Deep Compressed Sensing-Based Cascaded Channel Estimation for RIS-Aided Communication Systems

Wenwu Xie, Jian Xiao, Peng Zhu, Chao Yu, and Liang Yang

***Abstract*—To reduce the pilot overhead of cascaded channel estimation for RIS-aided Massive MIMO communication system, we proposed a deep compressed sensing-based channel estimation scheme, where U-shaped network (U-Net), an encoder-decoder with skip connection, is used to recover the high-dimensional cascaded channel matrix from limited pilot overhead. The skip connections between encoder and decoder can fuse features of different scales and semantic by concatenating the feature map, which enhance the reconstruction performance of cascaded channel. To further improve the feature extraction ability of U-Net, we design a ResU-Net architecture with stacked residual units to increase the depth of network. Simulation results show the channel estimation of ResU-Net is more accurate than conventional algorithm and other network model. Meanwhile, ResU-Net has good generalization and robustness for different pilot lengths and phase quantization errors.**

***Index Terms*—Reconfigurable intelligent surface, channel estimation, deep compressed sensing, encoder-decoder.**

I. Introduction

BY EMPLOYING a large number of sub-wavelength units with tunable electromagnetic response, reconfigurable intelligent surface (RIS) can artificially control the propagation of electromagnetic wave to realize the smart radio environment [1], which have been applied to the various communication scenarios [2], [3]. The most promising applications of RIS required the accurate channel state information (CSI) to design the reflection vector of RIS. However, in RIS-aided Massive MIMO system, high-dimensional cascaded channel estimation will bring more pilots overhead, which becomes a key challenge to adjust the passive beamforming of RIS in real time.

Many efficient design ideas have been proposed for reducing pilot overhead, including the semi-passive channel estimation [4], the compressed sensing (CS)-based sparsity channel estimation [5] and channel vary characteristic-based channel factorization [6]. In [4], a low-complexity channel estimation algorithm based on direction-of-arrival (DOA) estimation was proposed. However, few radio frequency (RF) chains are equipped with RIS for DOA estimation, which increased the energy consumption. Considering the sparsity in mmWave channel, the work of [5] deduced a sparse representation of cascaded channel based on the product of Katri-Rao and Kronecker, and CS algorithms are used to recover the sparse channel. In [6], the cascaded channel estimation was modeled as a matrix factorization problem according to the characteristic of slow-varying and fast-varying channel components.

Due to the powerful non-linear mapping ability of deep learning (DL) technology, data driven-based DL model is widely used to the wireless communication. The work of [7] first applied DL to the online configuration of RIS, which significantly enhanced the indoor signal focusing. Considering the multiple RISs-aided communication system, a supervised learning-based phase configuration scheme was proposed in [8], which can obtain the achievable rate closed to the optimal scheme. In [9], residual network (ResNet) was used to denoise the CSI obtained by least square (LS) algorithm. However, the LS-based initial channel estimation required a lot of pilot overhead. In [10], deep complex-value denoising network was used to improve the estimation accuracy of CS algorithm. In [11], [12], channel estimation was regarded as the super-resolution (SR) reconstruction in computer vision. The design ideas of [9]–[12] used traditional algorithm and DL to realize channel estimation stage by stage, which was not easy for the online deployment of general intelligent communication system. Based on the element-by-element reflecting protocol, convolutional neural network (CNN) was used to construct an end-to-end mapping from received pilot signal to CSI in [13], where the required pilot length was equal to LS algorithm. In the existed end-to-end channel estimation network, the number of neural nodes of input layer was determined by the pilot lengths, and thus a trained model was not compatible with the different pilot lengths.

