EasyChair Preprint

№ 15751

DSRNet: Hybrid Deep Learning-Based Channel Estimation for RIS-Aided Wireless Communication

Syed Samiul Alam, Haolin Tang, Guang Yang, Changqing Luo, Wei Wang and Yanxiao Zhao

January 23, 2025

DSRNet: Hybrid Deep Learning-based Channel Estimation for RIS-Aided Wireless Communication

Syed Samiul Alam[1], Haolin Tang[2], Guang Yang[3], Changqing Luo[3], Wei Wang[4], and Yanxiao Zhao[1]

[1] Department of Electrical and Computer Engineering,
Virginia Commonwealth University, Richmond, VA, USA
{alams10,yzhao7}@vcu.edu
[2] School of Engineering and Computing,
Fairfield University, Fairfield CT, USA
htang@fairfield.edu
[3] Department of Computer Science,
Virginia Commonwealth University, Richmond VA, USA
{yangg2,cluo}@vcu.edu
[4] Department of Computer Science,
San Diego State University, San Diego, CA, USA
wwang@sdsu.edu

Abstract. Reconfigurable intelligent surfaces (RIS) are widely perceived as a transformative technology for 5G and beyond, enabling dynamic programming of wireless propagation channels. However, acquiring accurate channel state information (CSI) remains a major challenge in RIS-assisted wireless communication systems. Most existing research assumes the availability of full CSI, which is often impractical due to the passive nature of the RIS elements and the high-dimensional nature of the channels. To fill this gap, we introduce a two-stage framework, called the denoising super-resolution network (DSRnet), to estimate full CSI from partial representations. Then the estimated full CSI is utilized to maximize the weighted sum-rate (WSR) via phase shift prediction. DSRnet employs a hierarchical architecture consisting of a super-resolution sub-network for initial estimation, followed by a denoising sub-network enhanced with spatial attention modules for refined processing. The proposed model achieves impressive channel estimation performance with an NMSE of -13.14 dB while maintaining computational efficiency, utilizing only $80,916$ parameters. When the estimated full CSI is used for phase shift prediction, it shows an approximately 15% improvement in WSR compared to the methods with partial CSI. These findings verify DSRnet as a practical and efficient solution for large-scale RIS deployments, effectively balancing high performance with reduced CSI overhead.

Keywords: Reconfigurable intelligent surfaces · deep learning · denoising · channel state information · super-resolution

1 Introduction

Reconfigurable intelligent surfaces (RIS) have emerged as a promising technology for beyond 5G wireless communication systems. A RIS typically consists of

a large array of nearly passive elements that can be individually controlled by an intelligent controller to dynamically adjust the phase shifts of incident signals [12], [15]. By forming a continuous electromagnetically active surface, these elements alter the direction of signal propagation, effectively enhancing throughput, spectral efficiency, and coverage in wireless systems. The energy efficiency of these surfaces, enabled by nearly passive components, further enhances their practical viability and appeal [13]. Furthermore, RIS is proposed to manipulate channel property, thereby enhancing the performance of multiple access technologies [9]. Among the potential applications of RIS, this paper focuses on the weighted sum-rate (WSR) maximization problem in the RIS-assisted downlink broadcast channel. The WSR maximization problem has been studied in the literature. In [5], authors proposed a low-complexity block coordinate descent algorithm to maximize the WSR in RIS-aided multi-input-single-output (MISO) systems through joint optimization of beamforming and phase shifts. In [6], authors proposed an alternating optimization algorithm that used majorization-minimization (MM) for RIS-aided MISO systems, establishing the fundamental framework for joint optimization of active beamforming and passive reflection coefficients. The authors in [19] extended this to multigroup-multicast scenarios. They developed two efficient algorithms, a second-order cone programming-based MM method and a low-complexity closed-form solution approach, to maximize the sum-rate while handling the more complex group-based transmission constraints. Furthermore, in active RIS deployments, successive refinement algorithms for passive beamforming enhancement and successive convex approximation techniques for signal-to-noise ratio (SNR) optimization. However, these iterative methods exhibit significant scalability limitations. Most implementations are constrained to handling 100 or fewer RIS elements, with even advanced approaches extending to only about 400 elements. This limitation presents a critical challenge, as practical deployments often require several hundred to thousands of RIS elements to achieve adequate link budgets in diverse and complex environments [8]. In addition, these methods often rely on suboptimal approximations, which may compromise overall system performance and limit their effectiveness in real-world scenarios.

