Deep Learning for Super-Resolution Channel Estimation in Reconfigurable Intelligent Surface Aided Systems

Wenhan Shen, Zhijin Qin, *Senior Member, IEEE* and Arumugam Nallanathan, *Fellow, IEEE*

Abstract

Reconfigurable intelligent surface (RIS) enables the configuration of the propagation environment. Channel estimation is an essential task in realizing the RIS-aided communication system. A RIS-aided multi-user multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) communication system involves cascaded channels with high dimensions and sophisticated statistics. Thus, implementing the optimal minimum mean square error (MMSE) with the integration computation is infeasible in practice. To accurately estimate channels with high accuracy in a RIS-aided multi-user MIMO-OFDM system, we model the channel state information (CSI) estimation as an image super-resolution (SR) problem to recover and denoise the channel matrix. Particularly, a convolutional neural network based on a super-resolution convolutional neural network (SRCNN) and denoising convolutional neural network (DnCNN), named SRDnNet, is then proposed. By taking estimated channels at pilot positions as a low-resolution image, the enhanced SRCNN can fully exploit the features of inputs to learn a suitable interpolation method and generate the coarse estimation of the channel matrix. The denoising model DnCNN with an element-wise subtraction structure can exploit features of the additive noise and recover channel coefficients from the coarse channel matrix. The simulation results demonstrate the effectiveness and excellent performance of the proposed SRDnNet.

Index Terms

Channel estimation, deep learning, OFDM, multi-user, reconfigurable intelligent surface.

I. INTRODUCTION

The rapid growth of communication types of equipment demands high data rates up to 1 terabyte per second (TB/s), super-low latency of less than 1 ms, super-high connectivity density of more than 10^7 devices/km^2, ten times spectrum efficiency of the fifth generation (5G) mobile network and other indices in energy efficiency, mobility, area traffic capacity, etc [2]. To meet these requirements, device-centric architectures, millimeter wave (mmWave), massive multiple-input multiple-output (MIMO) and other wireless technologies are proposed to lead to disruptive changes in both architectural and component design [3]. However, challenges from high hardware design requirements, excessive energy cost and complex signal processing prevent these technologies from implementation in practice [4]. Thanks to the development of materials, a novel concept of reconfigurable intelligent surface (RIS) has recently been proposed as a potential solution to these difficulties.

RIS, which can shape a wireless propagation environment by intelligently reflecting incident signals through its reconfigurable elements, has been considered a promising technology for the future smart radio environment [5]–[7]. Typically, a RIS consists of numerous nearly passive reflecting elements, which can be separately controlled by a smart controller to adapt its phase shift on incident signals and alter their propagation direction. By adjusting the phase shift of reflected signals, the RIS can program and reshape the wireless propagation environment to strengthen the received power at intended receivers [8]. The implementation of the RIS can improve throughput, coverage, energy efficiency (EE) and spectral efficiency (SE) [9] by adapting its phase shift. Therefore, RIS-aided communication systems and related studies, e.g., the throughput maximization with different reflection patterns [10], the EE and SE trade-off [9] and the RIS-aided physical layer security [11], [12], have attracted extensive interest from academia.

In RIS-aided wireless communication systems, reaping the promising performance gain requires precise channel state information (CSI) to perform beamforming. However, these studies mentioned above were completed with perfect CSI, which is usually unavailable. In fact, the acquisition of CSI is a crucial task in realizing RIS-aided communication systems. Compared with channel estimation in conventional systems, the RIS can only reflect signals without the ability to transmit/receive pilots and signal processing. In this case, acquiring the channel from RIS

to the user/base station (BS) is infeasible via pilot-based channel estimation methods. Only the cascaded channel of the BS-RIS-user can be estimated by adequately designing the transmission protocol and reflection pattern of the RIS. However, the estimation of the cascaded channel is challenging due to the following difficulties:

1) The cascaded channel does not follow the complex Gaussian distribution. In this case, it is infeasible to derive the optimal minimum mean square error (MMSE) estimator, which requires multiple integrals for practical implementation.

2) The achievable linear MMSE (LMMSE) and least squares (LS) estimators have a significant performance disparity compared to the optimal MMSE. Consequently, the channel estimation accuracy is limited for realizing RIS-aided communication systems.

3) The RIS generally consists of hundreds or even thousands of reflecting elements, leading to a highly dimensional cascaded channel. Therefore the pilot overhead is excessively high. The analytical channel estimation approaches, such as the LS estimator and the LMMSE estimator, bring a heavy workload on computation.

To address the difficulties of channel estimation in RIS-aided communication systems, extensive works have been carried out [13]–[17]. Mishra *et al.* [13] proposed a novel low-complexity optimal channel estimation protocol for the single-user system. Only one of the RIS reflecting elements is switched on and the rest are switched off simultaneously. In this case, the BS can receive pilots without inter-element interference of the RIS, and the estimation of cascaded channels can be conducted. However, the signal-to-noise ratio (SNR) obtained at the BS is limited for CSI estimation, as only one active reflecting element reflects the pilots in each time slot. Moreover, the massive number of reflecting elements introduces an exceedingly long latency to the system. Jensen *et al.* [14] activate all RIS elements to enhance the received SNR as well as to shorten the latency. Particularly, a novel RIS activation pattern has been designed to achieve the minimum variance unbiased estimation of the cascaded channel. The pilot cost in [13], [14] is determined by the product of the number of antennas, RIS elements and users, which could hinder the application of the RIS with massive reflecting elements to communication systems. On the other hand, some attempts have been dedicated to lowering the pilot overhead by designing efficient schemes. He *et al.* [15] first formulated the channel estimation in RIS-aided MIMO systems as recovering a sparse channel matrix, in which the sparsity of the spatial channel matrix is exploited by applying compressed sensing to reduce the pilot overhead.

To reduce the pilot overhead for channel estimation introduced by massive RIS reflecting elements, [16]–[21] have proposed novel CSI estimation frameworks in which they grouped those adjacent RIS elements into several sub-surfaces. Within every sub-surface, the RIS elements adopt the same configuration. Therefore, the computation power of channel estimation can be dramatically reduced. Zheng *et al.* [16] have proposed a transmission protocol for the channel estimation, while it is not an energy-efficient manner since all RIS elements are always active. Besides, we note that the pre-designed phase shift for each transmitted symbol should be configured when the symbol arrives at the RIS. Otherwise, the propagation environment is not the same as pre-designed. The above works significantly contribute to channel estimation for RIS-aided communication scenarios. However, the limited channel estimation performance for cascaded channels has not been well addressed. Thus a practical and efficient channel estimation scheme is more than desired to estimate the RIS channels accurately.

Deep learning (DL) [22] has been widely used in RIS-aided communication systems [23]–[25], such as the deep reinforcement learning (DRL)-based SNR maximization, DRL-based secure RIS multi-user communication systems and the DL-based active or passive beamforming design. It is noted that these promising DL-based approaches still demand perfect CSI, which is hard to obtain in practice. Up to now, DL-based channel estimation methods have been widely investigated in various conventional communication systems [26]–[28]. However, DL-based techniques have not been thoroughly investigated for channel estimation in RIS-aided wireless communication scenarios. Some attempts are devoted to designing DL-based frameworks to improve channel estimation performance [29]–[32].

A variety of DL-based algorithms have been proposed to improve the limited channel estimation performance in RIS-aided communication systems. Liu *et al.* [29] developed a deep residual network-based method and a convolutional neural network (CNN)-based network (CDRN) to improve channel estimation performance in RIS-aided multi-user communication systems. Although the estimation performance of CDRN is promising, we should note that inputs of the CDRN are generated by the LS estimator, which is costly computational because of the high-dimension cascaded channel. Kundu *et al.* [30] adopted two CNN-based methods, i.e., denoising CNN (DnCNN) and fast and flexible denoising Network (FFDNet), to perform channel denoising in a RIS-aided single user multiple input single output (MISO) system. The noise map (standard deviation of the additive noise in the LS estimate) is concatenated with the LS channel matrix as the input to the FFDNet. It provides flexibility to handle different noise

levels. However, prior Knowledge of the additive noise is usually unavailable in practice, and the LS estimator generates the inputs of both CNN-based methods. Liu *et al.* [31] proposed a complex-valued denoising convolution neural network (CV-DnCNN) to enhance compressive sensing-based channel estimation performance in a RIS-aided mmWave massive MIMO system. Jin *et al.* [32] proposed two practical residual neural networks, i.e., single-scale enhanced deep residual (EDSR) and multi-scale enhanced deep residual (MDSR), to recover channel matrix from the sparse RIS channels in a time division duplex (TDD) RIS-aided mmWave communication system. In particular, the RISs in [31] and [32] are hybrid in that a few reflecting elements are active and connected to the baseband unit (i.e., a radio frequency (RF) chain) for signal processing. Besides, the estimation of the RIS channel is processed at the receiver to generate the inputs of the CV-DnCNN, the EDSR and the MDSR.

It is noted that estimated channels in RIS-aided communication systems must be obtained as inputs of neural works, which is computationally intensive and highly affect the performance of neural works [29]–[32]. Pilot-based channel estimation is the most commonly used method to acquire the CSI in OFDM systems. The pilot channels can be first estimated by the pilots known at transmitters and receivers. Then the data channels can be calculated by interpolation methods, e.g., spline and linear interpolation methods, based on the estimated channels at pilot positions. Interestingly, the Image super-resolution (SR), which aims at recovering a high-resolution (HR) image from a low-resolution (LR) and noisy image, is comparable to channel estimation. Various attempts have been conducted to estimate channels in non-RIS communication scenarios based on the image SR algorithm [33]–[35]. Ouyang *et al.* [33] have proposed the channel super-resolution neural network (CSRNet) for channel estimation in underwater acoustic (UWA) orthogonal frequency division multiplexing (OFDM) communications. Soltani *et al.* [34] have developed a two-stage CNN-based channel estimation scheme (ChannelNet) to recover the channel matrix from the initial estimation of channels at pilot positions. ChannelNet is the concatenation of a super-resolution convolutional neural network (SRCNN) and a denoising convolutional neural network (DnCNN). Li *et al.* [35] have developed a deep residual channel estimation network (ReEsNet) to perform CSI estimation in downlink OFDM systems.

