The following publication S. Singh, A. Trivedi and D. Saxena, "Channel Estimation in IRS-Assisted OTFS Communication via Residual Attention Network," 2023 IEEE 7th Conference on Information and Communication Technology (CICT), Jabalpur, India, 2023, pp. 1-5 is available at https://doi.org/10.1109/CICT59886.2023.10455192.

Channel Estimation in IRS-Assisted OTFS Communication via Residual Attention Network

Shatakshi Singh
Department of IT
ABV-IIITM
Gwalior, India
shatakshi@iiitm.ac.in

Aditya Trivedi
Department of IT
ABV-IIITM
Gwalior, India
atrivedi@iiitm.ac.in

Divya Saxena
Department of Computing
The Hong Kong Polytechnic University
Kowloon, Hong Kong
divsaxen@comp.polyu.edu.hk

***Abstract*—For intelligent reflecting surface (IRS) based communication, channel estimation methods have predominantly focused on low-mobility and static scenarios. However, in dynamic scenarios where mobility and channel variations take place, accurate channel estimation becomes a challenging task. To address this limitation, this paper proposes a novel approach for channel estimation in dynamic IRS-aided communication scenarios by leveraging the advantages of orthogonal time-frequency space (OTFS) modulation. The proposed approach converts the time-frequency domain channel representation into the delay-Doppler (DD) domain using OTFS modulation. By doing so, the channel estimation problem is transformed into estimating the DD channel, which is more suitable for dynamic scenarios. To estimate the DD channel, a residual attention-based channel estimation (RACE) model is proposed. The RACE model outperforms existing deep learning methods and conventional approaches. It achieves a lower normalized mean square error compared to other methods.**

***Index Terms*—Residual attention channel estimation (RACE), intelligent reflecting surface (IRS), orthogonal time-frequency space (OTFS).**

I. Introduction

Intelligent reflecting surface (IRS) has gained significant attention for improving spectrum and energy efficiency in wireless communication. IRS is a planar metasurface equipped with cost-effective reflective elements that can dynamically modulate the phase and amplitude of incoming signals [1]. By optimizing the reflection coefficients of IRS, it becomes possible to steer the incident signal toward the desired direction, resulting in enhanced received signal power at the destination. In contrast to conventional active relays, IRS presents notable benefits, including minimized power consumption and reduced hardware expenses, achieved through the passive reflection of signals [2]. Due to these benefits, IRS has attracted significant attention in various wireless systems, including MIMO, cognitive radio, NOMA, and OFDM.

The precise and real-time channel state information (CSI) is essential for the effective design of IRS reflection [3]. Due to the numerous reflecting components and limited signal processing abilities, collecting this information is extremely difficult. Several successful channel estimation strategies have been put forth in the literature to address these problems. For estimating the cascaded IRS channel, the least squares (LS)-based estimator was presented in [4]. High pilot overhead, which worsens with the quantity of reflecting parts, is the main drawback of this strategy. To solve this problem, [5] suggested a method that clusters nearby elements and gives them the same reflection pattern, hence lowering pilot overhead.

To immediately estimate spatial angle information at both the base station (BS) and IRS, [6] and [7] created a one-stage channel prediction method for IRS-assisted millimeter wave (mmWave) systems. In addition, [8], [9] developed a two-stage cascaded channel estimation method where the BS angle information is calculated in the first stage and the cascaded channel coefficients and IRS angle information are estimated in the second stage.

IRS research has largely concentrated on static and low-mobility situations with quasi-static channels, which limits its applicability in real-world scenarios involving user movements, like vehicle-to-vehicle communication (V2V) and high-speed trains (HST). Due to a decreased channel coherence time and changes in the scattering environment, traditional static channel designs perform poorly in these dynamic situations.