The advent of deep learning (DL) has driven significant advances in various domains of wireless communication, including modulation recognition [16], channel state information (CSI) reconstruction [2], [11] and resource allocation [17]. In recent years, different DL-based models have been utilized to optimize RIS-aided communication. Although DL-based frameworks significantly reduce inference time post-training and offer higher flexibility due to the universal approximation theorem, existing implementations often struggle with scalability challenges. Current DL approaches, including conventional neural networks, reinforcement learning frameworks, and meta-learning paradigms, remain limited by the complexity of the model that grows proportionally to the RIS elements [9]. Consequently, most machine learning-based solutions operate with restricted RIS configurations, typically not exceeding 100 [4], [7] elements.

CSI acquisition presents another critical challenge in RIS-assisted communications. The majority of the existing literature assumes full CSI availability, an impractical requirement given the large number of RIS elements in realistic deployments. Although codebook-based RIS optimization methods have been proposed as potential solutions [14], the associated beam training overhead remains a significant implementation barrier. As RIS surfaces grow larger with hundreds or thousands of elements to meet coverage and capacity demands, acquiring full CSI becomes increasingly impractical and resource intensive due to the massive overhead. In [10], the authors addressed this challenge by introducing a scalable DL model called RISnet which was further improved in [9]. RISnet introduced two variants: one assumed the availability of the full CSI and another used partial CSI from just a few active elements to configure the phase shifts of the RIS elements. Although their studies showed that RISnet with partial CSI performs well compared to RISnet with full CSI scenarios in deterministic channels, the approach did not estimate the complete channel information.

In our paper, we are motivated by the fact that obtaining full CSI will enhance the RIS phase shift configuration, as it would provide comprehensive channel knowledge for the optimization process. Therefore, to improve the phase shift prediction and the WSR, we propose a novel two-stage DL model, called deep super-resolution network (DSRnet), to estimate the full CSI from the partial inputs. The DSRnet consists of a super-resolution sub-network and a denoising sub-network. In the first stage, the super-resolution sub-network estimates the full CSI from partial representation. Then, the estimated full CSI is refined by the denoising sub-network. Finally, the refined full CSI is input into the RISnet model introduced in [10] for phase shift prediction, aiming to maximize WSR. The main contributions of our work are summarized as follows:

- We propose a novel two-stage network that efficiently estimates the full CSI from partial measurements through a hierarchical design. It begins with a super-resolution sub-network to generate an initial estimation, followed by a denoising sub-network that employs refinement techniques. This cascade design enhances phase shift accuracy while minimizing the overhead associated with CSI measurements.
- The denoising sub-network employs inception-inspired parallel dilated convolution layers to effectively capture local and global CSI features across various scales. In addition, it integrates channel attention mechanisms to adaptively weight different channel regions based on their relevance. This dual-mechanism design enhances the accuracy of CSI estimation by efficiently suppressing noise while preserving critical channel characteristics.
- Extensive simulations are conducted to evaluate the performance of our proposed DSRNet-enhanced hybrid model in improving WSR. The comparative study reveals that feeding the estimated full CSI from our DSRnet approach into the existing RISnet. This results in an approximate 15% improvement in WSR.

The remainder of the paper is organized as follows. Section 2 discusses the system model, while Section 3 provides the architecture of the proposed model.

Section 4 presents numerical results and includes an ablation study to evaluate the performance of the proposed model. Finally, Section 5 offers concluding remarks on the findings of our study.

2 System Model

In this paper, we consider a multi-user MISO system where RIS serves a single multi-antenna base station (BS) and N_K users, as shown in Fig. 1. The BS is equipped with N_T antennas, and the RIS consists of N_M reflecting elements. The antenna arrays of the BS, as well as the RIS, are arranged in a uniform linear array. The channel from BS to RIS is denoted by $\mathbf{H}_{\text{BS-RIS}} \in \mathbb{C}^{N_M \times N_T}$. The channel from RIS to users is indicated by $\mathbf{H}_{\text{RIS-User}} \in \mathbb{C}^{N_K \times N_M}$ and the direct channel from BS to the user is indicated by $\mathbf{H}_{\text{Direct}} \in \mathbb{C}^{N_K \times N_T}$. The objective is to employ space-divison multiple access to maximize the users' sum-rate. The BS performs precoding under a maximum transmit power constraint E_{Tr}. Each RIS element receives the signal from the BS, applies a complex phase shift to it, and reflects the signal without altering its amplitude. The RIS applies phase shifts to the incident signals through a diagonal matrix $\mathbf{\Theta} \in \mathbb{C}^{N_M \times N_M}$, where the diagonal element $\theta_m = e^{j\phi_m}$, represents the phase shift, $\phi_m \in [0, 2\pi)$ applied by the m^{th} RIS element.