Note that the CSRNet and ChannelNet generate the complete channel matrix by the spline and bicubic interpolation, which are fixed and can not be trained. The ReEsNet adopted transposed convolution layer as the up-sampling layer and upscaled the LR channel matrix to the desired size. However, the up-sampling layer is the penultimate layer, and only a convolutional layer

is concatenated to generate the output without denoising. Moreover, the ChannelNet is not an end-to-end model, its estimation performance needs further improvement. Motivated by this, we formulate the channel estimation as an image SR problem and propose the revised SRCNN with a trainable interpolation method to generate the coarse channel matrix and concatenate it to the DnCNN to denoise the channel matrix and further improve channel estimation performance. In this work, we concentrate on channel estimation in RIS-aided multi-user MIMO-OFDM systems with the proposed DL-based scheme, SRDnNet. We modeled channel estimation as a denoising problem and proposed a deep residual learning-based channel estimation method for a downlink RIS-aided OFDM systems [36]. In [36], compared with this work, each user collects their datasets to train their local models and only transmits model weights to the BS for updating the global model. The updated global model is then broadcasted to all users for the next training turn until convergence. The main contributions of this paper are summarized as follows:

1) To separately estimate the direct and cascaded channels, we propose a three-stage transmission protocol in which we consider that the pre-designed phase shift in a time slot should correspond to transmitted symbols in the same time slot. By applying the transmission protocol, the mismatch of the pre-designed phase shift and symbols can be mitigated and eliminated. Besides, the RIS works energy-efficiently since RIS only activates its elements when receiving signals from users.

2) Compared with related works, we formulate the channel estimation problem in a RIS-aided multi-user MIMO-OFDM communication system as an image SR problem. Based on this, we develop a deep CNN-based algorithm, SRDnNet, to accurately recover the channel matrix from estimating channels at pilot positions. With the superior interpolation and denoising capabilities of the enhanced SRCNN and DnCNN, the SRDnNet could further improve channel estimation accuracy.

3) Comprehensive simulations validate the effectiveness of the SRDnNet concerning the effect of SNR and channel dimension. The results illustrate that the performance of the SRDnNet outperforms benchmarks without prior knowledge of direct and cascaded channels.

The rest of this paper is organized as follows. Section II presents the system model of RIS-aided multi-user MIMO-OFDM communication systems and the transmission protocol. Section III presents two conventional CSI estimation methods as benchmarks. Then the proposed SRDnNet is presented. Section IV provides simulation results of our proposed methods in RIS-
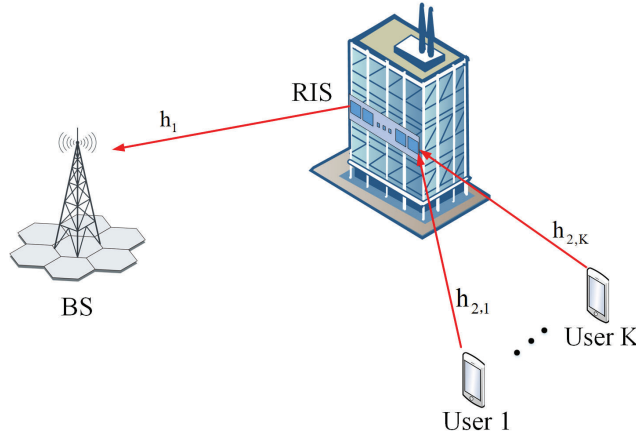
Fig. 1. The uplink of RIS-aided Multi-user MIMO-OFDM system.

aided wireless communication systems. Finally, Section V concludes this work.

In this paper, $\mathbf{I}_M$ denotes the identity matrix of size $M \times M$. $(\cdot)^{-1}$ is the matrix inverse. The $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and Hermitian transpose, respectively. $E(\cdot)$ is used to represent the statistical expectation operation. $\|\cdot\|_F$ represents the Frobenius norm of a matrix. $\odot$ denotes the Hadamard element-wise multiplication. $./$ denotes the right element-wise division, which divides each element of the left by the corresponding element of the right. $\mathbb{C}$ denotes the dimension of a complex variable. In addition, $\mathcal{CN}(\mu, \sigma^2)$ is the complex Gaussian distribution with μ-mean and variance σ^2.

II. System Model and Transmission Protocol

As shown in Fig. 1, we consider an uplink OFDM system, where the RIS serves one BS and K users. The BS is equipped with an N_{BS}-antenna array and the RIS comprising N_M reflecting elements serves the BS and K single-antenna users. Antenna arrays of the BS and reflecting elements of the RIS are arranged in a uniform linear array (ULA). The BS controls a controller to reconfigure the propagation environment. As shown in Fig. 1, $\mathbf{h}_1$, $\mathbf{h}_{2,k}$ and $\mathbf{d}_k$, $k \in \{1, 2, ..., K\}$ denote the channels of RIS-BS, U_k-RIS and U_k-BS, respectively. Since there are plenty of scatters distributed in the channel between the BS and users, the line of sight (LoS) path may not exist [37], we adopt Rayleigh fading channel model to formulate $\mathbf{d}_k \in \mathbb{C}^{N_M \times 1}$. For the channels $\mathbf{h}_1$ and $\mathbf{h}_{2,k}$, we model them as Rician fading channels.

A. System Model

We define $\mathbf{h}_1 \in \mathbb{C}^{N_{BS} \times N_M}$ as

$$\mathbf{h}_1 = \sqrt{\frac{\beta_1}{\beta_1 + 1}} \overline{\mathbf{h}}_1 + \sqrt{\frac{1}{\beta_1 + 1}} \widetilde{\mathbf{h}}_1, \tag{1}$$

where β_1 is the Rician factor of $\mathbf{h}_1$; $\overline{\mathbf{h}}_1 \in \mathbb{C}^{N_{BS} \times N_M}$ is the LoS component; $\widetilde{\mathbf{h}}_1 \in \mathbb{C}^{N_{BS} \times N_M}$ is the non-LoS (NLoS) component. The elements of $\widetilde{\mathbf{h}}_1$ follows the complex Gaussian distribution $\mathcal{CN}(0, 1)$. In the same way, we denote k-th user $\mathbf{h}_{2,k} \in \mathbb{C}^{N_M \times 1}$ as

$$\mathbf{h}_{2,k} = \sqrt{\frac{\beta_{2,k}}{\beta_{2,k} + 1}} \overline{\mathbf{h}}_{2,k} + \sqrt{\frac{1}{\beta_{2,k} + 1}} \widetilde{\mathbf{h}}_{2,k}, \tag{2}$$

where $\beta_{2,k}$ is the Rician factor of $\mathbf{h}_{2,k}$; $\overline{\mathbf{h}}_{2,k} \in \mathbb{C}^{N_M \times 1}$ is the LoS component; $\widetilde{\mathbf{h}}_{2,k} \in \mathbb{C}^{N_M \times 1}$ is the NLoS component. All elements of $\widetilde{\mathbf{h}}_{2,k}$ follow the complex Gaussian distribution $\mathcal{CN}(0, 1)$.

The LoS components can be expressed by the array response of a N_M-element ULA:

$$\mathrm{a}_{N_M}(\phi) = [1, e^{j2\pi \frac{d}{\lambda} \sin \phi}, ..., e^{j2\pi \frac{d}{\lambda}(N_M - 1) \sin \phi}], \tag{3}$$

where ϕ is the angle of departure (AoD) or angle of arrival (AoA) of a signal, d is the antenna spacing, and λ denotes the wavelength of the transmitted signal, respectively. Under this condition, $\overline{\mathbf{h}}_1$ is expressed as

$$\overline{\mathbf{h}}_1 = \mathrm{a}_{N_{BS}}^H(\phi_{\mathrm{AoD},1}) \mathrm{a}_M(\phi_{\mathrm{AoA},1}), \tag{4}$$

where $\phi_{\mathrm{AoD},1}$ is the AoD from the RIS reflecting elements, $\phi_{\mathrm{AoA},1}$ is the AoA to the BS. We express the LoS path $\overline{\mathbf{h}}_2$ in as

$$\overline{\mathbf{h}}_{2,k} = \mathrm{a}_{N_M}^H(\phi_{\mathrm{AoA},\mathrm{k}}), \tag{5}$$

where $\theta_{\mathrm{AoA},\mathrm{k}}$ is the AoA from the k-th user to the ULA at the RIS.