A new paradigm for communication, orthogonal time frequency space (OTFS) modulation has been presented as a solution to this problem [10]. OTFS multiplexes symbols in the delay-Doppler (DD) domain rather than the conventional time-frequency (TF) domain. This imparts remarkable resilience to Doppler shifts and delays spread, rendering it suited for highly dynamic environments. Unlike traditional modulation schemes where the channel gain may vary across different frequency and time bins, OTFS spreads the signal energy across the DD domain uniformly. This uniform spreading ensures that each multipath component receives equal energy, resulting in equal channel gains. As a result, the channel estimation task becomes easier because there is no need to estimate and compensate for varying channel gains across different subcarriers or time slots. The combination of IRS and OTFS presents a promising opportunity to leverage the advantages of both flexible channel configurations provided by the IRS and the resilience of OTFS in high-mobility communications.

In the study presented by authors of [11], a pilot transmission approach was employed where an entire OTFS frame was dedicated solely to pilot transmission. The estimated channel information obtained from this pilot transmission was subsequently utilized for data detection in the following frame. However, this method faces limitations if the channel estimation becomes outdated in the subsequent frame. To address this limitation, an OTFS channel prediction scheme for a single-input single-output (SISO) system is proposed in [12]. In this scheme, within each OTFS frame, a single pilot symbol is embedded along with guard symbols and data symbols to minimize interference and predict the channel using a threshold-based method.

The research presented in [13] introduces structured orthogonal matching pursuit (OMP) for three-dimensional space. This approach effectively leverages the block sparsity observed in the Doppler domain, the burst sparsity in the angle domain, and the regular sparsity in the delay domain. By considering the unique sparsity characteristics across these three dimensions, the proposed algorithm offers promising results for channel estimation in complex scenarios. Different from the model-based approaches, authors in [14] proposed a data-driven method for estimating channels in the IRS-OTFS system. The channel estimation is framed as a denoising problem and a residual neural network (ResNet) is applied to estimate the channel. ResNet has shown remarkable performance in various tasks, including image classification and denoising. However, when applied directly to channel estimation, ResNet may not fully capture the specific characteristics of the channel and could be limited in their generalizability.

To address this limitation and enhance the accuracy of channel estimation, we propose the residual attention-based channel estimation (RACE) model. The RACE model introduces attention mechanisms after the residual network, enabling the model to selectively focus on important channel features while suppressing noise. The RACE model consists of residual blocks with an attention module. The attention module enables the network to learn to focus on informative features while attenuating the impact of noise. By adaptively weighting the feature maps, the attention mechanism incorporated in the RACE model helps to improve the denoising performance. The main contributions of this work are listed as follows:

1) We initially estimate the DD channel with a conventional estimation method [15]. Further to improve the estimation accuracy and to eliminate noise in the estimated channel, a residual attention-based model is proposed.
2) To evaluate the performance of the proposed RACE model for channel estimation, we conducted experiments using the normalized mean square error (NMSE) as the performance metric. We investigated the impact of varying Signal-to-Noise Ratio (SNR), the number of IRS elements, and the number of DD paths on the model's performance.

The subsequent sections of the paper are structured as follows: Section II provides a description of the system model The proposed channel estimation method is explained in Section III. Experimental settings and results are presented in Section IV. The conclusion of the paper can be found in Section V.

Notations: Real numbers and complex numbers are denoted with the notations $\mathbb{R}$ and $\mathbb{C}$, respectively. The superscripts

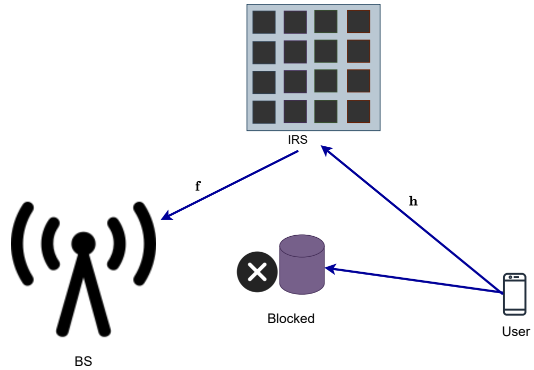

Fig. 1. IRS-aided Wireless Network

T and H indicate the transpose and conjugate transpose operations, respectively. The vector and matrix are represented by bold symbols, namely $\mathbf{v}$ for a vector and $\mathbf{V}$ for a matrix. Moreover, the Frobenius norm of a matrix is expressed as $|\cdot|_F$.