In order to reduce the pilot overhead and improve the flexibility of channel estimation network for different pilot lengths, we combine the deep compressed sensing (DCS) framework [14] and U-shaped network (U-Net) [15] to reconstruct the cascaded channel matrix from limited pilot overhead. Firstly, we boost the dimension of low-dimensional received

Manuscript received November 9, 2021; revised January 4, 2022; accepted January 26, 2022. Date of publication January 31, 2022; date of current version April 11, 2022. This work was supported in part by the Open Fund of Advanced Cryptography and System Security Key Laboratory of Sichuan Province under Grant SKLACSS-202115; in part by the Natural Science Foundation of Hunan Province under Grant 2021JJ40228 and Grant 2020JJ4341; in part by the Key Projects of Hunan Provincial Department of Education Department under Grant 21A0408; and in part by the Outstanding Youth Project of Hunan Provincial Education Department under Grant 20B267 and Grant 20B269. The associate editor coordinating the review of this article and approving it for publication was G. Alexandropoulos. *(Corresponding author: Chao Yu.)*

Wenwu Xie is with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China, and also with the Advanced Cryptography and System Security Key Laboratory of Sichuan Province, Chengdu 610103, China (e-mail: gavinxie2015@qq.com).

Jian Xiao, Peng Zhu, and Chao Yu are with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414006, China (e-mail: victor@vip.hnist.edu.cn; jianchongren@sina.com; 12012017@hnist.edu.cn).

Liang Yang is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: liangy@hnu.edu.cn).

Digital Object Identifier 10.1109/LWC.2022.3147590

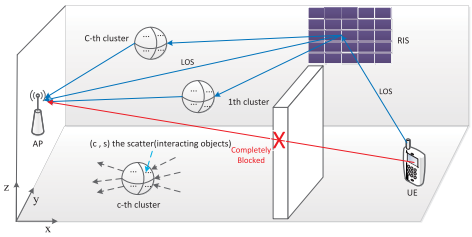

Fig. 1. The RIS-aided indoor communication system with C clusters between RIS and AP.

pilot signal by matrix multiplication, which is equal to the linear layer of neural network. Secondly, U-Net, an encoder-decoder architecture with skip connection, is adopted to compress the input data and reconstruct the cascaded channel matrix. Moreover, we introduce residual learning [16] to the original U-Net (called ResU-Net) to increase the depth of network, which can further improve the channel estimation accuracy.

II. System Model

A. Channel Model

Considering an indoor RIS-aided uplink mmWave communication system, Fig. 1 shows the three-dimensional (3D) geometry model. The access point (AP) with $M = M_1 \times M_2$ uniform planar array (UPA) antennas lies on the yz plane and the RIS with $N = N_1 \times N_2$ UPA reflection elements is mounted on the side wall (zx plane). Let $\boldsymbol{G} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{h} \in \mathbb{C}^{N\times 1}$ denote the RIS-AP channel and the user-RIS channel, respectively. Following the 3GPP standardization and 5G indoor physical channel modeling [17], [18], the clustered statistical MIMO model is adopted [19]. Suppose M_{S} scatters are grouped under C clusters between the AP and the RIS, and each cluster has $S_c (c = 1, 2, \ldots, C)$ scatters. The channel $\boldsymbol{G}$ can be decomposed of the non-line of sight (NLoS) component $\boldsymbol{G}_{\mathrm{NLoS}}$ and line of sight (LoS) component $\boldsymbol{G}_{\mathrm{LoS}}$. $\boldsymbol{G}_{\mathrm{NLoS}}$ can be represented as

$$\boldsymbol{G}_{\mathbf{NLoS}} = \gamma \sum_{c=1}^{C} \sum_{s=1}^{S_c} \bar{\beta}_{c,s} \sqrt{G_e\left(\alpha_{c,s}^{G_t}\right) L_{c,s}^{G_r}} \times \boldsymbol{a}_1\left(\phi_{c,s}^{G_r}, \alpha_{c,s}^{G_r}\right) \boldsymbol{a}_2{}^T\left(\phi_{c,s}^{G_t}, \alpha_{c,s}^{G_t}\right) \quad (1)$$