The usable bandwidth is evenly partitioned into N sub-carriers in the uplink RIS-aided multi-user MIMO-OFDM system. At the k-th user, the OFDM symbol $\mathrm{X}_k \triangleq [\mathrm{x}_{1,k}, \mathrm{x}_{2,k}, ..., \mathrm{x}_{N,k}]^T$ is first transformed into the time domain by an N-point inverse discrete Fourier transform (IDFT). A L_{cp}-length cyclic prefix which is longer than the delay spread of both cascaded and direct channels, is then added to transformed symbols. At the BS, we remove the CP of received signals transmitted from the k-th user and then perform N-point DFT to transform these signals into the frequency domain. Then, we denote the received signals at each antenna of the BS as

$$\mathbf{y}_k = \mathbf{X}_k \left(\sqrt{P_{L,r}^{(k)}} \sum_{m=1}^{N_M} \mathbf{H}_{1,m} \phi_{m,k} \odot \mathbf{H}_{2,m}^{(k)} + \sqrt{P_{L,d}^{(k)}} \mathbf{d}_k \right) + \mathbf{v}_k, \tag{6}$$

$$\frac{1}{P_{L,r}^{(k)}} = \frac{P_r}{P_t G_t G_r} = \frac{G d_x d_y \lambda^2}{64\pi^3} * \left| \sum_{m=1-\frac{M}{2}}^{\frac{M}{2}} \sum_{n=1-\frac{N}{2}}^{\frac{N}{2}} \frac{\sqrt{F_{combine}}}{r_{n,m}^t r_{n,m}^k} e^{\frac{-j2\pi(r_{n,m}^t + r_{n,m}^k)}{\lambda}} \right|^2, \tag{7}$$

$$\frac{1}{P_{L,r}^{(k)}} = \frac{P_r}{P_t G_t G_r} = \frac{G d_x d_y \lambda^2}{64\pi^3} * \left| M \frac{F(\theta^t, \varphi^t) F(\theta^k, \varphi^k)}{r^t r^k} e^{\frac{-j2\pi(r^t + r^k)}{\lambda}} \right|^2, \tag{8}$$

where N_M denotes the RIS elements, $\mathbf{y}_k \in \mathbb{C}^{N \times 1}$ and $\mathbf{X}_k = \mathrm{diag}(\mathbf{x}) \in \mathbb{C}^{N \times N}$ are the received and transmitted OFDM symbols. $\mathbf{H}_{1,m} \in \mathbb{C}^{N \times 1}$ is the channel frequency response (CFR) of the BS-RIS channel for the m-th element, $\mathbf{H}_{2,m}^{(k)} \in \mathbb{C}^{N \times 1}$ is the CFR of the channel between RIS and the k-th user for the m-th element, $\phi_{m,k} = e^{j\phi_{m,k}}$ is the phase shift introduced by the m-th element when the k-th user transmits signals, $\mathbf{d}_k \in \mathbb{C}^{N \times 1}$ is the CFR of the BS-user channel, and $\mathbf{v}_k \in \mathbb{C}^{N \times 1}$ is the additive white Gaussian noise (AWGN) vector with $\mathbf{v}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$, respectively.

To make the model more practical, we consider path loss [38]. $P_{L,r}^{(k)}$ is the path loss of the cascaded channel, and it is expressed as (7) presented at the top of the page. We consider the scenario in which the distance between the RIS and BS and that between users and the RIS is much larger than the size of the RIS. We can simplify the path loss of the cascaded channel as (8) at the top of the page. The detailed information in Eq.(7) and Eq.(8) can be found in [36].

We define the path loss of direct channel as $P_{L,d}^{(k)}$

$$\frac{1}{P_{L,d}^{(k)}} = \frac{\lambda^2}{4\pi^2 d_k^\gamma}, \tag{9}$$

where d_k denotes the distance between the BS and the k-th user and γ is the path loss exponent, respectively.

We use $\mathbf{H}_{m,k} \triangleq [H_{m,0}, H_{m,1}, ..., H_{m,N-1}]^T = \mathbf{H}_{2,m}^{(k)} \odot \mathbf{H}_{1,m}$ to express the equivalent CFR of the k-th user-RIS-BS channel associated with the m-th element, (6) can be rewritten as

$$\mathbf{y}_k = \mathbf{X}_k \left(\sqrt{P_{L,r}^{(k)}} \sum_{m=1}^{N_M} \phi_{m,k} \mathbf{H}_{m,k} + \sqrt{P_{L,d}^{(k)}} \mathbf{d}_k \right) + \mathbf{v}_k, \tag{10}$$

Moreover, by stacking $\mathbf{H}_{m,k}$ with $m = 1, ..., N_M$ into $\mathbf{H}_k = [H_{1,k}, H_{2,k}, ..., H_{N_M,k}]$ as the overall CFR of the cascaded channel, we can use a concise form to rewrite (10) as

$$\mathbf{y}_k = \mathbf{X}_k \underbrace{\left(\sqrt{P_{L,r}^{(k)}}\mathbf{H}_k\boldsymbol{\phi}_{\boldsymbol{k}} + \sqrt{P_{L,d}^{(k)}}\mathbf{d}_k\right)}_{\mathbf{h_k}} + \mathbf{v}_k, \tag{11}$$

where $\boldsymbol{\phi}_{\boldsymbol{k}} \triangleq [\phi_{1,k}, \phi_{2,k}, ..., \phi_{N_M,k}]^T$ denotes the phase-shift vector. Therefore, the main task is to estimate the superimposed CFR of the whole channel $\mathbf{h_k} = [H_{k,0}, H_{k,1}, ..., H_{k,N-1}]^T$. By stacking the received signals of each antenna at the BS, we can get a concise form

$$\mathbf{Y}_k = \mathbf{X}_k\mathbf{G}_k + \mathbf{v}_k, \tag{12}$$

where $\mathbf{Y}_k \triangleq [y_{k,1}, y_{k,2}, ..., y_{k,N_{BS}}] \in \mathbb{C}^{N \times N_{BS}}$ denotes the received signal transmitted by the k-th user at the BS. $\mathbf{G}_k \triangleq [\mathbf{h}_{k,1}, \mathbf{h}_{k,2}, ..., \mathbf{h}_{k,N_{BS}}] \in \mathbb{C}^{N \times N_{BS}}$ denotes the superimposed CFT of the whole channel between the k-th user and the BS. To practically estimate the cascaded and direct channel, we design a novel three-stage transmission protocol, which will be illustrated in the following section.

B. Transmission Protocol

To practically implement the RIS in wireless communication systems, we present a novel transmission protocol to estimate direct and cascaded channels separately. The reflecting elements embedded in the RIS are passive, so they cannot actively process the received signals to estimate the CSI. The RIS needs to know the timing when the user transmits signals through the uplink channel to prepare the proper phase shift $\phi_{m,k} = e^{j\phi_{m,k}}$ for communications. Otherwise, the BS cannot estimate the channel correctly because the propagation environment is not the same as the pre-designed, e.g., the pre-designed phase shift prepared for the m-th symbol meets the $(m + 1)$-th or the $(m - 1)$-th symbol.

The proposed transmission protocol with three sub-frames is illustrated in Fig. 2. The first sub-frame is to estimate the channel $\sqrt{P_{L,d}^{(k)}}\mathbf{d}_k$ between the k-th user and the BS, while the RIS is turned off. According to the feedback information of the first sub-frame, the RIS controller can know that the k-th user will establish a link through it, and then it is turned on for the upcoming signal from the user. This means the RIS does not need to be active all the time and it is an energy-efficient working mode. During the second sub-frame, to estimate the cascaded channel $\sqrt{P_{L,r}^{(k)}}\mathbf{H}_k\boldsymbol{\phi}_k + \sqrt{P_{L,d}^{(k)}}\mathbf{d}_k$, each pilot symbol is arranged with the pre-designed phase shift of the
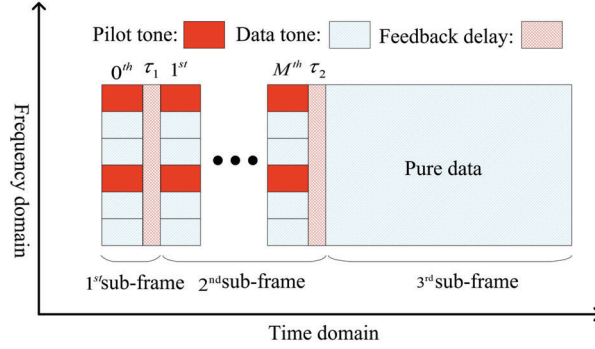
Fig. 2. The proposed three-stage transmission protocol for the RIS-aided wireless system.

RIS, both of which are known at the BS [16]. After removing the effect of the direct channel, the BS can estimate the channel $\sqrt{P_{L,r}^{(k)}}\mathbf{H}_k$ based on the received pilot symbols and phase shift. With the estimated cascaded channel and direct channel, we can design a reflection pattern of the RIS to maximize the system throughput, which will be done in the future. According to the feedback information of the second sub-frame, the RIS controller adjusts the phase shift of each element to reconfigure the propagation environment for data transmission in the third sub-frame.

The pilot overhead for channel estimation in this transmission protocol scales with the symbol number and pilot number N_p. With too few pilots, the accuracy of CSI cannot fulfill the requirement of phase shift design and precoding design, while too many pilots will lead to low SE and data rates. Thus, there is a trade-off between channel estimation accuracy and pilot overhead. To accurately estimate the channel with relatively low pilot overhead in RIS-aided multi-user MIMO-OFDM systems, in the next section, we propose a data-driven framework, i.e., SRDnNet, to tackle this problem.

III. SUPER-RESOLUTION-BASED CHANNEL ESTIMATION

As shown in Fig. 2, the first sub-frame is used to estimate the direct channel when the RIS is turned off. After performing the channel estimation, we can obtain the CSI of the direct link $\mathbf{d}_k$. A comb-type pilot arrangement is adopted to estimate the cascaded channel. For each user, the received OFDM symbol at pilot positions can be defined as

$$\mathbf{Y}_p = \mathbf{X}_p \mathbf{G}_p + \mathbf{v}_p, \tag{13}$$

where $\mathbf{G}_p \in \mathbb{C}^{N_P \times N_{BS}}$ denotes the superimposed CFRs of the whole channel. $\mathbf{X}_p = \mathrm{diag}(\mathbf{x}_p) \in \mathbb{C}^{N_P \times N_P}$ is the diagonal matrix of pilot sequence $\mathbf{x}_p$, and N_P is the number of pilot in each

OFDM symbol. $\mathbf{Y}_p \in \mathbb{C}^{N_P \times N_{BS}}$ and $\mathbf{v}_p \in \mathbb{C}^{N_P \times N_{BS}}$ denote received signals and the additive noise on pilot positions. With the pilot observations $\mathbf{Y}_p$, we can employ analytical methods, i.e., LS or LMMSE, both of which are derived as benchmarks, to acquire $\mathbf{G}$. After eliminating the effect of $\mathbf{d}$, we can acquire the cascaded channel $\mathbf{H}$. For simplicity, we only explain the estimation process of $\mathbf{G}$ in the following section.

