng, Z. Wei, and Z. Feng, "Adaptive waveform optimization for mimo integrated sensing and communication systems based on mutual information," in 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2022, pp. 472– 477.
- [7] X. Chen, Z. Feng, Z. Wei, X. Yuan, P. Zhang, J. A. Zhang, and H. Yang, "Multiple signal classification based joint communication and sensing system," *IEEE Transactions on Wireless Communications*, 2023.
- [8] L. Wang, Z. Wei, L. Su, Z. Feng, H. Wu, and D. Xue, "Coherent compensation based isac signal processing for long-range sensing," 2023.
- [9] C. Meng, D. Ma, X. Chen, Z. Feng, and Y. Liu, "Cramer-rao bounds for near-field sensing: A generic modular architecture," *arXiv preprint* arXiv:2404.07472, 2024.
- [10] Z. Wei, H. Liu, X. Yang, W. Jiang, H. Wu, X. Li, and Z. Feng, "Carrier aggregation enabled integrated sensing and communication signal design and processing," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 3, pp. 3580–3596, 2024.
- [11] X. Lu, Z. Wei, R. Xu, L. Wang, B. Lu, and J. Piao, "Integrated sensing and communication enabled multiple base stations cooperative uav detection," 2024.
- [12] J. Guan, A. Paidimarri, A. Valdes-Garcia, and B. Sadhu, "3-d imaging using millimeter-wave 5g signal reflections," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 6, pp. 2936–2948, 2021.
- [13] Z. Wei, H. Liu, Z. Feng, H. Wu, F. Liu, and Q. Zhang, "Deep cooperation in isac system: Resource, node and infrastructure perspectives," *arXiv* preprint arXiv:2403.02565, 2024.
- [14] Z. Wei, R. Xu, Z. Feng, H. Wu, N. Zhang, W. Jiang, and X. Yang, "Symbol-level integrated sensing and communication enabled multiple base stations cooperative sensing," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 724–738, 2024.
- [15] M. Ren, P. He, and J. Zhou, "Improved shape-based distance method for correlation analysis of multi-radar data fusion in self-driving vehicle," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24771–24781, 2021.
- [16] B. Nuss, Y. L. Sit, and T. Zwick, "3d radar image fusion using ofdmbased mimo radar," in 2016 German microwave conference (GeMiC). IEEE, 2016, pp. 209–212.
- [17] B. Gao, M. Jia, T. Zhang, and Q. Zhang, "Reliable target positioning in complicated environments using multiple radar observations," in 2021 *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.
- [18] A. J. Weiss, "Direct geolocation of wideband emitters based on delay and doppler," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2513–2521, 2011.
- [19] A. Weiss, "Direct position determination of narrowband radio frequency transmitters," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 513– 516, 2004.

- [20] Y. Wang, C. Yang, Z. Ren, Y. Sun, and M. Peng, "Sensing-aided hybrid precoding for efficient terahertz wideband communications in multiuser high-data-rate iot," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8253–8267, 2024.
- [21] X. Yang, Z. Wei, Y. Liu, H. Wu, and Z. Feng, "Ris-assisted cooperative multicell isac systems: A multi-user and multi-target case," *IEEE Tran. Wireless Commun.*, Early Access, 2024.
- [22] X. Zhang, P. Liu, B. Lu, Y. Wang, X. Chen, Y. Zhang, and Z. Wang, "Mtsfbet: A hand-gesture-recognition-based identity authentication approach for passive keyless entry against relay attack," *IEEE Transactions* on Mobile Computing, vol. 23, no. 2, pp. 1902–1913, 2024.
- [23] X. Wang, M. H. Ang Jr, and G. H. Lee, "Cascaded refinement network for point cloud completion," in *Proceedings of the IEEE/CVF conference* on computer vision and pattern recognition, 2020, pp. 790–799.
- [24] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," ACM Transactions on Graphics (tog), vol. 38, no. 5, pp. 1–12, 2019.
- [25] K. Wang, K. Chen, and K. Jia, "Deep cascade generation on point sets." in *IJCAI*, vol. 2, 2019, p. 4.
- [26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.
- [27] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6411–6420.
- [28] B. Lu, Z. Wei, L. Wang, R. Zhang, and Z. Feng, "Isac 4d imaging system based on 5g downlink millimeter wave signal," in 2023 IEEE Globecom Workshops (GC Wkshps), 2023, pp. 389–394.
- [29] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.
- [30] X. Chen, Z. Feng, J. A. Zhang, Z. Wei, X. Yuan, P. Zhang, and J. Peng, "Downlink and uplink cooperative joint communication and sensing," *IEEE Transactions on Vehicular Technology*, pp. 1–15, 2024.
- [31] C. A. Balanis, *Antenna theory: analysis and design*. John wiley & sons, 2016.
- [32] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Academic Press Library in Signal Processing*. Elsevier, 2014, vol. 3, pp. 651–717.
- [33] H. Xie, H. Yao, S. Zhou, J. Mao, S. Zhang, and W. Sun, "Grnet: Gridding residual network for dense point cloud completion," in *European Conference on Computer Vision*. Springer, 2020, pp. 365–381.
- [34] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19.* Springer, 2016, pp. 424–432.
- [35] L. P. Tchapmi, V. Kosaraju, H. Rezatofighi, I. Reid, and S. Savarese, "Topnet: Structural point cloud decoder," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 383–392.
- [36] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 605–613.
- [37] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "A papier-mâché approach to learning 3d surface generation," in *Proceed*ings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 216–224.
- [38] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "Pcn: Point completion network," in 2018 international conference on 3D vision (3DV). IEEE, 2018, pp. 728–737.
- [39] Y. Yang, C. Feng, Y. Shen, and D. Tian, "Foldingnet: Point cloud auto-encoder via deep grid deformation," in *Proceedings of the IEEE* conference on computer vision and pattern recognition, 2018, pp. 206– 215.
- [40] M. Liu, L. Sheng, S. Yang, J. Shao, and S.-M. Hu, "Morphing and sampling network for dense point cloud completion," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11596–11603.
- [41] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings* of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1912–1920.