where $\gamma = \sqrt{\frac{1}{\sum_{c=1}^{C} S_c}}$ is a normalization factor and $\bar{\beta}_{c,s} \sim \mathcal{CN}(0, 1)$ is the path gain. $\phi_{c,s}^{G_r}$ ($\alpha_{c,s}^{G_r}$) and $\phi_{c,s}^{G_t}$ ($\alpha_{c,s}^{G_t}$) represent the azimuth (elevation) angle of arrival (AoA) at the AP, and the azimuth (elevation) angle of departure (AoD) at the RIS for the (c, s)th path, respectively. The AoD $\phi_{c,s}^{G_t}$ and $\alpha_{c,s}^{G_t}$ follow conditionally Laplacian distribution. $G_e(\alpha_{c,s}^{G_t})$ denotes the RIS element radiation for the (c, s)th scatter. We adopt $\cos^q$ pattern to model $G_e(\alpha_{c,s}^{G_t})$, which is a popular representation for feed patterns in the reflectarray antenna design [20, Sec. 9.7.3]. $G_e(\alpha_{c,s}^{G_t})$ can be expressed as

$$G_e\left(\alpha_{c,s}^{G_t}\right) = \varepsilon \cos^{2q}\left(\alpha_{c,s}^{G_t}\right) \quad (2)$$

where q determines the element gain, and $\varepsilon = 2(2q + 1)$ is a normalization term to satisfy conservation of energy.

$L_{c,s}^{G_r}$ denotes the path loss for the (c, s)th path, and we adopt 5G path loss model to represent $L_{c,s}^{G_r}$ as [21]

$$L_{c,s}^{G_r} = -20\log_{10}\left(\frac{4\pi}{\lambda}\right) - 10n\left(1 + b\left(\frac{f - f_0}{f_0}\right)\right)\log_{10}(d_{c,s}) - X_{\sigma_x} \quad (3)$$

where λ is the carrier wavelength. n, b and f_0 denote model parameter, reference frequency and the path loss exponent, respectively. $d_{c,s}$ denotes the ray length of the (c, s)th scatter path and $X_{\sigma_x} \sim \mathcal{CN}(0, {\sigma_x}^2)$ is a shadow factor.

The array response $\boldsymbol{a}(\phi, \alpha)$ for $n_1 \times n_2$ UPA can be expressed as

$$\boldsymbol{a}(\phi, \alpha) = \left[1 \cdots e^{j2\pi d(x sin\alpha + y \sin\phi\cos\alpha)/\lambda} \cdots e^{j2\pi d((n_1-1) sin\alpha + (n_2-1)\sin\phi\cos\alpha)/\lambda}\right] \quad (4)$$

where $0 \le x \le n_1 - 1$ and $0 \le y \le n_2 - 1$. d represents the antenna spacing.

Similarly, $\boldsymbol{G}_{\mathrm{LoS}}$ can be expressed as

$$\boldsymbol{G}_{\mathrm{LoS}} = \sqrt{G_e\left(\alpha_{\mathrm{LoS}}^{G_r}\right) L_{\mathrm{LoS}}^{G_r}} e^{j\eta} \boldsymbol{a}\left(\phi_{\mathrm{LoS}}^{G_r}, \alpha_{\mathrm{LoS}}^{G_r}\right) \boldsymbol{a}^T\left(\phi_{\mathrm{LoS}}^{G_t}, \alpha_{\mathrm{LoS}}^{G_t}\right) \quad (5)$$

where $G_e(\alpha_{\mathrm{LoS}}^{G_r})$ represents the feed pattern of RIS elements in the LoS direction. $L_{\mathrm{LoS}}^{G_r}$ stands for the path loss of the LoS link. η follows the uniform distribution $\eta \sim \mathcal{U}[0, 2\pi]$. $\phi_{\mathrm{LoS}}^{G_r}$ ($\alpha_{\mathrm{LoS}}^{G_r}$) and $\phi_{\mathrm{LoS}}^{G_t}$ ($\alpha_{\mathrm{LoS}}^{G_t}$) denote the azimuth (elevation) AoA at the AP and the azimuth (elevation) AoD at the RIS in the LoS direction, respectively.