A. Conventional Channel State Information Estimator

1) LS Estimator: The LS channel estimator finds $\hat{\mathbf{G}}$ in such a way that can minimize the cost function. On pilot tones, it is given by

$$\hat{\mathbf{G}}_p^{LS} = \arg \min_{\hat{\mathbf{G}}} \left\| \mathbf{y}_p - \mathbf{X}_p \hat{\mathbf{G}} \right\|^2 = \mathbf{X}_p^{-1} \mathbf{y}_p = \mathbf{G}_p + \mathbf{X}_p^{-1} \mathbf{v}_\rho, \tag{14}$$

where $\hat{\mathbf{G}}_p^{LS} \in \mathbb{C}^{N_P \times 1}$ is the estimated channel on positions via LS estimator in each OFDM symbol. To obtain the channel coefficients on other positions, we need to apply some interpolation methods, i.e., linear interpolation and cubic spline interpolation [39], [40]. This method does not need prior knowledge of the channel. Thus the LS estimator has been broadly used.

Then we employ the discrete Fourier transform (DFT)-based channel estimation technique to further improve the channel estimation performance [41]. We use $\hat{\mathbf{G}}[k]$ to denote the estimated channel at k-th subcarrier, the IDFT of the channel estimate is given by

$$\text{IDFT}\{\hat{\mathbf{G}}[k]\} = g[n] + z[n] \triangleq \hat{g}[n], n = 0, 1, ..., N - 1 \tag{15}$$

where $z[n]$ denotes the noise component in the time domain. Ignoring the coefficients $\hat{g}[n]$ that go beyond the maximum channel delay L which only contains the noise

$$\hat{g}_{\text{DFT}}[n] = g[n] + z[n], n = 0, 1, ..., L - 1 \tag{16}$$

and transform these L elements back to the frequency domain as

$$\hat{\mathbf{G}}_{\text{DFT}}[k] = \text{DFT}\{\hat{g}_{\text{DFT}}[n]\}. \tag{17}$$

The maximum delay L should be known before applying the DFT LS estimator.

2) MMSE Estimator: Unlike the LS estimator where the estimation does not need any prior knowledge of the channel. The MMSE estimator assumes that the prior probability density function (PDF) $p(\mathbf{G})$ of the channel has been known. Therefore, according to the Bayesian

method, the estimation accuracy can be further improved by using prior knowledge and denoted as

$$\hat{\mathbf{G}}_{\text{MMSE}} = E(\mathbf{G}|\mathbf{X}) = \int \mathbf{G}p(\mathbf{G}|\mathbf{X})d\mathbf{G}, \tag{18}$$

where $p(\mathbf{G}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{G})p(\mathbf{G})}{\int \mathbf{G}p(\mathbf{G}|\mathbf{X})d\mathbf{G}}$ is the posterior PDF of $\mathbf{G}$ given $\mathbf{X}$ and $p(\mathbf{X}|\mathbf{G})$ is the conditional PDF.

3) LMMSE Estimator: It is noted that the MMSE estimator is infeasible to implement since we need the PDF of the channel, which is usually impossible to obtain. The LMMSE estimator, which only needs the second-order statistic, is always used in practice. Considering the solution of the equation (14), LMMSE estimator can be obtained by using a filtering matrix $\mathbf{W}_{\text{LMMSE}}$, the result is given by

$$\hat{\mathbf{G}}_{\text{LMMSE}} = \mathbf{W}_{\text{LMMSE}}\hat{\mathbf{G}}_{\text{LS}}, \tag{19}$$

Then, the LMMSE estimator finds an estimate in terms of $\mathbf{W}$ in such a way that can minimize the mean square error(MSE) in the following equation

$$\varepsilon = E\{\left\|\mathbf{G} - \mathbf{W}_{\text{LMMSE}}\hat{\mathbf{G}}_{\text{LS}}\right\|_2^2\}, \tag{20}$$

and solve the problem in equation (20) yields

$$\mathbf{W}_{\text{LMMSE}} = \mathbf{R}_{\mathbf{G}\hat{\mathbf{G}}_{\text{LS}}}(\mathbf{R}_{\hat{\mathbf{G}}_{\text{LS}}\hat{\mathbf{G}}_{\text{LS}}} + \frac{\sigma_v^2}{\sigma_x^2}\mathbf{I})^{-1}, \tag{21}$$

where $\hat{\mathbf{G}}_{\text{LS}}$ is the LS estimation on pilot tones, $\mathbf{R}_{\mathbf{G}\mathbf{G}_{\text{LS}}} = E\{\mathbf{G}\hat{\mathbf{G}}_{\text{LS}}^{\text{H}}\}$ denotes the cross-correlation matrix between the true channel vector and LS channel estimation vector in the frequency domain, and $\mathbf{R}_{\mathbf{G}_{\text{LS}}\mathbf{G}_{\text{LS}}} = E\{\hat{\mathbf{G}}_{\text{LS}}\hat{\mathbf{G}}_{\text{LS}}^{\text{H}}\}$ is the autocorrelation matrix of the LS channel estimation vector. The elements of $\mathbf{R}_{\mathbf{G}\mathbf{G}_{\text{LS}}}$ and $\mathbf{R}_{\mathbf{G}_{\text{LS}}\mathbf{G}_{\text{LS}}}$ can be obtained by calculating the exponentially-decreasing power delay profile. The LMMSE estimator can find a better linear estimation of the channel matrix by utilizing correlation between different channels on sub-carriers in the frequency domain.

Note that when the channel is Gaussian-distributed, the optimal MMSE estimator and LMMSE are uniform [42]. Both can improve the channel estimation accuracy by using the second-order statistic. However, the cascaded channel in RIS-aided communication systems usually does not follow the Gaussian distribution due to the phase shift introduced by the RIS. Thus the optimal MMSE estimator is computationally costly with the calculation of posterior PDF and integration of real channel, which is always unknown in practice. Besides, the performance of the LMMSE

estimator has a gap in contrast to the optimal MMSE estimator when channels are not Gaussian-distributed [42]. To deal with the estimation of the above problem, in the following subsection, we will adopt data-driven methods to design an adequate model, i.e., the SRDnNet.

B. Image Super-Resolution and Channel Estimation

The objective of this work is to estimate RIS channels by using received pilots. The DL method's ideal input and output to deal with pilot-based channel estimation are received pilots and estimated channels, which do not need to do any operation on these pilots and result in the least time and computation consumption. In the experiment stage, this method is tough to implement since the functionality of the model is too complex and the training has never converged. In this case, an algorithm, the input and output pairs of which are estimated channels at pilot positions and channel matrix, and only need once division operation on received pilots and transmitted pilots known at both BS and user sides, becomes an alternative solution.

Recently, SRCNN, which can transform low-resolution images into high-resolution images, has been proposed to deal with the image SR problem [43]. Image SR problem can be formulated as

$$\hat{I}_h = F(I_l; \theta), \tag{22}$$

where $\hat{I}_h$ is the recovered high-resolution image, I_l is the low-resolution and noisy image, and F is the super-resolution technique parameterized by θ.

Note that the input of the Pilot-based channel estimation method in RIS-aided multi-user OFDM communication systems is the initial estimation of channels at pilot positions, and the output is the estimated channel. This method is comparable to the problem formulation (22). Thus we formulate the problem as an image super-resolution (SR) problem. We consider the estimation of channels at pilot positions $\hat{\mathbf{G}}_{p,2}$ the second sub-frame which involves M symbols

$$\hat{\mathbf{G}}_{p,2} = \mathbf{Y}_{p,2}./\mathbf{X}_{p,2}, \tag{23}$$

where $\hat{\mathbf{G}}_{p,2} \in \mathbb{C}^{2 \times Q_1 \times N_{BS}}$ is the low-resolution 2D input image (for the real and the imaginary value of the inputs), $Q_1 = N_P \times M$ and N_{BS} are the length and width of the 2D image, M is the number of symbols in the second sub-frame as shown in Figure 2. $\mathbf{Y}_{p,2} \in \mathbb{C}^{2 \times Q_1 \times N_{BS}}$ and $\mathbf{X}_{p,2} \in \mathbb{C}^{2 \times Q_1 \times N_{BS}}$ are received and transmitted pilots in the second sub-frame. The problem of obtaining an estimation of the whole channel $\hat{\mathbf{G}} \in \mathbb{C}^{2 \times Q_2 \times N_{BS}}$ can be written as

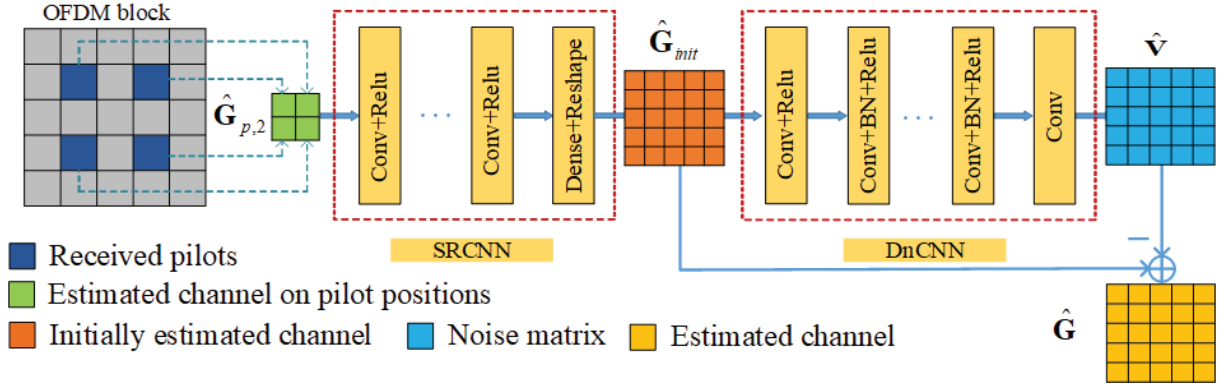$$\hat{\mathbf{G}} = F(\hat{\mathbf{G}}_{p,2}; \theta), \tag{24}$$

Fig. 3. The proposed SRDnNet with a concatenation of SRCNN and DnCNN for channel estimation in RIS-aided multi-user MIMO-OFDM systems.

where $Q_2 = N \times M$ and N_{BS} are the length and width of the output 2D image. It is noted that $\hat{\mathbf{G}}$ contains a cascaded channel, which is knotty to find a close expression of the MMSE estimator. In this case, we develop a data-driven model, named SRDnNet, to estimate the channel in the RIS-aided multi-user MIMO-OFDM system.

