Deep Learning Based CFO Estimation for Multi-user Massive MIMO with One-Bit ADCs

Yunqi Feng, Kangkai Zhou, Huimei Han, Weidang Lu, *Senior Member, IEEE*, Jie Tang, *Senior Member, IEEE*, Nan Zhao, *Senior Member, IEEE*, and Arumugam Nallanathan, *Fellow, IEEE*

Abstract—Low-resolution architectures represent a compelling and power-efficient approach for high-bandwidth communication in massive multiple-input multiple-output (MIMO) systems. In this letter, we present a novel residual convolutional neural network (CNN) with recurrent neural network (RNN) called ResR model to tackle the carrier frequency offset (CFO) problem in multi-user massive MIMO with one-bit analog-to-digital converters (ADCs). Leveraging the combined strengths of residual CNN and RNN, the ResR model can extract frequency-spatial characteristics of all users for CFO estimation. Moreover, it effectively addresses the vanishing gradient problem in CNN-based model while delivering superior accuracy with fewer parameters compared to exiting CNN or RNN models. Through extensive experimental evaluations, we consistently demonstrate the efficiency and robustness of the ResR model in multi-user CFO estimation for one-bit ADCs massive MIMO.

Index Terms—carrier frequency offset (CFO), deep learning, multiple-input multiple-output (MIMO), one-bit analog-to-digital converter (ADCs), residual network.

I. INTRODUCTION

With the rapid advancement of communication technology, massive multi-input multi-output (MIMO) systems, which can focus transmitted signal energy within limited ranges from the base station (BS), play a pivotal role and are currently undergoing significant development. This concentrated energy allocation yields substantial improvements in terms of overall system capacity and can simultaneously serve multiple users on the same time-frequency resources [1]. In the conventional massive MIMO systems, each radio frequency (RF) port is equipped with two high-resolution analog-to-digital converters (ADCs), where the cost and power consumption of the ADCs increases significantly with the number of quantization bits. As a result, the low cost and power consumption make multi-user massive MIMO with one-bit ADCs a promising technique for the next generation wireless systems [2].

However, the remarkable improvements of multi-user massive MIMO with one-bit ADCs in both spectral efficiency and power consumption depends on perfect frequency synchronization between the transceivers [3]. Since the joint recovery of multiple CFOs leads to a highly non-linear estimation problem, the multi-user frequency synchronization in the massive MIMO systems is quite different from the conventional single-CFO estimation [4]–[6]. In the past decade, there have been a few works on the CFO estimation for the conventional multi-user MIMO systems [7], [8]. For example, Zhang *et al.* in [8] proposed a joint spatial-frequency alignment procedure (JSFA) for the frequency synchronization in the massive MIMO with uniform linear array (ULA) at the BS. However, conventional methods designed for the high-resolution massive MIMO systems may suffer from substantial estimation performance loss in the one-bit ADCs due to the severe non-linear distortion.

Recently, with the potential to capture non-linear feature interaction, deep learning (DL) has emerged as a powerful tool for addressing complex tasks in the field of communications [9], [10]. Compared to the conventional methods, DL models excels at extracting relevant features from the received signal with reduced manual feature engineering. The authors in [12] utilized the convolutional neural network (CNN) and its derivatives to design a robust CFO estimator for single-user MIMO with one-bit ADCs. However, the traditional back propagation method may encounter challenges and gives rise to a vanishing gradient problem after extensive training due to the large number of antennas in massive MIMO systems. In [13], a hybrid CNN-attention-deep neural network (DNN) architecture was proposed for CFO estimation in the single-input single-output (SISO) orthogonal frequency division multiplexing (OFDM) systems, where the residual CNN module was employed to abstract the necessary features from the received OFDM signal. The residual CNN module can address the vanishing gradient problem in CNN caused by the escalation of convolutional layers. However, the above works only considered the conventional single-user CFO estimation and may not be applied in the multi-user massive MIMO systems.

In this letter, a residual CNN with recurrent neural network (RNN) called ResR model is designed to handle the frequency synchronization problem in multi-user massive MIMO with one-bit ADCs. The proposed ResR model outperforms both exiting CNN and RNN models in terms of achieving higher accuracy while using fewer parameters. For the first time, both the frequency and spatial characteristics are extracted in the proposed model for multi-user CFO estimation. By combining the residual CNN and RNN, the vanishing gradient problem has been avoided to a certain extent, resulting in better estimation accuracy. We further develop a structure that effectively utilizes the spatial characteristics extracted from the residual CNN to facilitate the user separation. The numerical results demonstrate the effectiveness and robustness of the proposed model and their superiority over existing approaches in the one-bit ADCs massive MIMO.