Following [19], we consider that the RIS are sufficiently close to the low-mobility user in the indoor office environment, so that a clear LoS link exists between RIS and user. The LoS-dominated channel $\boldsymbol{h}$ can be represented as

$$\boldsymbol{h} = \sqrt{G_e(\alpha^r) L^r} e^{j\eta} \boldsymbol{a}(\phi^r, \alpha^r) \quad (6)$$

where $G_e(\alpha^r)$ represents the radiation of RIS. L^r is the path loss. ϕ^r (α^r) denotes the azimuth (elevation) AoA at the RIS.

B. Problem Formulation

Let $\boldsymbol{\theta} = [e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}]^T \in \mathbb{C}^N$ denotes the reflecting vector at the RIS, where $\theta_i (i = 1, 2, \ldots, N)$ represents the phase shift at the i-th RIS element. Assuming Q pilot slots are used for channel estimation, the received signal $y_q (q = 1, 2, \ldots, Q)$ at the AP can be expressed as

$$\begin{aligned} y_q &= \boldsymbol{G}\,\mathrm{diag}(\boldsymbol{\theta}_q)\boldsymbol{h} s_q + w_q \\ &= \boldsymbol{G}\,\mathrm{diag}(\boldsymbol{h})\boldsymbol{\theta}_q s_q + w_q \end{aligned} \quad (7)$$

where s_q and $\boldsymbol{\theta}_q$ are the q-th transmitted pilot and q-th reflecting vector, respectively. $w_q \sim \mathcal{CN}(0, {\sigma_n}^2 I_M)$ is white Gaussian noise.

We define $\boldsymbol{H} = \boldsymbol{G}\,\mathrm{diag}(\boldsymbol{h}) \in \mathbb{C}^{M\times N}$ as the cascaded channel. Suppose $s_q = 1$ for each pilot, we can obtain the $M \times Q$ observation matrix $\boldsymbol{Y} = [y_1, y_2, \ldots, y_Q]$.

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{\Theta} + \boldsymbol{W} \quad (8)$$

where $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_Q] \in \mathbb{C}^{N\times Q}$ and $\boldsymbol{W} = [w_1, w_2, \ldots, w_Q] \in \mathbb{C}^{M\times Q}$.

The inverse estimation of (8) is ill-posed because the problem is underdetermined with $Q < N$. Since $\boldsymbol{H}$ has sparsity in a specific transform domain φ for mmWave Massive MIMO communication with limited scatters, e.g., angular domain. Let $\boldsymbol{H} = \varphi\tilde{\boldsymbol{H}}$, where φ denotes the sensing matrix and $\tilde{\boldsymbol{H}}$ is a sparsity matrix with $k \ll M \times N$ non-zero elements. The straightforward formulation of CS reconstruction can be expressed as [22]

$$\min_{\tilde{\boldsymbol{H}}} \|\tilde{\boldsymbol{H}}\|_p \quad s.t. \quad \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\Theta}\|_2 \le \xi \tag{9}$$

where ξ bounds the amount of noise in the data. The subscript p is usually set to 1 or 0, characterizing the sparsity of $\tilde{\boldsymbol{H}}$. Numerous methods have been proposed for solving the optimization problem of (9), e.g., convex optimization method and matching pursuit method based on greedy algorithms. However, there is no explicit theory to generate the sparse channel representation $\tilde{\boldsymbol{H}}$ with the least non-zero entries for complex communication environment [23]. Moreover, the mathematical model-based cascaded channel estimation algorithm, e.g., LS and CS, required the accurate prior knowledge of reflecting vector $\boldsymbol{\theta}$, which is not available for perfect phase configuration in practice [24]. Considering the phase quantization error of RIS elements, the actual reflecting vector can be expressed as $\hat{\boldsymbol{\theta}} = [e^{j\hat{\theta}_1}, e^{j\hat{\theta}_2}, \ldots, e^{j\hat{\theta}_N}]$ and $\hat{\theta}_i = \theta_i + \bar{\theta}_i (i = 1, 2, \ldots, N)$, where $\bar{\theta}_i \sim \boldsymbol{\mathcal{U}}[-2^{-b}\pi, 2^{-b}\pi]$ and b represent the phase quantization error and quantization bits, respectively.