The size of the input $\hat{\mathbf{G}}_{p,2} \in \mathbb{C}^{2 \times Q_1 \times N_{BS}}$ can be huge because of the large number of antennae and symbols. The complexity order of the neural network is directly related to the input size. The dimension reduction techniques, i.e., principal component analysis (PCA), can reduce the dimensions of the dataset, computation cost and data transmission overhead [44]. PCA is a popular multivariate approach that transforms the high-dimensional dataset into the low-dimensional dataset by using only the first several principal components while extracting the most important information from the dataset. We summarize the PCA method in algorithm 1, where P denotes $2 \times Q_1 \times N_{BS}$, $\dot{\mathbf{D}}$ denotes $\{\dot{\mathbf{G}}_{p,2}^{(1)}, ..., \dot{\mathbf{G}}_{p,2}^{(N_s)}\}^T \in \mathbb{C}^{N_s \times P}$, n is the number of dimensions of the output low-dimensional dataset. The output is the compressed dataset $\mathbf{D_n}$.

C. The Proposed SR-based Channel Estimation Scheme

Based on the image SR model, the ChannelNet [34] has been developed for channel estimation but with the following limits: 1) The choice of interpolation methods highly affects the final performance of this network; 2) After the interpolation at the very beginning, the input size becomes larger than the raw input $\hat{\mathbf{G}}_{p,2}$, and significantly increase the computation complexity; 3) The ChannelNet is not an end-to-end model and is trained in two stages.

Algorithm 1 The principal component analysis (PCA) algorithm

Input: Training dataset $\mathbf{D} = \{\hat{\mathbf{G}}_{\mathbf{p,2}}^{(\mathbf{1})}, ..., \hat{\mathbf{G}}_{\mathbf{p,2}}^{(\mathbf{N_s})}\}^{\mathbf{T}} \in \mathbb{C}^{\mathbf{N_s} \times \mathbf{P}}$; n the number of dimensions of the low-dimensional dataset.

Step:

1: Calculate mean of all samples $\bar{\mathbf{G}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{\mathbf{G}}_{p,2}^{(i)} \in \mathbb{C}^{1 \times P}$.

2: Centralize all samples $\dot{\mathbf{G}}_{p,2}^{(i)} = \hat{\mathbf{G}}_{p,2}^{(i)} - \bar{\mathbf{G}} \in \mathbb{C}^{1 \times P}$.

3: Calculate covariance matrix $\mathbf{S} = \dot{\mathbf{D}}\dot{\mathbf{D}}^{\mathrm{T}} \in \mathbb{C}^{P \times P}$.

4: Calculate eigenvalues and eigenvectors of $\mathbf{S}$, arrange them in descending order.

5: Select n eigenvectors with n largest eigenvalues to construct $\mathbf{U_n} \in \mathbb{C}^{\mathbf{P} \times \mathbf{n}}$.

Output: The compressed dataset $\mathbf{D_n} = \mathbf{D}\mathbf{U_n} \in \mathbb{C}^{\mathbf{N_s} \times \mathbf{n}}$.

To avoid these drawbacks of ChannelNet, we propose an end-to-end model named SRDnNet, which is illustrated in Fig. 3. The SRDnNet is a concatenation of SRCNN and DnCNN. To learn the features of complex input, we adopt two channels at the input layer. The first layer of SRCNN is the feature extraction and representation layer and can extract features of low-resolution figures and represent them in high-dimension. The second layer can non-linearly map each feature map to another high-dimension feature map representing a high-resolution patch. The third layer can aggregate these patches to generate a high-resolution representation that has a different size compared with the ground truth channel without the interpolation at the very beginning. We then use a dense layer and a reshape layer to generate an initial estimation $\hat{\mathbf{G}}_{init}$ as the same size as the ground truth channel but not an up-sampling layer as usual. Since the up-sampling layer is not learnable, the dense layer can be trained and find a learnable method to aggregate all features to estimate the ground truth. Channels in different transmitted symbols are not independent because the data transmission is within a coherence time. Each pixel in the output of the dense layer is related to the high-dimension representation of all estimated channels at pilot positions. In this case, the SRDnNet can outperform the ChannelNet. In a word, the SRCNN can extract high-dimensional features from low-resolution image $\hat{\mathbf{G}}_{p,2}$ and aggregates these features to recover a high-resolution image $\hat{\mathbf{G}}_{init}$. It is noted that the ChannelNet is a denoising model since the interpolation at the very beginning is not trainable. Meanwhile, the SRCNN has the capability to find a proper interpolation method beyond linear or polynomial constraints to improve the interpolation accuracy. We adopt the DnCNN, an efficient denoising

TABLE I

SETTINGS OF SRDNNET

Input:	
estimated channels at pilot positions.	

SRCNN:

Layers	Filter Size
1	$64 \times (9 \times 9 \times 2)$
2	$32 \times (1 \times 1 \times 64)$
3	$2 \times (5 \times 5 \times 32)$

DnCNN:

Layers	Filter Size
1	$64 \times (3 \times 3 \times 2)$
2∼19	$64 \times (3 \times 3 \times 64)$
20	$2 \times (3 \times 3 \times 64)$

Output:	
Estimated whole channel.	

model verified by academia, cascaded with SRCNN to enhance performance further. The DnCNN can learn features of the coarse channel matrix generated by SRCNN and the additive noise, its element-wise subtraction before the output layer could generate an accurate denoised estimation $\hat{\mathbf{G}}$. The hyperparameters of SRDnNet are listed in Table I. All the hyperparameters follow the default settings in [43] and [45].

1) Offline Training: We define

$$(\mathbf{X}, \mathbf{G}) = \{(\mathbf{X}^{(1)}, \mathbf{G}^{(1)}), (\mathbf{X}^{(2)}, \mathbf{G}^{(2)}), ..., (\mathbf{X}^{(N_s)}, \mathbf{G}^{(N_s)})\} \tag{25}$$

as training examples, where $\mathbf{X}^{(i)}$ and $\mathbf{G}^{(i)}$ are the received pilots and real channels in i-th frame, respectively. To obtain the initially estimated channels at pilot positions as defined in the image super-resolution (24), we sent $\mathbf{X}$ to LS estimator in (14) to generate estimated channels at pilot positions $\hat{\mathbf{X}}$ and training data set is

$$(\hat{\mathbf{X}}, \mathbf{G}) = \{(\hat{\mathbf{X}}^{(1)}, \mathbf{G}^{(1)}), (\hat{\mathbf{X}}^{(2)}, \mathbf{G}^{(2)}), ..., (\hat{\mathbf{X}}^{(N_s)}, \mathbf{G}^{(N_s)})\}, \tag{26}$$

where $\hat{\mathbf{X}}^{(\mathbf{i})}$ and $\mathbf{G}^{(\mathbf{i})}$ are the i-th training sample pair. We denote $f(\cdot)$ as the function of SRDnNet and the output is

$$\mathbf{O} = f(\mathbf{I}, \mathbf{\Theta}), \tag{27}$$

where $\mathbf{I}$ and $\mathbf{O}$ are the input and output of SRDnNet, Θ is the trainable parameters of SRDnNet. We adopt the MSE loss function to train the model, which is formulated as

$$E_{mse} = \mathrm{E}(\left\| f(\hat{\mathbf{X}}, \Theta) - \mathbf{G} \right\|_F^2), \tag{28}$$

where $\mathbf{G}$ denotes the estimation and the ground truth of the channel. Formally, the average MSE between the estimated channel and the ground truth is

$$\ell(\Theta) = \frac{1}{2N_s} \sum_{i=1}^{N_s} \left\| f(\hat{\mathbf{X}}^{(i)}, \Theta) - \mathbf{H}^{(i)} \right\|_F^2, \tag{29}$$

where $\ell(\cdot)$ is the loss function. Based on this, SRDnNet can update its trainable parameters by the backpropagation (BP) algorithm to get a well-trained model parameterized by $f(\mathbf{I}, \Theta^*)$, Θ^* is well-trained model parameters.

2) Online prediction: We denote

$$(\widehat{\mathbf{X}}, \mathbf{H}) = \{(\widehat{\mathbf{X}}^{(1)}, \mathbf{G}^{(1)}), (\widehat{\mathbf{X}}^{(2)}, \mathbf{G}^{(2)}), ..., (\widehat{\mathbf{X}}^{(N_t)}, \mathbf{G}^{(N_t)})\} \tag{30}$$

as test data set, where $\widehat{\mathbf{X}}^{(i)}$ and $\mathbf{G}^{(i)}$ are estimated channels at pilot positions and ground truth in i-th test sample, respectively. We send $\widehat{\mathbf{X}}$ to $f(\mathbf{I}, \Theta^*)$, and the estimated channel is

$$\hat{\mathbf{G}}_{SRDnNet} = f(\widehat{\mathbf{X}}, \Theta^*). \tag{31}$$

We summarize the SRDnNet-based channel estimation approach in algorithm 2. i denotes the loop indicator. $epoch$ is the maximum loop number.