Y. Feng, Z. Zhou, Huimei Han and W. Lu are with College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, P. R. China (e-mail: yqfeng@zjut.edu.cn, rr3855719hanx@163.com, hmhan1215@zjut.edu.cn, luweid@zjut.edu.cn).

Jie Tang is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: eejtang@scut.edu.cn).

Nan Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, P. R. China (e-mail: zhaonan@dlut.edu.cn).

Arumugam Nallanathan is with School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS, U.K. (e-mail: a.nallanathan@qmul.ac.uk).

II. System Model

The multi-user CFO estimation problem in the uplink massive MIMO systems with one-bit ADCs is considered in this work. In the massive MIMO systems, K single-antenna users are randomly distributed in the coverage area and communicating with the BS simultaneously. The BS is equipped with $M \gg 1$ antennas in the form of ULA and usually elevated at a very high altitude such that there are few surrounding scatterers at the end of the BS. Hence, the spatial channel model in [15] that exploits the array manifold and angle information of propagation signals is adopted in this work. Similarly, we assume that the channel is flat fading and time-invariant as in [15], while the proposed model can be extended to the frequency-selective scenario.

The propagation channel $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ between the BS and the kth user is assumed to be composed of P separable rays and can be given as:

$$\mathbf{h}_k = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \alpha_{p,k} \boldsymbol{a}(\theta_{p,k}), \qquad (1)$$

where $\alpha_{p,k} \sim \mathcal{CN}(0, \sigma_{h,p}^2)$ denotes the complex gain in the pth ray of the kth user with mean zero and variance $\sigma_{h,p}^2$. Moreover, the array manifold vector $\boldsymbol{a}(\theta_{l,k}) \in \mathbb{C}^{M \times 1}$ can be expressed as:

$$\boldsymbol{a}(\theta_{p,k}) = \left[1, e^{-\mathbf{j}\frac{2\pi d}{\lambda}\cos\theta_{p,k}}, \ldots, e^{-\mathbf{j}\frac{2\pi(M-1)d}{\lambda}\cos\theta_{p,k}} \right]^T, \quad (2)$$

where λ is the signal carrier wavelength, d is the antenna spacing and $\theta_{p,k}$ represents the angle-of-arrival (AoA). Similar to [15], we assume that the incident angles of the kth user are limited within a narrow angular range $\theta_{p,k} \in (\theta_k - \theta_{as}, \theta_k + \theta_{as})$, where θ_k and θ_{as} represent the mean AoA of the kth user and angular spread (AS), respectively. Then, the channel matrix of all users can be expressed by $\mathbf{H} \in \mathbb{C}^{M \times K}$ as:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_K]. \qquad (3)$$

Let us consider the classic problem of CFO estimation of a discrete-time set of samples. Denote $x_{k,n}, n = 0, 1, \cdots, N-1$ as the pilot symbol of the kth user with N being the length of pilot symbol. Let Δf_k and T_s stand for the real CFO of the kth user and the sampling interval, respectively. The CFO-induced phase shift over one symbol can be represented as $2\pi\omega_k = 2\pi N T_s \Delta f_k$. Denote $\mathbf{x}_k = [\bar{x}_{k,0}, \bar{x}_{k,1}, ..., \bar{x}_{k,N-1}]^T$, where $\bar{x}_{k,n} = x_{k,n} e^{\mathbf{j}(\frac{2\pi\omega_k n}{N} + \phi_{k,n})}$ with $\phi_{k,n}$ being the phase shift. After passing the channel, the received signal after one-bit quantization is expressed as

$$\mathbf{Y} = \text{sign}(\mathbf{HX} + \mathbf{N}), \qquad (4)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K]^T$ and $\mathbf{N} \in \mathbb{C}^{M \times N}$ denotes the corresponding additive white Gaussian noise (AWGN) matrix with $E[\mathbf{NN}^H] = N\sigma_n^2 \mathbf{I}_M$. In the one-bit ADCs massive MIMO, the signum function $\text{sign}(\cdot)$ is adopted as the element-wise operator applied separately to the real and imaginary parts of the received signal and is given by

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ -1, & \text{otherwise} \end{cases}. \qquad (5)$$

Here, the quantized received signal $\mathbf{Y}$ is considered as the input of proposed ResR model for multi-user CFO estimation, which is rather challenging due to the coexistence of different CFOs $\boldsymbol{\omega} = [\omega_1, \omega_2, ..., \omega_K]^T$ in the received signal.