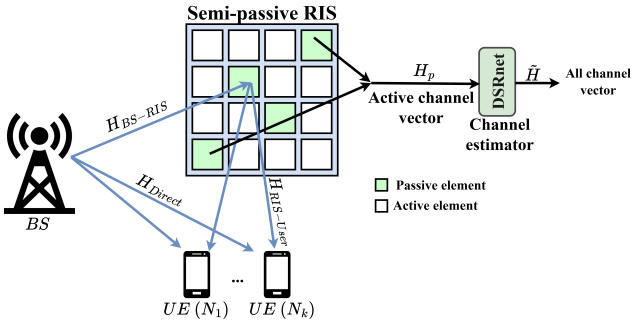

Fig. 1. System model of the RIS-aided MISO system with partial CSI.

The received signal vector $\mathbf{y} \in \mathbb{C}^{N_K \times 1}$ at the users can be represented as follows:

$$\mathbf{y} = \left(\mathbf{H}_{\text{RIS-User}} \mathbf{\Theta} \mathbf{H}_{\text{BS-RIS}} + \mathbf{H}_{\text{Direct}}\right) \mathbf{V}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{V} \in \mathbb{C}^{N_T \times N_K}$ is the precoding matrix, $\mathbf{x} \in \mathbb{C}^{N_K \times 1}$ is the transmitted signal, and $\mathbf{n} \in \mathbb{C}^{N_K \times 1}$ is the thermal noise.

In practical implementations, obtaining the full CSI for all RIS elements poses substantial overhead. Therefore, this paper considers partial CSI instead of full CSI. Specifically, we consider a system where only $N_S \ll N_M$ elements are equipped with RF chains for channel estimation, yielding partial CSI, $\mathbf{H}_{\text{Partial}} \in \mathbb{C}^{N_K \times N_S}$. To effectively process this partial CSI, we define channel features that capture both amplitude and phase information. For user k and reflecting element m, we construct a feature vector $\mathbf{f}_{k,m}$ as:

$$\mathbf{f}_{k,m} = \begin{bmatrix} |\mathbf{H}_{\text{Partial},k,m}| \\ \arg(\mathbf{H}_{\text{Partial},k,m}) \end{bmatrix} \in \mathbb{R}^{2\times 1}, \tag{2}$$

where $|\cdot|$ represents the amplitude and $\arg(\cdot)$ represents the phase of the complex channel gain from the m^{th} reflecting element to user k. The partial channel feature $\mathbf{X} \in \mathbb{R}^{2\times N_K \times N_S}$ forms a three-dimensional tensor containing all $\mathbf{f}_{k,m}$ features.

To handle the direct channel component, we define $\mathbf{\Gamma} = \mathbf{H}_{\text{Direct}}\mathbf{H}'_{\text{BS-RIS}}$ denotes the pseudo-inverse of the BS-RIS channel $\mathbf{H}'_{\text{BS-RIS}}$. For each user k and reflecting element m, we construct a feature vector $\boldsymbol{\delta}_{k,m}$ as:

$$\boldsymbol{\delta}_{k,m} = \begin{bmatrix} |\mathbf{\Gamma}_{k,m}| \\ \arg(\mathbf{\Gamma}_{k,m}) \end{bmatrix} \in \mathbb{R}^{2\times 1}. \tag{3}$$

The complete direct channel feature $\boldsymbol{\Delta} \in \mathbb{R}^{2\times N_K \times N_S}$ contains all $\boldsymbol{\delta}_{k,m}$ features. The channel matrix $\mathbf{H}$ is then estimated using the DSRNet model as:

$$\hat{\mathbf{H}} = \mathcal{F}_{\text{DSR}}(\mathbf{X}). \tag{4}$$

So, Eq. 1 becomes:

$$\mathbf{y} = \big(\hat{\mathbf{H}}\mathbf{\Theta}\mathbf{H}_{\text{BS-RIS}} + \mathbf{\Gamma}\big)\mathbf{V}\mathbf{x} + \mathbf{n}. \tag{5}$$