Compared with some existing channel estimation methods [26], [31], [46] deriving estimated channels as inputs, our proposed framework, SRDnNet, can accept rawer input with only one division operation on received pilots and transmitted pilots with less time and computation consumption. Based on the proposed SRDnNet with the capability of non-linear interpolation and denoising, we can estimate the cascaded channel, which does not follow Gaussian distribution without any prior knowledge of the channel. Inheriting the superior feature extraction and denoising ability from SRCNN and DnCNN, SRDnNet can accurately estimate channels which will be validated in the following section.

IV. NUMERICAL RESULTS

In the simulations, the system consists of an N_{BS}-antenna BS, an N_M-element RIS and K single-antenna users, as shown in Fig. 1. Without the further declaration, the RIS can continuously adjust its phase shift, and the default settings are $N_{BS} = 16$, $N_M = 144$ and $K = 5$. A practical

---

Algorithm 2 The proposed Channel Estimation with SRDnNet

Initialization: Initialize trainable parameters by HeUniform initializer

Offline Training:

1: **Input:** Training data set $(\hat{\mathbf{X}}, \mathbf{G})$

2:     **While** $i \leq epoch$

3:         Update trainable parameters Θ by BP algorithm

4:         $i = i + 1$

5: **Output:** Well-trained model $f(\mathbf{I}, \Theta^*)$

Online Prediction:

6: **Input:** Test data set $(\widehat{\mathbf{X}}, \mathbf{G})$

7: Estimate channel by using well-train SRDnNet

8: **Output:** Estimated channel $\hat{\mathbf{G}}_{SRDnNet}$.

---

path loss model illustrated in (7) has been considered in the simulation to better formulate the channels for our considered system. We refer to the parameter Settings for small RISs in the paper by Tang *et al.* [38]. In addition, Rician fading for the cascaded channel is formulated in (1) and (2), and a Rayleigh fading channel model is modeled for the direct channel. Note that (1) and (2) are general models. By adjusting their Rician factors, this general model can characterize the LoS channel model when Rician factors are close to the infinite and Rayleigh channel model when Rician factors are close to zero. Different users are randomly allocated to the area near the RIS within a range from 3.5m to 15m. The detailed settings of the simulations are summarized in Table II. Additionally, the second transmission sub-frame consists of 16 OFDM symbols, each of which consists of $N = 64$ sub-carriers appended with a CP with the length of $L_{cp} = 8$. The maximum delay spread of both direct and cascaded channels is $L = 5$. We set γ as 3.5 because of the long distance and the variety of obstructions in the considered propagation environment.

To evaluate the performance of the SRDnNet, we derive the following estimation methods as benchmarks: the LMMSE channel estimator, the LS, the DFT LS channel estimator and the ChannelNet [34]. The proposed neural network SRDnNet adopts the Adam optimizer to update the network parameters with a learning rate of 0.001 and a batch size of 100 samples. In total, 10,000 channel realizations for different users are collected at the BS. The training, validation and testing sample numbers are 6000, 2000, and 2000 respectively. In addition, the

TABLE II

SIMULATION SETTINGS

Description	Settings
The Nearest Distance between users and BS	$d = 102m$
Distance of RIS-BS	$r^r = 100m$
The Nearest Distance of User-RIS	$r^t = 3.5m$
Rician factor	$\beta_1 = 10$
Rician factor	$\beta_2 = 10$
Width of Reflecting elements	$d_x = 0.012m$
Length of Reflecting elements	$d_y = 0.012m$
Antenna gain of Reflecting elements	$G = 8dB$
Antenna gain of Transmitter	$G_t = 21dB$
Antenna gain of Receiver	$G_r = 21dB$

SNR is defined as $\mathrm{SNR} = \frac{E_b}{N_0}$, in which E_b is the energy of each user data bit and N_0 is the noise spectral density. In the following, we will illustrate the performance of the SRDnNet in terms of normalized MSE (NMSE) defined as $\mathrm{NMSE} = \frac{E(\|\mathbf{G} - \hat{\mathbf{G}}_{SRDnNet}\|_F^2)}{E(\|\mathbf{G}\|_F^2)}$, where $\hat{\mathbf{G}}_{SRDnNet}$ and $\mathbf{G}$ denote the channel estimated by a well-trained SRDnNet and the ground truth channel, respectively.

A. NMSE versus SNR

As shown in Fig. 2, we train two models for the first and second sub-frame to obtain the direct and whole channels, which is the superposition of the direct and cascaded channels. Then we calculate the cascaded channel by estimating the direct and the whole channel. Simulation results under different SNRs for the whole channel and direct channel are presented in Fig. 4 and Fig. 5a, Fig. 5b shows the NMSEs after calculation of the cascaded channel. The cascaded channel is not a Gaussian-distributed channel. Thus it is not easy to find the posterior PDF to design a closed-form expression for the MMSE estimator. Only the LMMSE estimator (21) can be derived to obtain an estimated channel based on the exponentially-decreasing power delay profile. As shown in Fig. 4, Fig. 5a, and Fig. 5b, the performance of all estimators decreases with the enhancement of SNR since the impact of noise can be mitigated by higher transmit power. In the following, we will analyze the simulation results of conventional methods and DL-based methods.
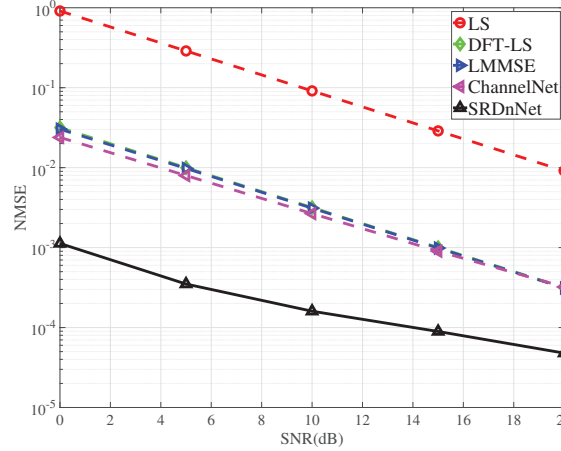
Fig. 4. NMSE performance for the whole channel with different SNRs under 32 pilots, $N_{BS} = 16$, $M = 144$ and $K = 5$.

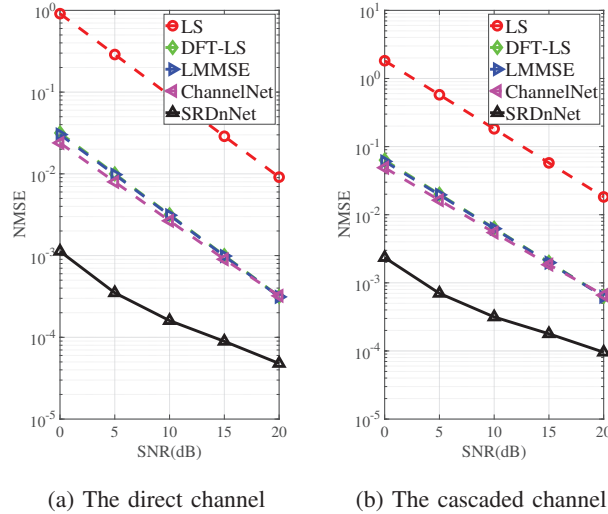(a) The direct channel   (b) The cascaded channel

Fig. 5. NMSE performance for the direct and cascaded channel with different SNRs under 32 pilots.

First, we investigate the estimation performance of the LS, DFT LS, and LMMSE estimator in terms of NMSE. It can be seen that LMMSE significantly outperforms the LS method. The reason is that the LMMSE can utilize the second-order statistical knowledge of the channel, and the LS estimator could only estimate unknown but deterministic constants without prior knowledge. In contrast to the LS, the DFT LS estimator can reduce the noise effect by removing channel coefficients beyond the maximum channel length, which only contains noise, to improve the channel estimation performance. It is shown that the DFT LS estimator significantly outperforms the LS method but is still worse than the LMMSE estimator.
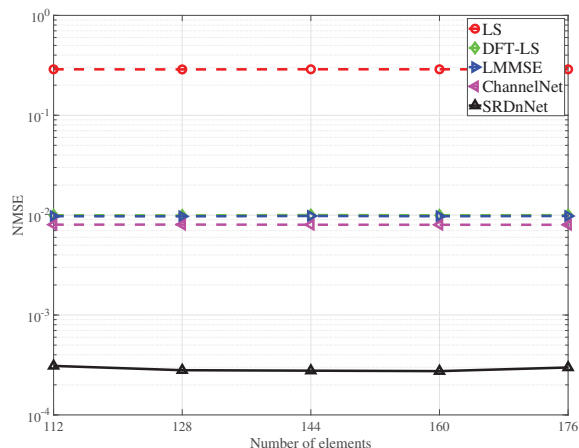
Fig. 6. NMSE performance for the whole channel under SNR = 5dB, 32 pilots, $N_{BS} = 16$ and $K = 5$.

Second, we will discuss the performance of DL-based methods, i.e., ChannelNet and SRDnNet. It is noticed that the ChannelNet is comparable to the LMMSE and DFT LS estimator since the ChannelNet is a denoising model for the LS estimator as the same as these two estimators. Another reason is that the ChannelNet is not trained in an end-to-end manner. In contrast, the SRDnNet can learn a non-linear interpolation method to improve channel estimation accuracy and perform better. For example, in Fig. 4, the SRDnNet achieves a performance gain of 13 dB at 10 dB SNR compared with the LMMSE estimator and 12 dB compared with the ChannelNet. The reason is the SRDnNet has the capability of non-linear mapping a low-resolution image into a high-resolution image and learning the distinguishable features of the additive noise to improve the channel estimation accuracy.