#### REFERENCES

- [1] Z. Wei, H. Qu, Y. Wang, X. Yuan, H. Wu, Y. Du, K. Han, N. Zhang, and Z. Feng, "Integrated sensing and communication signals toward 5ga and 6g: A survey," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11068–11092, 2023.
- [2] S. Sun, A. P. Petropulu, and H. V. Poor, "Mimo radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98– 117, 2020.
- [3] A. Santra, A. R. Ganis, J. Mietzner, and V. Ziegler, "Ambiguity function and imaging performance of coded fmcw waveforms with fast 4d receiver processing in mimo radar," *Digital Signal Processing*, vol. 97, p. 102618, 2020.
- [4] S. Sun and Y. D. Zhang, "4d automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE Journal of Selected Topics* in Signal Processing, vol. 15, no. 4, pp. 879–891, 2021.
- [5] Y. Niu, Z. Wei, L. Wang, H. Wu, and Z. Feng, "Interference management for integrated sensing and communication systems: A survey," *arXiv* preprint arXiv:2403.16189, 2024.
- [6] C. Meng, Z. Wei, and Z. Feng, "Adaptive waveform optimization for mimo integrated sensing and communication systems based on mutual information," in 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2022, pp. 472– 477.
- [7] X. Chen, Z. Feng, Z. Wei, X. Yuan, P. Zhang, J. A. Zhang, and H. Yang, "Multiple signal classification based joint communication and sensing system," *IEEE Transactions on Wireless Communications*, 2023.
- [8] L. Wang, Z. Wei, L. Su, Z. Feng, H. Wu, and D. Xue, "Coherent compensation based isac signal processing for long-range sensing," 2023.
- [9] C. Meng, D. Ma, X. Chen, Z. Feng, and Y. Liu, "Cramer-rao bounds for near-field sensing: A generic modular architecture," *arXiv preprint* arXiv:2404.07472, 2024.
- [10] Z. Wei, H. Liu, X. Yang, W. Jiang, H. Wu, X. Li, and Z. Feng, "Carrier aggregation enabled integrated sensing and communication signal design and processing," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 3, pp. 3580–3596, 2024.
- [11] X. Lu, Z. Wei, R. Xu, L. Wang, B. Lu, and J. Piao, "Integrated sensing and communication enabled multiple base stations cooperative uav detection," 2024.
- [12] J. Guan, A. Paidimarri, A. Valdes-Garcia, and B. Sadhu, "3-d imaging using millimeter-wave 5g signal reflections," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 6, pp. 2936–2948, 2021.
- [13] Z. Wei, H. Liu, Z. Feng, H. Wu, F. Liu, and Q. Zhang, "Deep cooperation in isac system: Resource, node and infrastructure perspectives," *arXiv* preprint arXiv:2403.02565, 2024.
- [14] Z. Wei, R. Xu, Z. Feng, H. Wu, N. Zhang, W. Jiang, and X. Yang, "Symbol-level integrated sensing and communication enabled multiple base stations cooperative sensing," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 724–738, 2024.
- [15] M. Ren, P. He, and J. Zhou, "Improved shape-based distance method for correlation analysis of multi-radar data fusion in self-driving vehicle," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24771–24781, 2021.
- [16] B. Nuss, Y. L. Sit, and T. Zwick, "3d radar image fusion using ofdmbased mimo radar," in 2016 German microwave conference (GeMiC). IEEE, 2016, pp. 209–212.
- [17] B. Gao, M. Jia, T. Zhang, and Q. Zhang, "Reliable target positioning in complicated environments using multiple radar observations," in 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021, pp. 1–6.
- [18] A. J. Weiss, "Direct geolocation of wideband emitters based on delay and doppler," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2513–2521, 2011.
- [19] A. Weiss, "Direct position determination of narrowband radio frequency transmitters," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 513– 516, 2004.
- [20] Y. Wang, C. Yang, Z. Ren, Y. Sun, and M. Peng, "Sensing-aided hybrid precoding for efficient terahertz wideband communications in multiuser high-data-rate iot," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8253–8267, 2024.
- [21] X. Yang, Z. Wei, Y. Liu, H. Wu, and Z. Feng, "Ris-assisted cooperative multicell isac systems: A multi-user and multi-target case," *IEEE Tran. Wireless Commun.*, Early Access, 2024.
- [22] X. Zhang, P. Liu, B. Lu, Y. Wang, X. Chen, Y. Zhang, and Z. Wang, "Mtsfbet: A hand-gesture-recognition-based identity authentication ap-