III. Proposed Method

In this section, we first construct a general dataset to make the neural network compatible with different pilot lengths. And then the ResU-Net is proposed to recover the high-dimensional cascaded channel matrix from limited pilot overhead.

A. Data Preprocessing

For general end-to-end channel estimation models, the received pilot signal is directly used as the input of network. When we adjust the pilot length for different communication scenarios, the number of neural node of input layer needs to be correspondingly changed to adapt the pilot length. Hence, the trained model is not flexible for various set of communication system. Motivated by [14], a linear layer is used to boost the dimension $\mathbb{C}^{M\times Q}$ of the observation matrix $\boldsymbol{Y}$ to the $\mathbb{C}^{M\times N}$, where the weights of linear layer are fixed and are set to the transpose of reflecting vector Θ. In fact, the operation of linear layer with fixed weight is equivalent to the matrix multiplication. Consequently, we denote received pilot signal proxy $\bar{\boldsymbol{Y}} = \boldsymbol{Y}.\boldsymbol{\Theta}^T$ as the input of the network. The received pilot signal and the cascaded channel are normalized by the maximum absolute value of their elements. We treat $\bar{\boldsymbol{Y}}$ and $\boldsymbol{H}$ as two-channel image with $M \times N \times 2$, where each channel is the real and imaginary part of complex data, respectively.

B. Network Architecture

In the data preprocessing stage, we boost the dimension of pilot matrix through mathematical operation, which increases the redundant features and computational complexity of the network. Consequently, we design an encoder-decoder architecture to recover the desired cascaded channel matrix from the feature map compressed by encoder. In the regular encoder-decoder architecture, e.g., Autoencoder (AE), the reconstruction performance of high-dimensional data is limited. Fig. 2 shows the proposed ResU-Net architecture, where the semantic information at different scales is fused by designing skip connections between encoder and decoder, so that ResU-Net can reduce information loss in the process of feature compression.

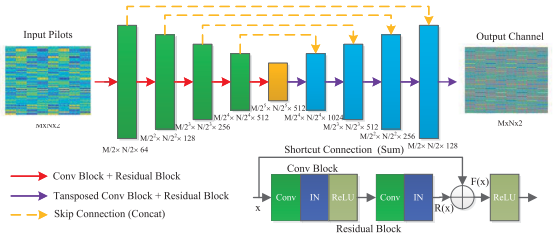

Fig. 2. The network architecture of ResU-Net.

In the stage of encoder, E encoder blocks are used to compress the input data with dimensions of $M \times N \times 2$ into $M/2^E \times N/2^E \times 512, E \le \lfloor log_2 M \rfloor$, where each encoder block consists of a convolutional block and a residual block. Each convolutional block compose of a convolutional layer with stride 2 to downsample the feature map, an instance normalization layer and a LeakyReLU layer. The number of filter $f_i^e = 64 \times 2^{i-1}, (i = 1, 2, \ldots, 4)$ and $f_i^e = 512$ when $i \ge 5$ for convolutional layer, where i denotes the sequence number of the encoder block from outer to inner. The deeper convolutional network is, the more high-level semantic information of input data can be extracted. ResNet is widely used as the backbone of the deep network, and the basic idea of residual learning can be expressed as $F(x) = R(x) + x$, where $R(x)$ represents the output after convolution operation and x denotes the initial input data. The residual learning introduces shortcut connections to avoid the vanishing gradient problem. Hence, we add residual blocks into the U-Net architecture to improve the feature extraction ability of the network.