II. System Model

We consider a scenario consisting of an IRS with I passive elements, BS, and a user with a single antenna, where a direct link between the user and the BS is blocked, as depicted in Fig. 1. The DD channel from BS-IRS and IRS-user are given as: $f(\tau_i,\rho_i) = \sum_{q=1}^{Q_f} f_{i,q}\delta(\tau_i^f - \tau_{i,q}^f)\delta(\rho_i^f - \rho_{i,q}^f)$ and $h(\tau_i,\rho_i) = \sum_{q=1}^{Q_h} h_{i,q}\delta(\tau_i^h - \tau_{i,q}^h)\delta(\rho_i^h - \rho_{i,q}^h))$, respectively. A time division duplexing (TDD) protocol is considered, where the channel can be estimated in the uplink, and channel reciprocity is assumed.

The incident signal on the i-th element of the IRS from the user is given as:

$$r^i(t) = \sum_{q=1}^{Q_h} h_{i,q} e^{j2\pi\rho_{i,q}^h (t-\tau_{i,q}^h)} s(t-\tau_{i,q}^h), \tag{1}$$

where $h_{i,q}$, $\rho_{i,q}^h$, and $\tau_{i.q}^h$, are the coefficients of the channel, the Doppler, and delay of the q^{th} path of the user to the i^{th} IRS element. s(t) is the time-domain signal obtained by performing Heisenberg's Transform [16].

The signal received at the BS after reflection from the i^{th} IRS with $\beta_i e^{j\psi_i}$ being the reflection coefficient is represented as:

$$\begin{aligned} r_{out}^i = \sum_{p=1}^{Q_f} f_{i,p}\beta_i e^{j\psi_i} \sum_{q=1}^{Q_h} e^{j2\pi\rho_{i,p}^f (t-\tau_{i,p}^h - \tau_{i,p}^f)} h_{i,q} \\ e^{j2\pi\rho_{i,q}^h (t-\tau_{i,q}^h - \tau_{i,p}^f)} s(t-\tau_{i,p}^h - \tau_{i,q}^f), \end{aligned} \tag{2}$$

where $f_{i,p}$, $\rho_{i,p}^f$, and $\rho_{i,p}^f$, are the channel coefficient, the Doppler, and delay value of the i^{th} IRS element to the BS for p^{th} path. The signal received at the BS from the user via I elements of the IRS can be obtained as:

$$r_{out}(t) = \sum_{i=1}^{I} \beta_i e^{j\psi_i} \sum_{q=1}^{Q_f}\sum_{p=1}^{Q_h} g_{i,qp} e^{j2\pi\rho_{i,qp}(t-\tau_{i,qp})} s(t-\tau_{i,qp}), \tag{3}$$

where $\rho_{i,qp} = (\rho^f_{i,p} + \rho^h_{i,q})$ is the Doppler shift, $\tau_{i,qp} = (\tau^f_{i,p} + \tau^h_{i,q})$ is the delay of the cascaded channel, and $g_{i,qp} = f_{i,p} h_{i,q}$ is the effective channel gain of the cascaded DD channel. The discrete-time representation of (3) is obtained as:

$$r[n] = = \sum_{i=1}^{I} \beta_i e^{j\psi_i} \sum_{p=1}^{Q_f} \sum_{q=1}^{Q_h} g_{i,qp} e^{j2\pi \frac{k_{i,qp}(n-l_{i,qp})}{NM}} s[[n - l_{i,qp}]_{NM}], \tag{4}$$

where n = 1, ..., NM. In the matrix form, (4) can be expressed as:

$$\mathbf{r} = \sum_{i=1}^{I} \beta_i e^{j\psi_i} \tilde{\mathbf{G}}_i \mathbf{s} + \mathbf{z} = \tilde{\mathbf{G}} \mathbf{s} + \mathbf{w}, \tag{5}$$

where $\mathbf{w} \in \mathbb{C}^{NM\times 1}$ is the additive white Gaussian noise (AWGN) vector. The transmitted data vector from the user and the received data vector at the BS are given as $\mathbf{s} \in \mathbb{C}^{NM\times 1}$ and $\mathbf{r} \in \mathbb{C}^{NM\times 1}$. When the ideal-pulse shaping is employed in OTFS, and after applying the ISFFT/SFFT and Heisenberg/Wigner transforms, equation (12) can be restated as follows:

$$\mathbf{y} = \sum_{i=1}^{I} \beta_i e^{j\psi_i} \mathbf{G}_i \mathbf{x} + \mathbf{w}. \tag{6}$$