proach for passive keyless entry against relay attack," *IEEE Transactions on Mobile Computing*, vol. 23, no. 2, pp. 1902–1913, 2024.

- [23] X. Wang, M. H. Ang Jr, and G. H. Lee, "Cascaded refinement network for point cloud completion," in *Proceedings of the IEEE/CVF conference* on computer vision and pattern recognition, 2020, pp. 790–799.
- [24] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," ACM Transactions on Graphics (tog), vol. 38, no. 5, pp. 1–12, 2019.
- [25] K. Wang, K. Chen, and K. Jia, "Deep cascade generation on point sets." in *IJCAI*, vol. 2, 2019, p. 4.
- [26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.
- [27] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6411–6420.
- [28] B. Lu, Z. Wei, L. Wang, R. Zhang, and Z. Feng, "Isac 4d imaging system based on 5g downlink millimeter wave signal," in 2023 IEEE Globecom Workshops (GC Wkshps), 2023, pp. 389–394.
- [29] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.
- [30] X. Chen, Z. Feng, J. A. Zhang, Z. Wei, X. Yuan, P. Zhang, and J. Peng, "Downlink and uplink cooperative joint communication and sensing," *IEEE Transactions on Vehicular Technology*, pp. 1–15, 2024.
- [31] C. A. Balanis, *Antenna theory: analysis and design.* John wiley & sons, 2016.
- [32] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Academic Press Library in Signal Processing*. Elsevier, 2014, vol. 3, pp. 651–717.
- [33] H. Xie, H. Yao, S. Zhou, J. Mao, S. Zhang, and W. Sun, "Grnet: Gridding residual network for dense point cloud completion," in *European Conference on Computer Vision*. Springer, 2020, pp. 365–381.
  [34] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ron-
- [34] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19.* Springer, 2016, pp. 424–432.
- [35] L. P. Tchapmi, V. Kosaraju, H. Rezatofighi, I. Reid, and S. Savarese, "Topnet: Structural point cloud decoder," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 383–392.
- [36] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 605–613.
- [37] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "A papier-mâché approach to learning 3d surface generation," in *Proceed*ings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 216–224.
- [38] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "Pcn: Point completion network," in 2018 international conference on 3D vision (3DV). IEEE, 2018, pp. 728–737.
- [39] Y. Yang, C. Feng, Y. Shen, and D. Tian, "Foldingnet: Point cloud auto-encoder via deep grid deformation," in *Proceedings of the IEEE* conference on computer vision and pattern recognition, 2018, pp. 206– 215.
- [40] M. Liu, L. Sheng, S. Yang, J. Shao, and S.-M. Hu, "Morphing and sampling network for dense point cloud completion," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11596–11603.
- [41] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings* of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1912–1920.