The combined channel matrix $\mathbf{C} = \hat{\mathbf{H}}\mathbf{\Theta}\mathbf{H}_{\text{BS-RIS}} + \mathbf{\Gamma}$ captures the overall effect of both direct and RIS-assisted paths. The signal-to-noise-plus-interference ratio (SNIR) is:

$$\text{SINR}_{n,k} = \frac{|\mathbf{C}_k\mathbf{v}_k|^2}{\sum_{n\neq k}|\mathbf{C}_k\mathbf{v}_n|^2 + \sigma^2}, \tag{6}$$

where $\mathbf{v}_k$ is the kth column of $\mathbf{V}$, σ^2 is the noise power, and $\mathbf{C}_k$ represents the kth row of the combined channel matrix $\mathbf{C}$. The system aims to maximize the weighted sum-rate subject to transmit power and RIS phase shift constraints:

$$\text{WSR} = \sum_{k=1}^{N_K} \omega_k \log_2(1 + \text{SINR}_k), \tag{7}$$

where ω_k is a weight for user k, with the constraint:

$$\text{tr}(\mathbf{V}^{\mathsf{H}}\mathbf{V}) \leq E_{\text{Tr}}. \tag{8}$$

The weighted minimum mean squared error precoder is considered for the precoding at the BS and RISnet is used to predict $\mathbf{\Theta}$ from the full CSI.

3 Structure of the Proposed Model

This section introduces the architecture of the proposed DSRnet model for CSI estimation. The network employs a dual-network design, divided into two primary components: the super-resolution sub-network and the denoising sub-network. Both components will be elaborated on in the following subsections.

3.1 Super Resolution Sub-Network

The super-resolution sub-network (see Fig. 3(a)) is designed to estimate full CSI from partial CSI data by leveraging a series of convolutional (conv) blocks with varying kernel sizes and upsampling layers. This architecture enables the generation of high-resolution CSI from lower-resolution inputs. Initially, the partial

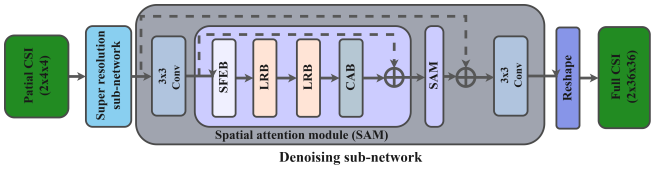

Fig. 2. Architecture of the proposed DSRnet for CSI estimation.

CSI input, with dimensions of 4×4, passes through the first conv block, featuring a kernel size of 9×9, followed by a Leaky ReLU (LReLU) activation function. The LReLU is defined as:

$$\text{LReLU}(z) = \begin{cases} z, & \text{if } z > 0, \\ \beta z, & \text{otherwise.} \end{cases} \tag{9}$$

where β is a hyper-parameter that scales the negative values of z. This initial layer captures a broad spatial context. The resulting feature map is then upsampled to a resolution of 12×12. Following this, a 1×1 conv layer is applied to refine the upsampled features without increasing the spatial dimensions. Subsequently, the output undergoes an additional upsampling step to achieve a 36×36 resolution. The final step involves a conv block with a 5×5 kernel, which refines and accurately reconstructs the high-resolution CSI output to match the original full-resolution dimension.

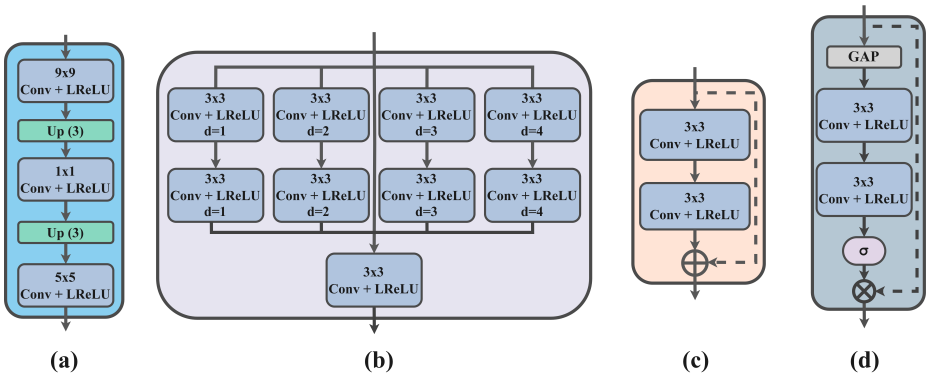

Fig. 3. (a) Super-resolution sub-network, (b) Spatial feature extraction block, (c) Local residual block, (d) Channel attention block.