III. Proposed ResR Model

In this section, we present the ResR method for multi-user CFO estimation in one-bit ADCs massive MIMO systems. As illustrated in Fig. 1, the ResR comprises two key components: the multi-user separation (MUS) module and the multi-CFO estimation (MCE) module. The MUS module, comprising a residual stack and a specially designed convolution layer for feature combining (FC), is intended to extract the frequency and spatial characteristics for multi-user separation with significantly reduced number of parameters. The MCE module is designed to jointly estimate multi-user CFOs from a set of approximate single-user models, leveraging the long short-term memory(LSTM)-gated recurrent units (GRU) to address the vanishing gradient problem during training and to reduce the computational complexity of the entire method.

A. Model Structure of MUS Module

The MUS module consists of two main components: feature extraction (FE) module and FC module. In the network architecture design, CNN has been chosen due to its advantage of effectively extracting fundamental features from localized repetitive fields [12]. The multi-user CFO estimation of one-bit ADCs massive MIMO can be considered as an image-to-image translation as shown in [11]. Hence, the quantized received signal can be conceptualized as two low-resolution images with the dimension of $M \times N$, and represented as:

$$\mathbf{Y}_1 = [\Re(\mathbf{Y}), \Im(\mathbf{Y})], \qquad (6)$$

where the real part and imaginary part of quantized received signal $\mathbf{Y}$ are separated and stacked together, resulting in the final input $\mathbf{Y}_1 \in \mathbb{C}^{M \times N \times 2}$. As shown in Fig. 1, the network architecture of FE module consists of the convolutional layers and a residual stack. To extract the frequency and spatial characteristics of received signal, the convolution layer performs convolution operation on $\mathbf{Y}_1$ and abstracts the features

$$\mathcal{F}(\mathbf{Y}_1) = \text{SeLU}(Conv(\mathbf{Y}_1)) = \text{SeLU}(\mathbf{Y}_1 * \boldsymbol{\Phi} + \mathbf{b}_1), \quad (7)$$

where $\boldsymbol{\Phi}$ represents the weights of the convolution kernel and $\mathbf{b}_1$ denotes the bias. Here, the scaled exponential linear unit (SeLU) is employed and $*$ is the convolutional operator. Then, the abstracted feature are passed down to the residual stack with three stages. In the first stage, the output and the input of convolution layer is combined together as the

$$\mathbf{U}_2 = \mathcal{F}(\mathbf{U}_1) + \mathbf{U}_1, \qquad (8)$$

where $\mathbf{U}_1$ is the input of the first stage. Then, $\mathbf{U}_2$ serves as the input of the second stage with two convolution layers. Similarly, the output of the second stage is combined with the input $\mathbf{U}_2$, i.e., $\mathbf{U}_3 = \mathcal{F}(\mathcal{F}(\mathbf{U}_2)) + \mathbf{U}_2$. Then, the procedure is continued in the third stage with three convolution layers and the final output of residual stack can be expressed as

$$\mathbf{U}_4 = \mathcal{F}(\mathcal{F}(\mathcal{F}(\mathbf{U}_3))) + \mathbf{U}_3 + \mathbf{U}_1, \qquad (9)$$
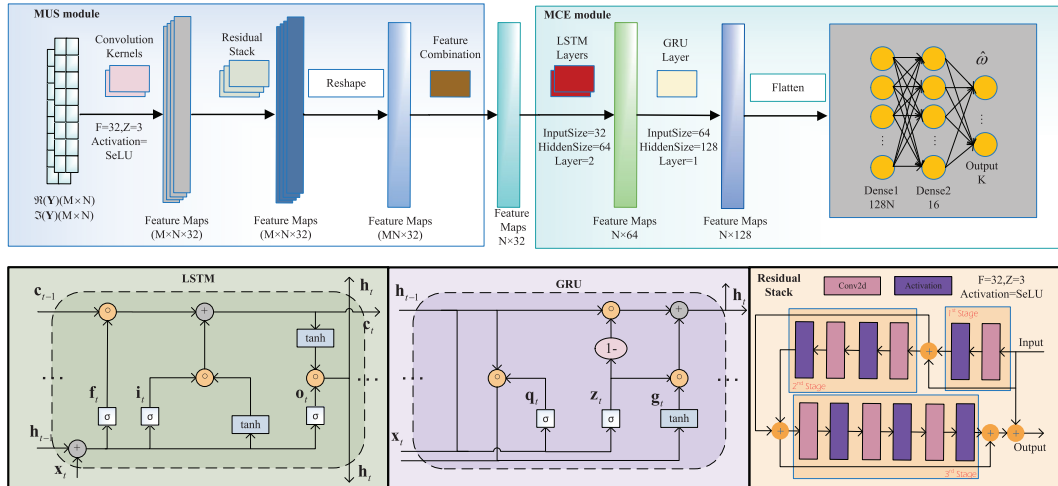
Fig. 1: ResR architecture for multi-user CFO estimation in one-bit ADCs massive MIMO.