The received symbol in the $[u, v]$ bin is expressed as:

$$y[u,v] = \left(\sum_{u'=1}^{M} \sum_{v'=1}^{N} \sum_{i=1}^{I} \beta_i e^{j\psi_q} g_i[u',v'] \right) x[[u-u']_M, [u-v']_N]. \tag{7}$$

After limiting the (7) to maximum delay and the Doppler tap present in the channel, we can obtain the relation as:

$$y[u,v] = \sum_{v'=-v_\rho}^{v_\rho} \sum_{u'=0}^{u_\tau} g_{\text{eff}}[u',v'] x[[u-u']_M, [v-v']_N] + z[u,v]. \tag{8}$$

$g_{eff}[u', v']$ is the element at the $[u', v']$ tap of the $M \times N$ grid of the effective channel. By taking into account all the IRS elements, the variables u_τ and v_ρ represent the maximum delay and Doppler taps, respectively.

III. Channel estimation using proposed method

A. Initial channel estimation

We consider a pilot embedded in data to perform an initial estimation of the cascaded channel. The channel for a particular DD tap $[u', v']$ is estimated as $g_{eff[u',v']} = y[u, v]/x_p$. Let Q be the total number of DD paths that need to be estimated, then, we have $\widehat{\mathbf{g}}_{eff} = [g_{i,1} ... g_{i,Q}]^T$ for $i = 1, 2, ..., I$. Thus, $\widehat{g}_{eff}$ can be represented as:

$$\hat{\mathbf{g}}_{\text{eff}} = [\mathbf{g}_1 \ \dots \mathbf{g}_I] \begin{bmatrix} e^{j\psi_1} \\ \vdots \\ e^{j\psi_I} \end{bmatrix} = \mathbb{G}\boldsymbol{\psi}, \tag{9}$$

where $\mathbb{G} \in \mathbb{C}^{Q\times I}$ is the channel matrix for IRS elements. We can estimate the $\mathbb{G}$ as:

$$\begin{aligned} &=[\hat{\mathbf{g}}^{(1)}_{\text{eff}} \ \hat{\mathbf{g}}^{(2)}_{\text{eff}} \ \dots \ \hat{\mathbf{g}}^{(I)}_{\text{eff}}][\boldsymbol{\psi}^{(1)} \ \boldsymbol{\psi}^{(2)} \ \dots \ \boldsymbol{\psi}^{(I)}]^{-1} \\ &=\hat{\mathbb{G}}_{\text{eff}} \boldsymbol{\Psi}^{-1}. \end{aligned} \tag{10}$$

The scheme utilized in the estimation process has the presence of additive white Gaussian noise. However, in order to enhance the accuracy of channel estimation in more complex noise scenarios, we introduce a data-driven approach that eradicates the noise from the initial estimate of the channel by combining the residual network and attention module.

B. Residual attention-based channel estimation

The architecture of the RACE model is depicted in Fig. 2. The RACE combines the residual blocks and attention module to create the complete residual attention architecture. The following operations are involved:

1) Residual Network: Noisy observation $\mathbf{X} \in \mathbb{R}^{Q\times I}$ is provided to the residual network. The residual network incorporates multiple residual blocks for eliminating noise from the noisy observation. Each residual block has three layers, the initial two layers perform Convolution (Conv) (with 64 filters of kernel size (3, 3)) + Batch Normalization (BN) + ReLU operations. The last layer of the residual block performs the Conv operation with (2 filters). The output of the b-th residual block is denoted as:

$$\mathbf{X}'_b = \mathbf{X}_{b-1} - F(\mathbf{X}_{b-1}). \tag{11}$$

The output of the last residual block $\mathbf{X}'$ is the noiseless observation, which is given to the attention block for further processing.