3.2 Denoising Sub-Network

The denoising sub-network is designed to refine the estimated full CSI. This sub-network consists of two sequential spatial attention modules (SAM). The output from the super-resolution sub-network is first passed through the initial SAM, which enhances critical spatial features while suppressing noise. This refined output is then additively combined with the original output of the super-resolution sub-network, forming an intermediate representation. This combined

representation is subsequently fed into the second SAM for further refinement. The detailed architecture of each SAM is described as follows:

The SAM consists of a spatial feature extraction block (SFEB), two local residual blocks (LRB) a channel attention block (CAB). Fig. 3(b), 3(c), and 3(d) depicts the detailed architectures of SFEB, LRBs and CAB, respectively. The SFEB includes multiple parallel conv blocks with 3×3 kernels, each configured with different dilation rates to capture varying spatial details. Specifically, a 3×3 conv kernel with a dilation rate (d) of 2 acts as an effective 5×5 kernel; similarly, for $d = 3$ and $d = 4$, the effective kernel sizes are 7×7 and 9×9, respectively. These varied kernel sizes allow the module to extract a broad range of spatially denoised features from the reconstructed CSI provided by the SR sub-network.

The spatial features extracted across different channels are concatenated with the super-resolution CSI and passed to subsequent layers. This concatenated feature map is further processed through two LRBs, enhancing feature learning. Global average pooling (GAP) is applied to incorporate global contextual information. A gating mechanism further refines this process by capturing channel dependencies: the GAP output is first downsampled using a convolutional layer and then processed through a sigmoid activation function (σ), producing soft attention weights ranging between 0 and 1. These weights adaptively rescale the input to the SAM module, enhancing feature selectivity.

The SAM also utilizes multiple long and short residual connections to retain previously extracted features, which help effectively mitigate gradient vanishing and explosion issues. The output from the two successive SAMs undergoes upsampling via a convolutional block, which is combined with the input and further processed through a final convolutional block. This output is then fed to the RISNet model for phase shift prediction, ensuring optimized performance.

4 Result and Discussion

4.1 Dataset and Training Strategy

The training and evaluation datasets are generated using the DeepMIMO ray-tracing simulator [1], which provides realistic CSI for RIS-aided communication scenarios. The dataset consists of channel data between a base station equipped with 9 antennas, a 36×36 RIS array (1296 elements), and multiple user terminals. For training purposes, we use $20,480$ samples while reserving $1,024$ samples for testing. In the partial CSI scenario, only 16 RIS elements (4×4) are used for channel estimation, representing approximately 1.2% of the total RIS elements. The network is trained using Adam Optimizer with an initial learning rate of 9×10^{-4}. As our objective is jointly optimizing the CSI estimation as well as the WSR, we employ a hybrid loss function that dynamically balances between channel reconstruction accuracy and system throughput optimization:

$$\mathcal{L} = \begin{cases} \alpha \cdot \mathcal{L}_{\text{rec}} + (1-\alpha) \cdot \mathcal{L}_{\text{wsr}} & \text{for } 0 \leq \text{epoch} < T \\ \mathcal{L}_{\text{wsr}} & \text{for epoch} \geq T \end{cases} \tag{10}$$

where, $\mathcal{L}_{\text{rec}}$ is the reconstruction loss and $\mathcal{L}_{\text{wsr}}$ is the WSR loss. α and T are the weighting factor and transition epoch, respectively. In our experiments we

have considered $\alpha = 0.5$ and $T = 200$. These values were chosen via extensive hyper-parameter tuning. The ($\mathcal{L}_{\text{rec}}$) can be represented as: $\mathcal{L}_{\text{rec}} = \|\hat{H} - H\|_2^2$. The $\mathcal{L}_{wsr}$ can be defined as: $\mathcal{L}_{\text{wsr}} = -\log_2(1 + \text{SINR})$, where SINR is the signal-to-interference-plus-noise ratio.

4.2 Evaluation Metrics

In order to evaluate the performance of the proposed CSI estimation model and facilitate meaningful comparisons, we employ three metrics that capture distinct aspects of CSI estimation quality. They are normalized mean square error (NMSE), peak signal-to-noise ratio (PSNR), and complex correlation coefficient (ρ). The NMSE measures the normalized difference between the estimated channel matrix and the ground truth, defined as:

$$\text{NMSE } (dB) = 10 \log_{10} \left(\mathbb{E} \left[\frac{\|\hat{\mathbf{H}} - \mathbf{H}\|_2^2}{\|\mathbf{H}\|_2^2} \right] \right), \tag{11}$$

where $\mathbb{E}$ denotes the values of expectation. PSNR quantifies the noise level sensitivity of the CSI estimation model and can be defined as:

$$\text{PSNR } (dB) = 10 \log_{10} \left(\frac{\max(\mathbf{H})}{\text{MSE}^2} \right), \tag{12}$$