Simulation results for the direct channel are shown in Fig. 5a, and with the estimated direct and whole channel, we can calculate and obtain the cascaded channel as shown in Fig. 5b. Fig. 5a shows that SRDnNet can achieve the best channel estimation performance compared with benchmarks. In Fig. 5b, it is noted that the estimation error of the direct channel and whole channel accumulate in the calculation of the cascaded channel. For example, in Fig. 5a and Fig. 4, the channel estimation performance gaps in terms of NMSE for SRDnNet and LMMSE at 10 dB SNR are -38dB and -38.1dB, but in Fig. 5b the NMSE performance of the cascaded channel is -35dB. In comparison, the performance gap between the SRDnNet and LMMSE remains 13dB, which shows the optimality of SRDnNet.

B. NMSE versus Number of Element, Antenna and Pilot

Note that the number of elements, antennas, and pilots in each OFDM symbol is also an important parameter that can affect channel estimation accuracy. To further verify the scalability of SRDnNet in different communication scenarios, simulations are conducted to evaluate the estimation performance with different system settings. For simplicity, we only demonstrate the NMSE performance on the estimation of the whole channel in the following, the accumulation of channel estimation error of the whole channel and direct channel on the cascaded channel remains similar to Fig. 4 and Fig. 5. As shown in Fig. 6, the NMSE performance is not sensitive to the change in the number of elements denoted by M. The reason is that in our system model (6), the dimensionality of channels to be estimated is a constant no matter M changes; thus, the size of input pairs (26) remains the same. Based on this, the NMSE performance of these considered methods just slightly changes as the increase of M. In addition, our proposed SRDnNet still achieves the best performance because of its superiority in non-linear interpolation and denoising. The NMSE performance gap between SRDnNet and ChannelNet is 13.6dB under 128 elements.

Besides M, the number of antennae denoted by N_{BS}, which determines channel dimension, also impacts the channel estimation performance. Therefore we investigate the NMSE performance with the increase of N_{BS} and demonstrate it in Fig. 7. It is noted that NMSEs of analytical channel estimators, i.e., the LS, DFT LS, and LMMSE estimator, do not change with the increase of N_{BS}, while NMSEs of DL-based estimators, i.e., the ChannelNet and SRDnNet decrease. This is because, in the simulation, we separately estimate the channels of each antenna and average their performance. In this case, the received signal power of each antenna does not change, and the NMSEs remain almost constant. For the DL-based estimators, and the input data contains received signals from all antennas, the input data size and received signal power become larger with the increase of N_{BS}. Thus the NMSEs for DL-based estimators decrease when adding more antennae to the BS. The SRDnNet works with the best-performing in terms of NMSE. Moreover, the pilot number can directly affect the dimension of input data size for DL-based estimators and consequently affect the channel estimation performance of DL-based estimators. As illustrated in Fig. 8, the performance of all considered algorithms decreases as the number of pilots increases as expected. This is because, for those analytical estimators, i.e., the LS, DFT LS and LMMSE, the interpolation accuracy will increase with more received pilots. For DL-based estimators, the ChannelNet and SRDnNet, convolutional neural networks can exploit more
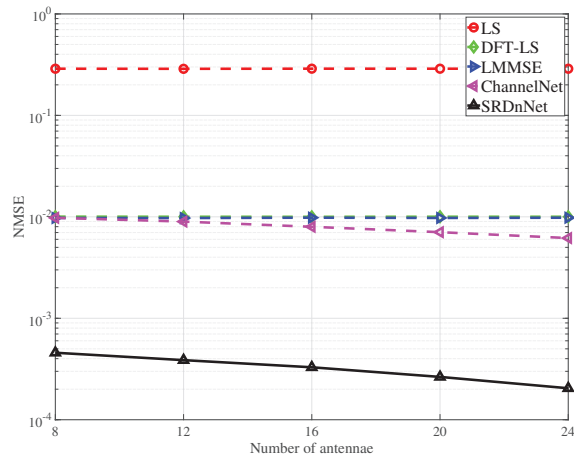
Fig. 7. NMSE performance for the whole channel under SNR = 5dB, 32 pilots, $M = 144$ and $K = 5$.
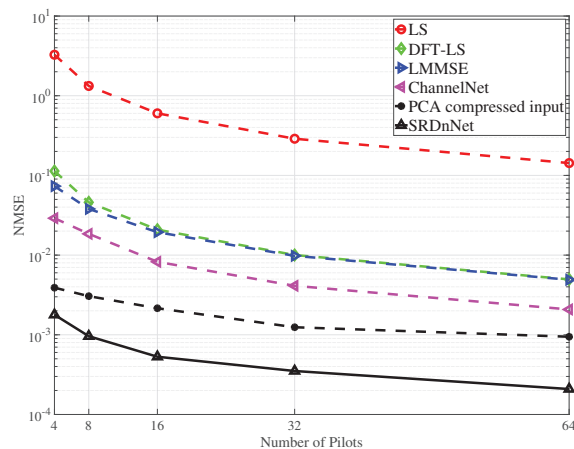

Fig. 8. NMSE performance for the whole channel with different pilots under SNR = 5dB, $N_{BS} = 16$, $M = 144$ and $K = 5$.

spatial features from larger input data and improve their performance. The simulation results of the line PCA compressed data adopt PCA to generate the compressed input of the SRDnNet and the compression ratio is fifty percent leading to a significant reduction in complexity order. However, the performance of the PCA compressed data is worse than the SRDnNet. Because the PCA extracts low-dimensional data as well as eliminates part of the useful information in the input dataset. Fig. 8 shows the superiority of SRDnNet in non-linear interpolation and denoising. For example, with only four pilots in each OFDM symbol, SRDnNet can still achieve NMSE performance gain of -27.5dB, -15.4dB for ChannelNet and -11.3dB for LMMSE. Moreover, this

simulation result shows that our proposed DL-based estimator can still achieve high channel estimation accuracy even with a few pilots under a relatively low SNR.

C. Complexity Analysis

We further investigate the computation complexity of the SRDnNet. Table I summarizes the parameters of each convolutional layer. The complexity of the convolutional layer is

$$\mathcal{C}_{CL} = \mathcal{O}(W_x W_y F^2 N_I N_O), \tag{32}$$

where W_x and W_y are the length and width of output feature maps, F is the length of the side of filters, and N_I and N_O are the number of input and output feature maps, respectively. Thus the complexity of SRCNN (without the dense layer) is

$$\mathcal{C}_{SR} = \mathcal{O}(2 \cdot N_{BS} \cdot N_P \cdot M(9^2 \cdot 2 \cdot 64 + 1^2 \cdot 64 \cdot 32 + 5^2 \cdot 32 \cdot 2))$$
$$= \mathcal{O}(128 \cdot 219 \cdot N_{BS} \cdot N_P \cdot M), \tag{33}$$

where N_P is the pilot number in each OFDM symbol, and M is the number of symbols in the second sub-frame. And the complexity of DnCNN and the output layer becomes approximately

$$\mathcal{C}_{Dn} = \mathcal{O}(2 \cdot N_{BS} \cdot N \cdot M(3^2 \cdot 2 \cdot 64 \cdot 2 + 3^2 \cdot 64 \cdot 64 \cdot 18))$$
$$\approx \mathcal{O}((6 \cdot 192)^2 \cdot N_{BS} \cdot N \cdot M), \tag{34}$$

The time complexity of the fully connected layer is

$$\mathcal{C}_{FC} = \mathcal{O}(I_x I_y N_{FC}), \tag{35}$$

where I_x and I_y are the input size, N_{FC} is the number of units of the fully connected layer, respectively. Therefore the complexity of the fully connected layers is approximately

$$\mathcal{C}_{FC} \approx \mathcal{O}(2 \cdot N_{BS} \cdot N \cdot M \cdot 1024) \quad = \mathcal{O}(2048 \cdot N_{BS} \cdot N \cdot M), \tag{36}$$

The total time complexity of the SRDnNet is $\mathcal{C} = \mathcal{C}_{SR} + \mathcal{C}_{Dn} + \mathcal{C}_{FC}$, which approximately is

$$\mathcal{C} = \mathcal{C}_{SR} + \mathcal{C}_{Dn} + \mathcal{C}_{FC}$$
$$= \mathcal{O}(((6 \cdot 192)^2 + 2048) \cdot N_{BS} \cdot N \cdot M + 128 \cdot 219 \cdot N_{BS} \cdot N_P \cdot M). \tag{37}$$

The bicubic interpolation method can be formulated as a convolutional layer. The time complexity of the bicubic interpolation is

$$\mathcal{C}_B = \mathcal{O}(2 \cdot N_{BS} \cdot N_P \cdot M(4^2 \cdot 2 \cdot 64))$$
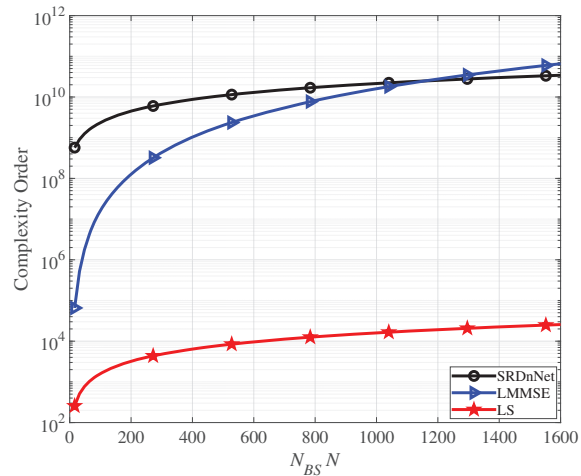$$= \mathcal{O}(4096 \cdot N_{BS} \cdot N_P \cdot M). \tag{38}$$

Fig. 9. Complexity order for the LMMSE, LS and the SDRnNet with $N_P = 32$ and $N_S = 64$.