The decoder of the ResU-Net is symmetric with the encoder, where the feature map is upsample by transposed convolution with stride 2. E decoder blocks are used to recover the compressed feature with dimensions of $M/2^E \times N/2^E \times 512$ into $M \times N \times 2$, where each decoder block consists of a transposed convolutional block and a residual block. The number of filter of transposed convolutional layer is set to $f_i^d = 512/2^{i-1}, (i = 1, 2, \ldots, 4)$, where i denotes the sequence number of the decoder block from inner to outer. The filters of last transposed convolutional layer is set to 2 to satisfy the output with two-channel. To release the checkboard artifacts when feature map is upsampled, we adopt 4×4 convolutional kernel. Different from the AE, the low-level feature maps of encoder and high-level feature maps of decoder are concatenated by skip connection in the ResU-Net, which help the decoder to reconstruct more accurate cascaded channel and accelerate the convergence of network. In fact, channel

estimation is a low-level regression task, where the low-level senmantic information extracted from shallow convolutional layer can facilitate the network to complete the final task.

C. Network Training

We use normalized mean squared error (NMSE) as the performance evaluation metric of channel estimation, which can be represented as

$$\mathrm{NMSE} = \mathbb{E}\left\{ ||\hat{\boldsymbol{H}} - \boldsymbol{H}||_F^2 / ||\boldsymbol{H}||_F^2 \right\} \tag{10}$$

where $\hat{\boldsymbol{H}}$ is the estimated cascaded channel, and $||\boldsymbol{\cdot}||_F$ denotes the Frobenius norm.

Although L_2 loss function is directly related with the NMSE performance metric, L_1 loss has faster convergence than L_2 loss in practice [25], where $L_1 = \mathbb{E}\{|\boldsymbol{H} - f(\bar{\boldsymbol{Y}})|\}$ and $f(\bar{\boldsymbol{Y}})$ represents the output of ResU-Net. In the training process, we use the cosine learning rate decay schedule to achieve stable convergence of network. The learning rate η_i at the i-th epoch can be represented as

$$\eta_i = \frac{1}{2}\eta_0 \left(1 + \cos\left(\frac{i}{I}\pi\right)\right), \ 0 \le i \le I \tag{11}$$

where η_0 and I represent initial learning rate and the total number of epochs, respectively.

IV. Numerical Simulation

In our simulation, $M = 8 \times 8$, $N = 16 \times 16$, $E = 5$, $I = 150$ and $\eta_0 = 0.02$. The coordinates of AP, RIS and user are set to $(x^{AP}, y^{AP}, z^{AP}) = (0, 25, 2)$, $(x^{RIS}, y^{RIS}, z^{RIS}) = (40, 50, 2)$ and $(x^{UE}, y^{UE}, z^{UE}) = (38, 48, 1)$. C follows the Poisson distribution $C \sim \max\{P(\lambda_p), 1\}$, where $\lambda_p = 1.8$ for 28 GHz frequency band [18], and $S_c \sim \boldsymbol{\mathcal{U}}[1, 30]$ [26]. We adopt the path loss model of InH Indoor Office, where $n = 3.19$, $b = 0.06$, $\sigma = 8.29$ dB and $f_0 = 24.2$ GHz for NLoS component, while $n = 1.73$, $b = 0.06$ and $\sigma = 3.02$ dB for LoS component [21]. We generate 20000 channel samples, which are randomly divided into training, validation and test datasets by the ratios of 60%, 20%, and 20%, respectively. In training phase, the SNR varies from 0 dB to 30 dB with an interval of 5dB and the pilot length of training set is set to $Q = 32$. The trained model can be tested for different SNR and Q. The network is optimized by Adam optimizer and batch size is set to 64, where the coefficients to compute gradient are set to $(\beta_1, \beta_2) = (0.9, 0.99)$. All simulations are performed on the Xeon Silver 4210R CPU and Nvidia TITAN RTX GPU.