where the input of residual stack is combined together in the final stage. The purpose of incorporating the residual CNN structures is to address the vanishing gradient problem, which is a common challenge in deep convolutional layers. Moreover, these structures can improve the estimation accuracy with the increased depth of the convolutional layers.

Finally, the FC module is proposed to facilitate the user separation in the angle domain and reshape the output of FE module into the sequences that suitable for the subsequent module. First, the FC module transforms the feature map $\mathbf{U}_4$ from its original shape of $M \times N$ in each channel into a one-dimensional (1D) representation, denoted as $\mathbf{u}_f = [u_0^{(f)}, u_1^{(f)}, ..., u_{MN-1}^{(f)}]^T \in \mathbb{C}^{MN \times 1}, f = 1, 2, \cdots, F$. Here, F is the number of convolutional kernels. Subsequently, $\mathbf{u}_f$ is fed into a 1D convolutional layer with a kernel size of M and a stride of M. Moreover, a SeLU activation function is applied to the spatial characteristics of the received signal, resulting in a combination process that partitions the signal into distinct user components. This process can be expressed as follows:

$$s_n = \text{SeLU}(\sum_{m=0}^{M} w_m u_{nM+m}^{(f)} + b_n), \qquad (10)$$

where w_m and b_n denote the weight and bias of the SeLU layer, respectively. Then, the feature map $\mathbf{U}_4$ is converted into a 2D matrix of size $N \times F$ and then broken into F sequences.

B. Model Structure of MCE Module

We propose to employ the LSTM-GRU-DNN structure in the MCE module to precisely estimate the multi-user CFOs, which contains two LSTM layers, one GRU layer and one DNN layer.

In the massive MIMO systems, the number of sequences after FC module is enormous, which causes a vanishing gradient and leads to the training failure. The LSTM can effectively process MIMO data and overcome the vanishing gradient problem, ensuring superior estimation performance. Two LSTM layers are utilized to effectively denoise the input

sequences with the memory cells and three gates. The input gate $\mathbf{i}_t$ stores information to the cell state and the forget gate $\mathbf{f}_t$ determines how much information should be removed from the cell state. Finally, the output gate $\mathbf{o}_t$ selects valuable information from the current cell state and presents it as an output. The hidden state of tth cell $\boldsymbol{h}_t$ can be computed as:

$$\begin{aligned}
\mathbf{i}_t &= \sigma\left(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i\right), \\
\mathbf{f}_t &= \sigma\left(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f\right), \\
\mathbf{c}_t &= \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \circ \tanh\left(\mathbf{W}_{xc}\mathbf{x_t} + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c\right), \\
\mathbf{o}_t &= \sigma\left(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o\right), \\
\mathbf{h}_t &= \mathbf{o}_t \circ \tanh\left(\mathbf{c}_t\right), \qquad (11)
\end{aligned}$$

where $\mathbf{x}_t$ represents the input of the cell and $\mathbf{c}_t$ denotes the cell state. Here, $\mathbf{W}$ and $\mathbf{b}$ are the learnable network parameters in the LSTM, respectively. $\sigma(x) = \frac{1}{1+e^{-x}}$ and $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ are nonlinear activation functions, and the symbol $\circ$ denotes the Hadamard product.