where MSE is the cumulative squared error between $\hat{\mathbf{H}}$ and $\mathbf{H}$. The complex correlation coefficient ρ represents the similarity between the actual and estimated channels in both magnitude and phase:

$$\rho = \frac{\mathbb{E}\left[(\mathbf{H} - \bar{\mathbf{H}}) \cdot (\hat{\mathbf{H}} - \bar{\hat{\mathbf{H}}})^* \right]}{\sqrt{\mathbb{E}\left[|\mathbf{H} - \bar{\mathbf{H}}|^2 \right] \cdot \mathbb{E}\left[|\hat{\mathbf{H}} - \bar{\hat{\mathbf{H}}}|^2 \right]}}, \tag{13}$$

where $\bar{\mathbf{H}}$ and $\bar{\hat{\mathbf{H}}}$ represent the mean values of the actual and the estimated channel matrices and $(\cdot)^*$ denotes the complex conjugate.

4.3 Ablation Study

To understand the effectiveness of our proposed DSRnet architecture and investigate the impact of the model depth, we conducted an ablation study focusing on the number of SAMs. Table 1 presents the comparative analysis of models with one, two, and three SAM configurations. The results demonstrate that increasing the number of SAM modules generally improves the estimation performance. The model with one SAM achieves an NMSE of -11.19 dB and a PSNR of 23.08 dB with $44,692$ parameters. Adding a second SAM significantly improves performance, reaching an NMSE of -13.14 dB and a PSNR of 25.75 dB, while requiring $80,916$ parameters. This represents an improvement of approximately 2 dB in NMSE and 2.67 dB in PSNR. Further increasing to three SAM modules yields marginal improvements, with NMSE reaching -13.65 dB and PSNR improving to 26.01 dB. However, this configuration requires $117,140$ parameters, representing a significant increase in model complexity. Moreover,

the ρ saturates at 0.934 for both two and three SAM configurations, which means that, two modules are sufficient for capturing the complex channel characteristics. Based on these results, we selected the double-SAM configuration as our final architecture, as it provides an optimal balance between performance and computational efficiency for practical RIS deployments.

Table 1. Ablation Study

Method	**NMSE (dB)**	**PSNR**	ρ	**Number of Parameters**
Single SAM	-11.19	23.08	0.91	$44,692$
double SAM	-13.77	26.39	0.934	$80,916$
Triple SAM	-13.80	26.38	0.934	$117,140$

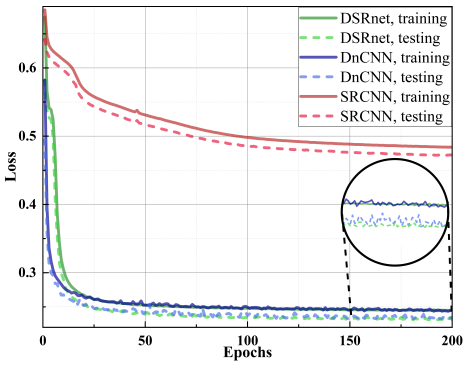

Fig. 4. Comparison between training and testing loss of DSRnet with SRCNN and DnCNN

4.4 Chanel Estimation Performance

In order to evaluate our proposed DSRnet architecture regarding CSI estimation performance, we compare it with two baseline methods: super-resolution convolutional neural network (SRCNN)[3] and denoising convolutional neural network (DnCNN)[18]. The training and testing loss curves (see Fig. 4) demonstrate that DSRnet and DnCNN achieve a similar convergence performance with reconstruction loss values of approximately 0.23, significantly outperforming SRCNN which converges to a higher loss of 0.48. This highlights DSRnet's remarkable stability and generalization capabilities, exhibiting minimal fluctuations in both training and testing phases compared to DnCNN. As shown in Table 2, our proposed model outperforms across all evaluation metrics while maintaining a compact and efficient model size. The DSRnet achieves an NMSE of -13.14 dB, an improvement of approximately 2.9 dB over DnCNN (-10.24 dB) and nearly 11 dB over SRCNN (-2.41 dB). This reduction in NMSE indicates that our model demonstrates superior accuracy in reconstructing the CSI from partial observations. Regarding PSNR, DSRnet achieves 25.75 dB, outperforming DnCNN (25.02 dB) and SRCNN (15.01 dB) by considerable margins. The higher PSNR value indicates that our model generates reconstructions with superior signal quality and reduced distortion compared to the baseline methods. Furthermore,

the ρ of 0.934 achieved by our model verifies superior accuracy in reconstructing both amplitude and phase information of the complex-valued channel states, surpassing both DnCNN (0.930) and SRCNN (0.664). This high complex correlation demonstrates that our model better preserves complex-valued channel characteristics, which is crucial for accurate CSI estimation in wireless communications.