We compared the NMSE performance of proposed ResU-Net with conventional LS estimator, CS algorithm, i.e., Oracle LS and OMP, and other DL model, i.e., plain CNN, AE and U-Net. Fig. 3 shows ResU-Net outperform the other method in the most of SNR ranges. In the channel estimation, neural network can be regarded as a universal approximator to realize the mapping from pilot signal to channel matrix. Since the estimation error of LS is the inverse of noise, LS estimator can obtain ideal performance under the high SNR, while the approximation error of neural network will be larger. Since the sparity k is variable and relatively large in our clustered MIMO channel model, we set sparity $\bar{k} = 400$ in OMP algorithm for better NMSE performance. However, the estimation accuracy of CS is limited even if we increase the number of iterations. Due to the lack of skip connection, the decoder of AE cannot use the low-level features obtained by the encoder, so it is difficult to recover accurate cascaded channel. ResU-Net fuses feature maps of different scales by skip connection and makes the learned features more refined by stacking residual blocks. Consequently, ResU-Net can obtain superior NMSE performance under a few pilot overhead.

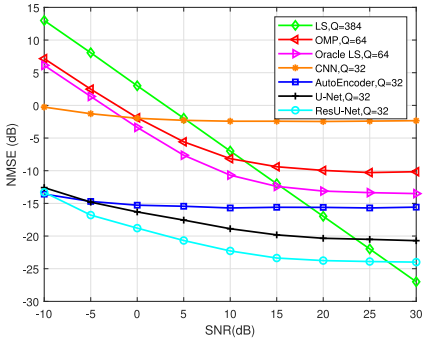

Fig. 3. NMSE vs. SNR for different algorithm.

TABLE I
Comparison of Computation Time on CPU and GPU

Algorithm	pilot length(Q)	CPU(ms)	GPU(ms)
LS	257(384)	22.6(30.3)	/
Oracle LS	64(128)	10.8(12.5)	/
OMP	64(128)	1583(1812)	/
CNN	32($\cdots$)	463	6.16
AutoEncoder	32($\cdots$)	40.1	1.75
U-Net	32($\cdots$)	52.1	1.82
ResU-Net	32($\cdots$)	95.7	4.27

Table I shows the computation time of channel estimation module for different algorithm, which is obtained by averaging 200 times running. The computation time of all algorithms is compared in Python3.7, where PyTorch 1.10 is used as the development frame of deep learning. The oracle LS algorithm can be fast performed due to the known supports of $\tilde{\boldsymbol{H}}$, while OMP based on iterative optimization strategy takes a lot of computation time. As a linear estimator, the large pilot overhead increases the computation time of LS estimator. Since the size of feature map is always $M \times N$ in CNN, the computation of network bring high computational complexity. Compared with AE and U-Net, the computational complexity of ResU-Net is relatively large due to introduce more network layers. The computation time of LS estimator and CS algorithm will increase with the pilot length Q. However, the computation time of proposed model is the same for different Q since the dimension of input data has been fixed during data preprocessing. Meanwhile, the computation time of DL model can be greatly reduced by utilizing the parallel computing ability of graphics processing unit (GPU).

In Fig. 4, we compared the NMSE performance of different algorithm for different pilot lengths. With the increase

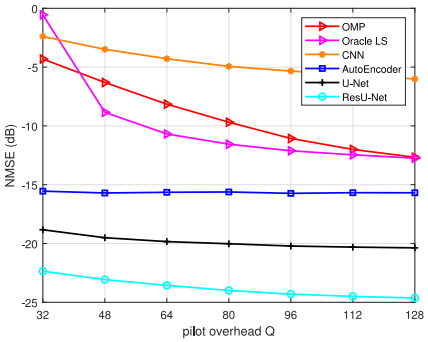

Fig. 4. NMSE vs. pilot overhead Q when the SNR is set to 10 dB.

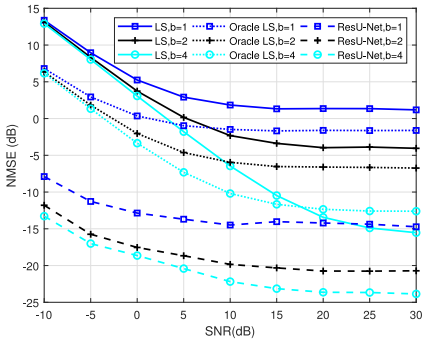

Fig. 5. NMSE vs. SNR for different quantization bits b.

of Q, the performance of CS will be improved, while the performance variation of encoder-decoder architecture is not obvious for different pilot lengths. In Fig. 5, the NMSE performance of mathematical model-based conventional channel estimation algorithm significantly reduced for phase shift with quantization error. Since the neural network is robust for some disturbances of input data, e.g., quantization noise, and the downsampling of encoder can reduce the risk of overfitting, ResU-Net is superior to the LS estimator and the oracle LS algorithm for phase shift with different quantization bits b.