The total time complexity of the ChannelNet is

$$\mathcal{C}_{CN} = \mathcal{C}_{SR} + \mathcal{C}_{Dn} + \mathcal{C}_B$$
$$= \mathcal{O}(((6 \cdot 192)^2 + 4096) \cdot N_{BS} \cdot N \cdot M + 128 \cdot 219 \cdot N_{BS} \cdot N_P \cdot M). \tag{39}$$

It is obvious that $C_B > C_{FC}$. Thus the total time complexity of the SRDnNet is less than that of the ChannelNet.

The time complexity of the LMMSE and LS estimators are $\mathcal{O}(N_{BS} \cdot (N \cdot M)^3)$ and $\mathcal{O}(N_{BS} \cdot N \cdot M)$ [47]. Fig.9 shows the time complexity of the proposed SRDnNet, LMMSE and LS with respect to $N_{BS} \cdot N$. It is shown that the SRDnNet has higher complexity than the LS estimator. As the number of channel dimensions, i.e., $N_{BS} \cdot N$ increases, the complexity of the LMMSE estimator becomes close to that of the SRDnNet, and it becomes larger approximately after $N_{BS} \cdot N \geq 1200$. While it seems that the complexity of the SRDnNet is comparable to the LMMSE estimator, the SRDnNet can run efficiently on GPUs, which can significantly reduce the testing time [48]. However, implementing the LMMSE and LS estimators is not straightforwardly supported by graphics processing units (GPUs) and requires an application-specific processor. Although the proposed SRDnNet involves higher computational complexity than the LS and LMMSE estimators, the required running time of the SRDnNet can be reduced by the parallel computing of a GPU. This conclusion can be verified in Table III, where each method runs on a desktop computer with an i9-9940X 3.30 GHz central processing unit (CPU) and an Nvidia GeForce RTX 2080Ti GPU.

TABLE III

COMPUTATION TIME OF DIFFERENT ESTIMATION METHODS

Method	Prediction Time
LS	$4.01 \times 10^{-4} s$
LMMSE	$2.12 \times 10^{-2} s$
SRDnNet	$1.61 \times 10^{-2} s$

V. Conclusion

This paper proposed a deep learning (DL)-based method to accurately perform channel estimation in a reconfigurable intelligent surface (RIS)-aided multi-user multiple-input multiple-output (MIMO)-orthogonal frequency division multiplexing (OFDM) system. We formulated channel estimation as an image super-resolution (SR) problem and proposed an image super-resolution network, named SRDnNet to recover the channel matrix from the coarse estimation of channels. By inheriting the abilities of the revised super-resolution convolutional neural network (SRCNN) and the denoising convolutional neural network (DnCNN) in feature extraction and denoising, the SRDnNet can further improve the channel estimation performance. We evaluated the performance of the proposed approach via different signal-to-noise ratios (SNRs), number of elements, antennae, and pilots. Simulation results demonstrated that the proposed SRDnNet outperforms analytical channel estimators and DL-based methods with more than 10 dB performance gain in terms of the normalized mean square error (NMSE). We showed that at least 8 pilots were required for reliable channel estimation performance under 5dB SNR. Furthermore, the proposed SRDnNet can reduce the size of input data and the computation cost for channel estimation in complicated scenarios.

References

[1] W. Shen, Z. Qin, and A. Nallanathan, "Deep Learning Enabled Channel Estimation for RIS-Aided Wireless Systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) 2022*, [To appear].

[2] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[4] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental Green Tradeoffs: Progresses, Challenges, and Impacts on 5G Networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 33–56, First quarter 2017.

[5] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A New Wireless Communication Paradigm through Software-Controlled Metasurfaces, year=2018," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep.

[6] M. Di Renzo, A. Zappone, M. Debbah, M. S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and The Road Ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[7] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective Multiple Antenna Technologies for Beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug 2020.

[8] Q. Wu and R. Zhang, "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[9] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[10] S. Gong, C. Xing, S. Wang, L. Zhao, and J. An, "Throughput Maximization for Intelligent Reflecting Surface Aided MIMO WPCNs With Different DL/UL Reflection Patterns," *IEEE Trans. Signal Process.*, vol. 69, pp. 2706–2724, Apr. 2021.

[11] Z. Chu, W. Hao, P. Xiao, and J. Shi, "Intelligent Reflecting Surface Aided Multi-Antenna Secure Transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 108–112, Jan. 2020.

[12] X. Guan, Q. Wu, and R. Zhang, "Intelligent Reflecting Surface Assisted Secrecy Communication: Is Artificial Noise Helpful or Not?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 778–782, June 2020.

[13] D. Mishra and H. Johansson, "Channel Estimation and Low-complexity Beamforming Design for Passive Intelligent Surface Assisted MISO Wireless Energy Transfer," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brighton, UK, May. 2019, pp. 4659–4663.

[14] T. L. Jensen and E. De Carvalho, "An Optimal Channel Estimation Scheme for Intelligent Reflecting Surfaces Based on a Minimum Variance Unbiased Estimator," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, May. 2020, pp. 5000–5004.

[15] Z. He and X. Yuan, "Cascaded Channel Estimation for Large Intelligent Metasurface Assisted Massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[16] B. Zheng and R. Zhang, "Intelligent Reflecting Surface-Enhanced OFDM: Channel Estimation and Reflection Optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[17] C. You, B. Zheng, and R. Zhang, "Channel Estimation and Passive Beamforming for Intelligent Reflecting Surface: Discrete Phase Shift and Progressive Refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, Nov. 2020.

[18] N. K. Kundu and M. R. Mckay, "Large Intelligent Surfaces With Channel Estimation Overhead: Achievable Rate and Optimal Configuration," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 986–990, May 2021.

[19] A. Zappone, M. Di Renzo, X. Xi, and M. Debbah, "On the Optimal Number of Reflecting Elements for Reconfigurable Intelligent Surfaces," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 464–468, March 2021.

[20] B. Zheng, C. You, and R. Zhang, "Intelligent Reflecting Surface Assisted Multi-User OFDMA: Channel Estimation and Training Design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec 2020.

[21] N. K. Kundu, Z. Li, J. Rao, S. Shen, M. R. McKay, and R. Murch, "Optimal Grouping Strategy for Reconfigurable Intelligent Surface Assisted Wireless Communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 1082–1086, May 2022.

[22] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep Learning in Physical Layer Communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 93–99, April 2019.

[23] K. Feng, Q. Wang, X. Li, and C. Wen, "Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, May. 2020.

[24] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Secure Wireless Communications," *IEEE Trans. Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

[25] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, 2021.

[26] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel Estimation for Cell-Free mmWave Massive MIMO Through Deep Learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10 325–10 329, Oct. 2019.

[27] M. B. Mashhadi and D. Gündüz, "Pruning the Pilots: Deep Learning-Based Pilot Design and Channel Estimation for MIMO-OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6315–6328, Oct 2021.

[28] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, "Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 989–1000, Sep. 2019.

[29] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, 2021, doi=10.1109/TWC.2021.3100148.

[30] N. K. Kundu and M. R. McKay, "Channel Estimation for Reconfigurable Intelligent Surface Aided MISO Communications: From LMMSE to Deep Learning Solutions," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 471–487, 2021.

[31] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep Denoising Neural Network Assisted Compressive Channel Estimation for mmWave Intelligent Reflecting Surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug 2020.

[32] Y. Jin, J. Zhang, X. Zhang, H. Xiao, B. Ai, and D. W. K. Ng, "Channel Estimation for Semi-Passive Reconfigurable Intelligent Surfaces With Enhanced Deep Residual Networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11 083–11 088, Oct 2021.

[33] D. Ouyang, Y. Li, and Z. Wang, "Channel Estimation for Underwater Acoustic OFDM Communications: An Image Super-Resolution Approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2021, pp. 1–6.

[34] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep Learning-Based Channel Estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, April. 2019.

[35] L. Li, H. Chen, H. Chang, and L. Liu, "Deep Residual Learning Meets OFDM Channel Estimation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 615–618, May. 2020.

[36] W. Shen, Z. Qin, and A. Nallanathan, "Federated learning enabled channel estimation for ris-aided multi-user wireless systems," in *Proc. IEEE Int. Conf. Commun. Workshops, (ICC Workshops) 2022*, May 2022, pp. 812–817.

[37] Y. Han, W. Tang, S. Jin, C. K. Wen, and X. Ma, "Large Intelligent Surface-Assisted Wireless Communication Exploiting Statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

[38] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless Communications With Reconfigurable Intelligent Surface: Path Loss Modeling and Experimental Measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.

[39] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[40] M.-H. Hsieh and C.-H. Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Trans. Consumer Electron.*, vol. 44, no. 1, pp. 217–225, Feb 1998.

[41] Y. Cho, J. Kim, W. Yang, and C. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. Wiley Publishing, 2010, pp.190-195.

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing, volume I: Estimation Theory*. Prentice Hall, 1993.

[43] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[44] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.

[45] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, July. 2017.

[46] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep Residual Learning for Channel Estimation in Intelligent Reflecting Surface-Assisted Multi-User Communications," *arXiv preprint arXiv:2009.01423*, 2020.

[47] Y. Liao, Y. Hua, and Y. Cai, "Deep Learning Based Channel Estimation Algorithm for Fast Time-Varying MIMO-OFDM Systems," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 572–576, March 2020.

[48] P. Dong, H. Zhang, G. Y. Li, N. NaderiAlizadeh, and I. S. Gaspar, "Deep CNN for Wideband Mmwave Massive Mimo Channel Estimation Using Frequency Correlation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May. 2019, pp. 4529–4533.