2) Attention block: In the attention block, the relationship between $\mathbf{X}'$ and $\mathbf{X}$ is calculated, to figure out the important features in the noisy observation. Further, attention weights are calculated to identify the important features in the noisy observation for accurately estimating channels. The attention weights are calculated from the following steps:
   a) We compute key $\mathbf{K}$ = Conv($\mathbf{X}$), query $\mathbf{Q}$ = Conv($\mathbf{X}'$), and value $\mathbf{V}$= Conv($\mathbf{X}$) matrix.
   b) By applying the Softmax function to the dot product of $\mathbf{K}$ and $\mathbf{Q}$, attention weights $\mathbf{A}$ are calculated as

$$\mathbf{A} = \texttt{Softmax}(\mathbf{K}^T\mathbf{Q}). \tag{12}$$

The obtained attention weights are then multiplied to the $\mathbf{V}$ to obtain the noise-free observation $\widehat{\mathbf{X}}$. The Conv layer is applied at the end to $\widehat{\mathbf{X}}$ to obtain the original shape as $\widehat{\mathbf{X}} \in \mathbb{R}^{Q\times I\times 2}$.

To train the RACE model for channel estimation, the following loss function is minimized

$$\frac{1}{S} \sum_{s=1}^{S} \left\| \mathbf{X}_s - \widehat{\mathbf{X}}_s \right\|_F^2, \tag{13}$$

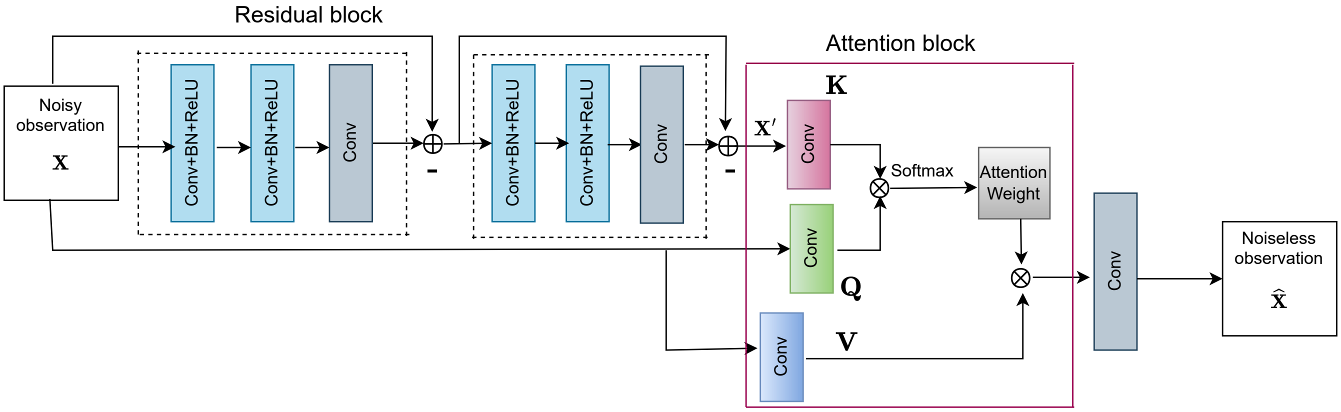

Fig. 2. Architecture of the RACE Model.

where $\mathbf{X}_s$ is the s^{th} ground truth sample in the training dataset and $\widehat{\mathbf{X}}_s$ is the predicted sample by the RACE model. For training the model Adam optimizer is used with total training samples S = 1000. At inference time, the RACE model takes a noisy channel estimate as an input and applies the learned RACE model to enhance the channel estimation accuracy.

IV. Experimental settings and results

We consider an OTFS grid with M = N= 32, f_c = 4 GHz, and Δf = 15 KHz. The Doppler and delay taps are randomly selected, and Rayleigh fading is used to model user-IRS and IRS-BS channels. The phase shift of IRS with I = 16 is obtained from the DFT matrix. The value of Q_f, Q_h = 2, is selected. Lastly, QAM modulation is adopted for bit mapping. We adopt the normalized mean square error (NMSE) to evaluate channel estimation performance. It is defined as

$$NMSE = \frac{E\|\widehat{\mathbf{G}} - \mathbf{G}\|_F^2}{E\|\mathbf{G}\|_F^2}, \tag{6}$$

where $\widehat{\mathbf{G}}$ is the estimated channel matrix and $\mathbf{G}$ is the true channel matrix.