V. Conclusion

Considering the high pilot overhead of cascaded channel estimation for RIS aided Massive MIMO communication system, we proposed a DCS-based channel estimation scheme, where ResU-Net is designed to reconstructed the high-dimensional cascaded channel. By introducing the skip connections to fuse the features of different scales, the NMSE performance of ResU-Net outperforms other deep neural network and conventional channel estimator. The generalization and robustness of ResU-Net have been verified for different pilot lengths and quantization bits.

References

[1] 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.

[2] L. Yang, J. Yang, W. Xie, M. O. Hasna, T. Tsiftsis, and M. D. Renzo, "Secrecy performance analysis of RIS-aided wireless communication systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12296–12300, Oct. 2020.

[3] L. Yang, F. Meng, J. Zhang, M. O. Hasna, and M. D. Renzo, "On the performance of RIS-assisted dual-hop UAV communication systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10385–10390, Sep. 2020.

[4] X. Chen, J. Shi, Z. Yang, and L. Wu, "Low-complexity channel estimation for intelligent reflecting surface-enhanced massive MIMO," in *Proc. IEEE ICASSP*, Brighton, U.K., May 2019, pp. 4659–4663.

[5] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, May 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9103231

[6] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, Nov. 2020.

[7] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE SPAWC*, Cannes, France, Jul. 2019, pp. 1–5.

[8] G. C. Alexandropoulos, S. Samarakoon, M. Bennis, and M. Debbah, "Phase configuration learning in wireless networks with multiple reconfigurable intelligent surfaces," in *Proc. GLOBECOM (Workshops)*, Taipei, Taiwan, 2020, pp. 1–6.

[9] 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," May 2021, *arXiv:2009.01423*.

[10] 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.

[11] 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. 11083–11088, Oct. 2021.

[12] Y. Wang, H. Lu, and H. Sun, "Channel estimation in IRS-enhanced mmwave system with super-resolution network," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2599–2603, Aug. 2021.

[13] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-Wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, Sep. 2020.

[14] A. Mousavi and R. G. Baraniuk, "Learning to invert: Signal recovery via deep convolutional networks," in *Proc. IEEE ICASSP*, 2017, pp. 2272–2276.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[17] "3GPP TR 38.901 V16.1.0-study on channel model for frequencies from 0.5 to 100 GHz," ETSI, Sophia Antipolis, France, ETSI TR 138 901–2018, Dec. 2019.

[18] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Quart., 2018.

[19] E. Basar, I. Yildirim, and F. Kilinc, "Indoor and outdoor physical channel modeling and efficient positioning for reconfigurable intelligent surfaces in mmWave bands," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8600–8611, Dec. 2021.

[20] P. Nayeri, F. Yang, and A. Z. Elsherbeni, *Reflectarray Antennas: Theory, Designs, and Applications*. New York, NY, USA: Wiley, 2018.

[21] N. Docomo, "White On Paper 5G Channel Model For Bands Up To 100 GHz," 2016. [Online]. Available: http://www.5gworkshops.com/5GCMSIG White

[22] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[23] E. Balevi, A. Doshi, A. Jalal, A. Dimakis, and J. G. Andrews, "High dimensional channel estimation using deep generative networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 18–30, Jan. 2021.

[24] M.-A. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 184–188, Feb. 2020.

[25] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[26] M. K. Samimi and T. S. Rappaport, "Statistical channel model with multi-frequency and arbitrary antenna beamwidth for millimeter-wave outdoor communications," in *Proc. IEEE GLOBECOM Workshops*, 2015, pp. 1–7.