Meanwhile, the convergence rate of MCE module is accelerated by GRU, which is another variant of the basic RNN and computationally efficient by combining the memory cell and the forget gate into a single update gate. Moreover, the GRU layer can also work as a compensator that retrieves the lost information resulting from the denoising process.

$$\begin{aligned}
\mathbf{q}_t &= \sigma\left(\mathbf{W}_{xq}\mathbf{x}_t + \mathbf{W}_{hq}\mathbf{h}_{t-1} + \mathbf{b}_q\right), \\
\mathbf{z}_t &= \sigma\left(\mathbf{W}_{xz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z\right), \\
\mathbf{g}_t &= \tanh\left(\mathbf{W}_{xg}\mathbf{x}_t + \mathbf{q}_t \circ \left(\mathbf{W}_{hg}\mathbf{h}_{t-1}\right) + \mathbf{b}_g\right), \\
\mathbf{h}_t &= \left(\mathbf{1} - \mathbf{z}_t\right) \circ \mathbf{h}_{t-1} + \mathbf{z}_t \circ \mathbf{g}(t),
\end{aligned} \qquad (12)$$

where the reset gate $\mathbf{q}_t$ determines which information from the previous state $\mathbf{h}_{t-1}$ should be discarded or forgotten, and the update gate $\mathbf{z}_t$ selects new information from both the input vector and the previous state $\mathbf{h}_{t-1}$ to be integrated into a new state. The candidate for the future hidden state is determined by the candidate state gate $\mathbf{g}_t$. Finally, the output of GRU is further processed by the subsequent fully connected layer, where DNN is utilized to establish the relationship between the inputs and target values to estimate the multi-user CFOs $\hat{\boldsymbol{\omega}} = [\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_K]^T$.

TABLE I: SETTINGS FOR DATA GENERATION

Parameter	Value
CFO ω_k	$[0.2, 0.25]$
Phase shift $\phi_{k,n}$	$[0, 2\pi]$
Number of channel rays P	2
d/λ	0.5
Pilot lengths N	$\{8, 16, 32\}$
Numbers of antennas M	$\{8, 16, 32\}$
Numbers of users K	$\{2, 4, 6\}$
SNR	$[-20, 40]$

IV. SIMULATION RESULTS

In this section, we outline the process of generating the training data and configuring the parameters for the proposed ResR method. Then, we evaluate the CFO estimation performance of proposed ResR under various conditions by comparing it with the conventional two stage approach (TSA) in [6], the JSFA in [8], the exiting CNN-based and RNN-based methods in [12] and the ResD (Residual CNN with DNN).

A. Data Generation

In the simulation, the training datasets $\mathcal{D} = \{\boldsymbol{Y}_{1,i}, \boldsymbol{\omega}_i\}_{i=1}^{N_{tr}}$ are generate based on (4), where the parameter $\boldsymbol{\omega}_i$ denotes the true CFO of each user which is utilized as the label during training process and N_{tr} is the total number of training samples. We assume that the CFO ω_k is randomly generated from 0.2 to 0.25, and the phase shift $\phi_{k,n}$ follows the uniform distribution within the range of $[0, 2\pi]$. The number of channel ray is set as $P = 2$ and $d = \lambda/2$. The channel vector of different users are formulated according to (1). The average power of pilot symbol is normalized to 1, and the signal-to-noise ratios (SNR) is defined as $1/\sigma_n^2$. The angular spread $\theta_{as} = 10°$ is considered. We introduce the noise with different SNR ranging from -20dB to 40dB. Moreover, we consider various configurations with different qualities: the number of pilot symbols, the number of antennas, the number of users and quantized/unquantized training data as shown in Table I.

B. Training Model

For the MUS module, two convolutional layers is utilized with the kernel size $Z = 3$ and the number of convolutional kernels $F = 32$. Three blocks are considered in the residual stack of MUS module to avoid the vanishing gradient problem and increase the estimation performance with 32 output channel. The network is trained by the Adam optimizer and learning rate decay, employing a batch size of $N_b = 32$ instances in each iteration and an initial learning rate of 0.001. The network parameters are updated by minimizing the mean square error (MSE) loss function.

For the MCE module, the objective is to minimize the loss function, which is the discrepancy between the output of proposed model $\hat{\boldsymbol{\omega}}$ and the correct label $\boldsymbol{\omega}$. The regularization term is added to avoid overfitting. The loss function for our proposed model can be formulated as follows:

$$\text{Loss} = \frac{1}{N_b} \sum_{i=1}^{N_b} \|\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}_i\|^2 + \lambda \|\boldsymbol{\beta}\|^2, \tag{13}$$

where λ is set as 5×10^{-8} and $\boldsymbol{\beta}$ is the parameters of network with $\|\cdot\|$ being the Euclidean norm.