We compare the proposed method with threshold-based [15] and residual network (ResNet) [14] that are adopted in the literature for OTFS channel estimation.

In Fig. 3, we compare the NMSE of the proposed method with the other methods against a range of SNR. It is clear from Fig. 3 that the ResNet with an attention model outperforms the ResNet alone in effectively eliminating noise from observations. Moreover, the integration of the attention module into the ResNet significantly enhances the accuracy of channel estimation, even under low SNR conditions. These results highlight the benefits of incorporating an attention mechanism into deep residual networks, as it leads to improved noise reduction and more accurate channel estimation, particularly in challenging SNR scenarios.

To assess the scalability of our proposed model, we conducted experiments to investigate the impact of varying the number of DD paths, denoted as $Q = Q_f + Q_h$, on the NMSE. This analysis enables us to evaluate the performance of our method across different channel conditions. As shown in Fig. 4, our method consistently outperforms other existing methods in terms of NMSE, even as the number of DD paths increases. The trend is evident: as we increase the number of DD paths, the NMSE remains significantly lower for our method compared to alternative approaches.

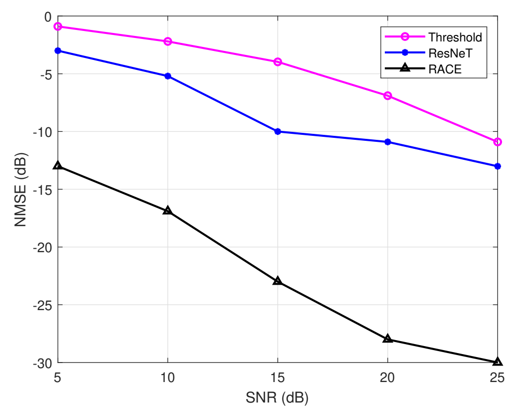

Fig. 3. NMSE vs SNR

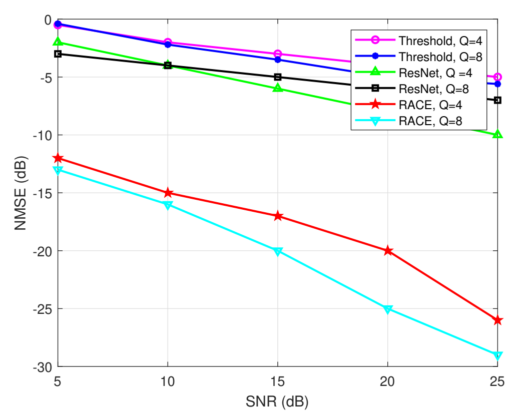

Fig. 4. NMSE vs SNR for different DD paths

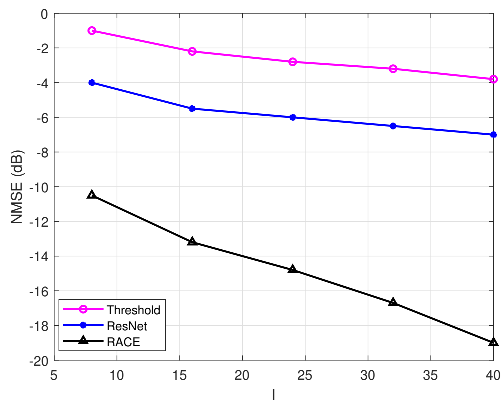

Fig. 5. NMSE vs Number of IRS elements

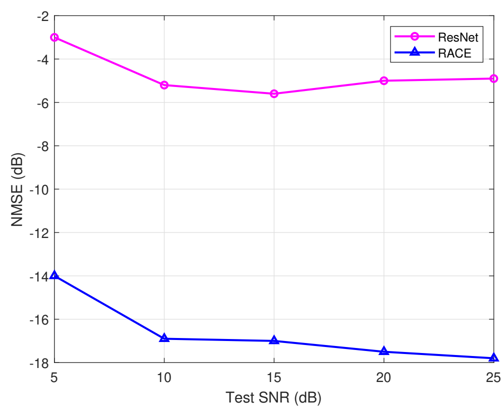

Fig. 6. NMSE vs varying SNR of test data at training SNR = 10 dB

Fig. 5 illustrates the impact of varying the number of elements of the IRS on the NMSE. By increasing the number of elements, we observe a decrease in NMSE. This improvement can be attributed to the adaptability of the RACE model to handle complex inputs.