Moreover, our proposed model achieves these superior results with only 80,916 parameters, significantly outperforming DnCNN (121, 792) and striking a better balance of performance and complexity compared to the simpler SRCNN (7, 874). This parameter efficiency is important for practical deployment in resource-constrained environments and real-time applications in RIS-aided communication systems.

Table 2. Comparison of the Methods

Method	**NMSE (dB)**	**PSNR**	ρ	**Number of Parameters**
SRCNN	−2.41	15.01	0.664	**7,874**
DnCNN	−12.40	25.02	0.930	121, 792
Proposed Model	**-13.77**	**26.39**	**0.934**	80, 916

4.5 Comparison with RISnet Model

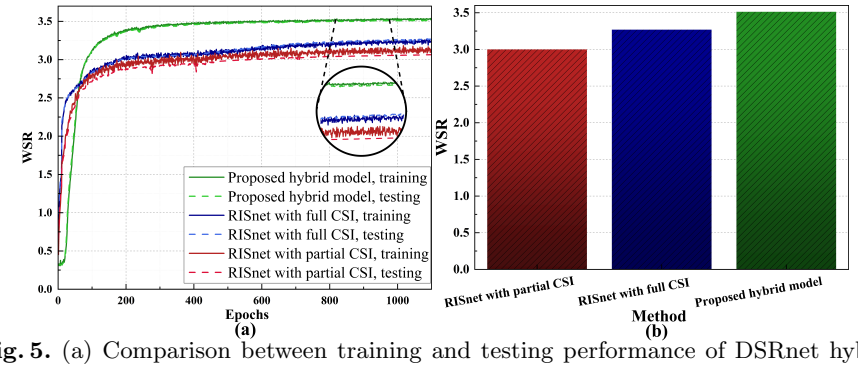

Fig. 5. (a) Comparison between training and testing performance of DSRnet hybrid optimization with RISnet partial and full CSI, (b) The comparison of test results for different methods

The ultimate objective of this work is to optimize phase shift configurations for a large-scale RIS through joint channel estimation and WSR maximization. Therefore, we compare our proposed DSRnet-enhanced hybrid model with the RISnet models in terms of WSR maximization. Note that here we consider two RISnet variants, one with partial CSI and another with pre-known full CSI. In the DSRnet-enhanced hybrid model, the reconstructed full CSI from the DSRnet was fed into RISnet. The training and testing curves of the hybrid model shown in Fig. 5 (a), exhibit minimal overfitting compared to RISnet, highlighting the superior generalization capability of the proposed approach. Moreover, Fig. 5(b) shows that the proposed hybrid approach achieves considerably higher WSR performance than the original RISnet approach. The DSRnet-enhanced hybrid

model shows a WSR of 3.51 bits/s, approximately 15% improvement, when compared to the RISnet model with partial CSI. The improved performance of the hybrid model can be attributed to its superior channel estimation capabilities, which effectively utilize limited CSI information to generate more accurate and reliable channel information. The exceptional performance of the hybrid model compared to the RISnet with full CSI can be attributed to the denoising effect of the DSRnet model. By reconstructing channel information from fewer but more reliable measurements, DSRnet can filter out noise while preserving the essential channel characteristics, leading to a more robust RIS phase configuration.

5 Conclusion

This paper presents a novel two-stage architecture named DSRnet to reconstruct full CSI from partial CSI for RIS-aided wireless communication. DSRNet incorporates a super-resolution sub-network for coarse estimation and a denoising sub-network enhanced with advanced spatial attention mechanisms, delivering superior CSI estimation accuracy while maintaining parameter efficiency. Specifically, it is efficient while requiring CSI from only 1.2% of the RIS elements. This makes it particularly suitable for practical large-scale RIS deployments. We have conducted extensive simulations, and the results demonstrated that DSRnet successfully reconstructs full CSI from minimal partial measurements of a large RIS array, achieving strong performance metrics (NMSE: -13.14 dB, PSNR: 25.75 dB, ρ: 0.934) while maintaining computational efficiency with only $80,916$ parameters. More importantly, when the reconstructed CSI is fed into the phase shift prediction network RISnet, it demonstrates a WSR of 3.51 bits/s, which is approximately 15% improvement compared to the RISnet with partial CSI. This substantial WSR improvement, combined with minimal CSI overhead and efficient scaling to large RIS arrays, verifies DSRnet's practical value for real-world deployments.