C. Performance Comparations

In Fig. 2, we compare the proposed ResR with the exiting methods in the scenario of two users with $M = 32$ and $N = 32$. The mean AoA of two users are fixed as $\{50°, 70°\}$. It is evident that the ResR significantly outperforms the existing methods due to the superior architecture that combines the residual CNN and RNN together. Moreover, the performances of RNN and TSA are poor since it cannot discriminate the spatial characteristics between different users. Similarly, Fig.2 depicts that the proposed method achieves performance gain as compared to the JSFA, which depends heavily on the massive ULA. Furthermore, the ResR surpasses the CNN and ResD since the LSTM and GRU layers in the ResR can effectively capture long-term dependencies in sequential data.
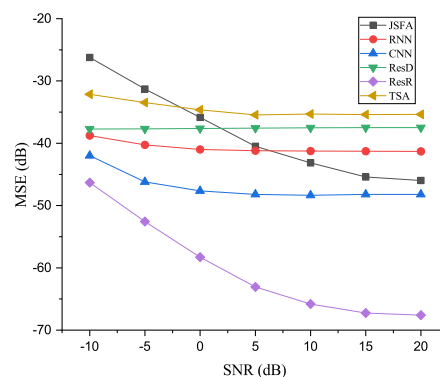
Fig. 2: MSE performance comparison of different methods ($K = 2$).

In Fig. 3, we present the MSE performance of our proposed ResR under different scenarios for both $K = 2$ and $K = 6$. Three scenarios are defined by varying angular distances θ_{ad} between the adjacent users as in Table II. It can be found out that as the angular distance θ_{ad} decreases, the overlapping region θ_{ol} expands from non-overlapping to severe overlapping among users, resulting in similar spatial characteristics. For $K = 2$, the mean AoAs are distributed as $\{(70 - \theta_{ad})°, 70°\}$. For $K = 6$, the mean AoAs are fixed as $\{(70 - \theta_{ad})°, 70°, 90°, \underline{(90 + \theta_{ad})°},$ $\underline{(110 + \theta_{ad})°}, (110 + 2\theta_{ad})°\}$, where the spatially proximate users are underlined. It can be observed from Fig. 3 that the ResR exhibits remarkable stability and effectively handles scenarios with severe overlapping, including the scenario with a 90% overlapping. Moreover, the MSE performance decreases as the number of users K increases since the dataset becomes enriched with more similar spatial characteristics in the training and testing process.

TABLE II: DIFFERENT SCENARIOS FOR AoA-OVERLAPPING

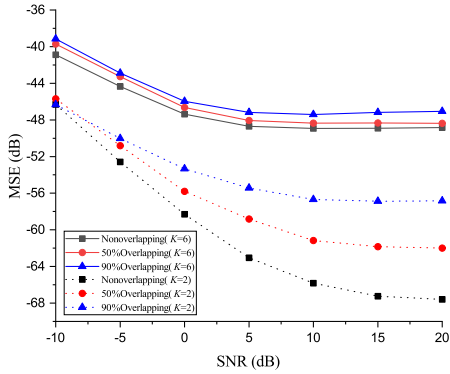| Scenario	Angular Distance	Overlapping
Nonoverlapping	20	0 $(0\%\theta_{as})$
50%overlapping	15	5 $(50\%\theta_{as})$
90%overlapping	11	9 $(90\%\theta_{as})$

Fig. 3: MSE performance of the ResR under different overlapping scenarios.

TABLE III: COMPARATIVE COMPLEXITY ANALYSIS

Method	Parameters	Flop counts
ResR	342,750	58,607,248
CNN	398,861,984	595,778,400
RNN	16,771,333	23,918,688
ResD	33,558,575	25,909,968