To show the robustness of the RACE model, we vary the SNR value in the test dataset while keeping the SNR in the train data fixed at 5 dB. Fig. 6 shows that, when there is a significant gap between the SNR of the training and test data, ResNet may struggle to generalize well. This is because ResNet relies on the underlying patterns and features in the training data to make predictions. If the test data has a different SNR distribution or characteristics, ResNet's performance may deteriorate. However, our method, which incorporates an attention mechanism, can still perform well even in the presence of such gaps between the SNR of the training and test data. The attention mechanism allows the model to selectively focus on relevant information and adapt to varying SNR levels.

V. Conclusion

Our proposed approach offers an effective solution for channel estimation in dynamic scenarios of IRS-aided communication systems. By leveraging OTFS modulation, we convert the channel representation from the time-frequency domain to the DD domain, which is better suited for dynamic scenarios. Through the utilization of a RACE model, we achieve superior performance compared to existing deep learning methods and conventional approaches. The RACE model demonstrates its capability to accurately estimate the DD channel, resulting in a significantly lower NMSE when compared to alternative methods.

References

[1] W. Mei, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless networks: From single-reflection to multireflection design and optimization," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1380–1400, 2022.

[2] A. Gupta, S. Singh, A. Trivedi, and B. Prasad, "Secrecy rate maximization for irs-aided uav system using bde algorithm," in *2022 IEEE 6th Conference on Information and Communication Technology (CICT)*, 2022, pp. 1–5.

[3] S. Singh, A. Trivedi, and D. Saxena, "Unsupervised los/nlos identification in mmwave communication using two-stage machine learning framework," *Physical Communication*, p. 102118, 2023.

[4] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted miso wireless energy transfer," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4659–4663.

[5] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets ofdm: Protocol design and rate maximization," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4522–4535, 2020.

[6] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE signal processing letters*, vol. 27, pp. 905–909, 2020.

[7] H. Liu, J. Zhang, Q. Wu, H. Xiao, and B. Ai, "Admm based channel estimation for riss aided millimeter wave communications," *IEEE communications letters*, vol. 25, no. 9, pp. 2894–2898, 2021.

[8] K. Ardah, S. Gherekhloo, A. L. de Almeida, and M. Haardt, "Trice: A channel estimation framework for ris-aided millimeter-wave mimo systems," *IEEE signal processing letters*, vol. 28, pp. 513–517, 2021.

[9] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for irs-assisted millimeter-wave mimo systems: Sparsity-inspired approaches," *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 4078–4092, 2022.

[10] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.

[11] M. K. Ramachandran and A. Chockalingam, "Mimo-otfs in high-doppler fading channels: Signal detection and channel estimation," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 206–212.

[12] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for otfs in delay–doppler channels," *IEEE transactions on vehicular technology*, vol. 68, no. 5, pp. 4906–4917, 2019.

[13] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, "Channel estimation for orthogonal time frequency space (otfs) massive mimo," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4204–4217, 2019.

[14] X. Zhang, W. Yuan, and C. Liu, "Deep residual learning for otfs channel estimation with arbitrary noise," in *2022 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2022, pp. 320–324.

[15] A. Thomas, K. Deka, S. Sharma, and N. Rajamohan, "Irs-assisted otfs system: Design and analysis," *IEEE Transactions on Vehicular Technology*, 2022.

[16] Z. Li, W. Yuan, B. Li, J. Wu, C. You, and F. Meng, "Reconfigurable intelligent surface aided otfs: Transmission scheme and channel estimation," *IEEE Internet of Things Journal*, 2023.