References

1. Alkhateeb, A.: DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications. In: Proc. of Information Theory and Applications Workshop (ITA). pp. 1–8. San Diego, CA (Feb 2019)
2. Chakma, A., Alam, S.S., Rahman, M.H., Jang, Y.M.: Deep decoder csinet for fdd massive mimo system. IEEE Wireless Communications Letters **12**(12), 2073–2077 (2023). https://doi.org/10.1109/LWC.2023.3307164
3. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision – ECCV 2014. pp. 184–199. Springer International Publishing (2014)
4. Feng, K., Wang, Q., Li, X., Wen, C.K.: Deep reinforcement learning based intelligent reflecting surface optimization for miso communication systems. IEEE Wireless Communications Letters **9**(5), 745–749 (2020). https://doi.org/10.1109/LWC.2020.2969167
5. Guo, H., Liang, Y.C., Chen, J., Larsson, E.G.: Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks. IEEE Transactions on Wireless Communications **19**(5), 3064–3076 (2020). https://doi.org/10.1109/TWC.2020.2970061

6. Huang, C., Zappone, A., Debbah, M., Yuen, C.: Achievable rate maximization by passive intelligent mirrors. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3714–3718 (2018). https://doi.org/10.1109/ICASSP.2018.8461496
7. Jung, M., Saad, W.: Meta-learning for 6g communication networks with reconfigurable intelligent surfaces. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8082–8086 (2021). https://doi.org/10.1109/ICASSP39728.2021.9413598
8. Najafi, M., Jamali, V., Schober, R., Poor, H.V.: Physics-based modeling and scalable optimization of large intelligent reflecting surfaces. IEEE Transactions on Communications **69**(4), 2673–2691 (2021)
9. Peng, B., Besser, K.L., Raghunath, R., Jamali, V., Jorswieck, E.A.: Risnet: A scalable approach for reconfigurable intelligent surface optimization with partial csi. In: GLOBECOM 2023 - 2023 IEEE Global Communications Conference. pp. 4810–4816 (2023). https://doi.org/10.1109/GLOBECOM54140.2023.10437049
10. Peng, B., Siegismund-Poschmann, F., Jorswieck, E.A.: Risnet: a dedicated scalable neural network architecture for optimization of reconfigurable intelligent surfaces. In: WSA SCC 2023; 26th International ITG Workshop on Smart Antennas and 13th Conference on Systems, Communications, and Coding. pp. 1–6 (2023)
11. Samiul Alam, S., Chakma, A., Imran, A., Min Jang, Y.: Dcsinet: Effective handling of noisy csi in fdd massive mimo system. IEEE Wireless Communications Letters **13**(9), 2556–2560 (2024). https://doi.org/10.1109/LWC.2024.3427316
12. Sarp, S., Tang, H., Zhao, Y.: Use of intelligent reflecting surfaces for and against wireless communication security. In: 2021 IEEE 4th 5G World Forum (5GWF). pp. 374–377. IEEE (2021)
13. Shen, W., Qin, Z., Nallanathan, A.: Deep learning for super-resolution channel estimation in reconfigurable intelligent surface aided systems. IEEE Transactions on Communications **71**(3), 1491–1503 (2023). https://doi.org/10.1109/TCOMM.2023.3239621
14. Taha, A., Alrabeiah, M., Alkhateeb, A.: Enabling large intelligent surfaces with compressive sensing and deep learning. IEEE Access **9**, 44304–44321 (2021)
15. Tang, H., Sarp, S., Zhao, Y., Wang, W., Xin, C.: Security and threats of intelligent reflecting surface assisted wireless communications. In: 2022 International Conference on Computer Communications and Networks (ICCCN). pp. 1–9. IEEE (2022)
16. Tang, H., Zhao, Y., Kuzlu, M., Luo, C., Catak, F.O., Wang, W.: Automatic modulation recognition using parallel feature extraction architecture. In: Wireless Artificial Intelligent Computing Systems and Applications. pp. 216–228 (2025). https://doi.org/10.1007/978-3-031-71467-2˙18
17. Ye, H., Li, G.Y., Juang, B.H.F.: Deep reinforcement learning based resource allocation for v2v communications. IEEE Transactions on Vehicular Technology **68**(4), 3163–3173 (2019)
18. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017). https://doi.org/10.1109/TIP.2017.2662206
19. Zhou, G., Pan, C., Ren, H., Wang, K., Nallanathan, A.: Intelligent reflecting surface aided multigroup multicast miso communication systems. IEEE Transactions on Signal Processing **68**, 3236–3251 (2020). https://doi.org/10.1109/TSP.2020.2990098