In Fig. 4, we evaluate the impact of both quantization noise and the number of antennas on the CFO estimation performance of the proposed ResR. The performance of the high-resolution case has be provided as the upper bound. In this example, the mean AoA of six users are fixed as $\{50°, 70°, 90°, 110°, 130°, 150°\}$. Both one-bit and ∞-bit training models are included with $M = [16, 32, 64]$. The MSE results clearly demonstrate that the ResR performs effectively under both high-resolution and low-resolution conditions. Moreover, we can also observe that the MSE performance improve as the number of antennas M increases since adding more antennas can provide higher spatial resolution.
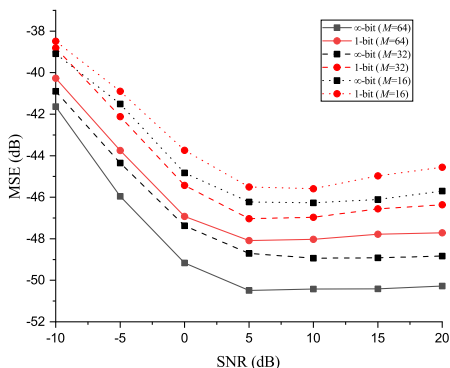

Fig. 4: MSE performance of the ResR with one-bit/∞-bit training models.

The required complexities of different methods are provided in Table III. Notably, the proposed ResR can achieve much better estimation performance with a lower order of computational burden than the existing CNN. Though the RNN and ResD have smaller flop counts, the proposed ResR achieves much better performance with the MUS module.

V. Conclusions

In this letter, we design a novel residual CNN with RNN (ResR) model to address the multi-user CFO estimation problem in one-bit ADCs massive MIMO. By leveraging the advantages of both residual CNN and RNN, the ResR method not only overcomes the vanishing gradient problem but also achieves superior estimation accuracy, surpassing the performance of standalone CNN and RNN methods. The MUS module can effectively extract both the frequency and spatial characteristics from highly quantized received signals for CFO estimation. Moreover, the combination of LSTM and GRU can addresses the vanishing gradient problem and accelerate the convergence rate, resulting in superior performance in both denoising and CFO estimation. The simulation results demonstrated the effectiveness and robustness of the ResR over exiting methods.

References

[1] E. Larsson, O. Edfors, F. Tufvesson, and T.Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[3] M. Morelli and M. Moretti, "Carrier frequency offset estimation for OFDM direct-conversion receivers," *IEEE Trans. Wirel. Commun,* vol. 11, no. 7, pp. 2670–2679, Jul. 2012.

[4] N. J. Myers and R. W. Heath, "Message passing-based joint CFO and channel estimation in mmWave systems with one-bit ADCs," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 6, pp. 3064–3077, Jun. 2019.

[5] D. Zhu, R. Bendlin, S. Akoum, A. Ghosh, and R. W. Heath, "Double sequence frequency synchronization for wideband millimeter-wave systems with few-bit ADCs," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 2, pp. 1357–1372, Nov. 2019.

[6] J. Zhu, H. Cao, Z. Xu, "A two-stage approach to estimate CFO and channel with one-bit ADCs," *Signal Process.* vol. 168, no. 7, pp. 107336, Mar. 2020.

[7] Y. R. Tasi, H. Y. Huang, Y. C. Chen, and K. J. Yang, "Simultaneous multiple carrier frequency offset estimation for coordinated multi-point transmission in OFDM systems," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4558–4568, Sept. 2013.

[8] W. Zhang, F. Gao, S. Jin, and H. Lin, "Frequency synchronization for uplink massive MIMO Systems," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 15, pp. 235–249, Jan. 2018.

[9] M. Zhou, X. Huang, Z. Feng, and Y. Liu, "Coarse frequency offset estimation in MIMO systems using neural networks: A solution with higher compatibility," *IEEE Access*, vol. 7, pp. 121 565–121 573, 2019.

[10] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive MIMO systems," *IEEE Wirel. Commun. Lett.*, vol. 7, no. 5, pp. 852–855, 2018.

[11] I. Helmy, P. Tarafder and W. Choi, "LSTM-GRU model-based channel prediction for one-bit massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 11053-11057, Aug. 2023.

[12] R. M. Dreifuerst, R. W. Heath, M. N. Kulkarni and J. Charlie, "Deep learning-based carrier frequency offset estimation with one-bit ADCs," *IEEE SPAWC.*, Atlanta, GA, USA, pp. 1-5, 2020.

[13] M. S. Chaudhari, S. Majhi and S. Jain, "CNN-attention-DNN design for CFO estimation of non-pilot-assisted OFDM system," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 551-555, Feb. 2023.

[14] Y. Feng, W. Zhang and Y. Ge, "Frequency Synchronization With Beamforming Network Optimization for Uplink Massive MIMO Systems," *IEEE Trans. Veh. Technol.*,vol. 69, no. 3, pp. 3486–3490, Mar. 2020.

[15] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.