# Multi-Scale Supervised Learning-Based Channel Estimation for RIS-Aided Communication Systems

Jian Xiao*, Ji Wang*, Wenwu Xie†, Xinhua Wang‡, Chaowei Wang§, and Hongbo Xu*
*Department of Electronics and Information Engineering, Central China Normal University, Wuhan, China
†School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China
‡College of Electrical Engineering, Qingdao University, Qingdao, China
§School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China
Email: jianx@mails.ccnu.edu.cn, {jiwang, xuhb}@ccnu.edu.cn, xhwang@qdu.edu.cn, wangchaowei@bupt.edu.cn

2023 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-9122-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCNC55385.2023.10119023

***Abstract*—Motivated by the development of single image super-resolution (SR) reconstruction in computer version, classic SR networks have been widely applied to the channel estimation of wireless communication system. To capture the spatial correlations in the reflection element-domain of reconfigurable intelligent surface (RIS), we propose a multi-scale supervised learning-based Laplacian pyramid wide residual network (LapWRes) to achieve the progressive reconstruction of cascaded channel in a coarse-to-fine fashion. The LapWRes can be divided vertically into feature extraction branch (FEB) and channel reconstruction branch (CRB), while it can also be viewed horizontally as multiple channel reconstruction modules (RMs) at different scales. In the FEB, the wide activation residual blocks are stacked to extract the high-frequency information of cascaded channel. In the CRB, the high-frequency and low-frequency information of cascaded channel is fused by utilizing the residual learning. Simulation results show that the LapWRes can achieve better estimation accuracy than other channel estimation schemes and faster convergence than existing SR network-based channel estimation models.**

## I. Introduction

Considering lots of communication bandwidth available at high frequency, millimeter wave (mmWave) is regard as the promising communication frequency for the future wireless communication system [1]. However, the significant path loss of high-frequency electromagnetic waves limits the coverage of mmWave communication. The intuitive solutions are to deploy denser base stations or to integrate more antennas into communications equipment, which will bring expensive hardware cost and much energy consumption. Reconfigurable intelligent surface (RIS) is a metasurface composed of sub-wavelength units, whose electromagnetic response of each unit can be tunable by adjusting the size or spatial arrangement of units. By utilizing the unique electromagnetic properties, RIS provide a possibility to enhance mmWave communication with low cost and energy [2]. However, high-dimensional channel estimation is the key challenge for RIS with a large of passive reflection elements [3].

To reduce pilot overhead of high-dimensional channel estimation for RIS-aided communication system, many works have provided various design ideas, e.g., the semi-passive channel estimation by equipping with few radio frequency chains in RIS [4], the compressed sensing-based channel estimation by exploiting the sparsity of RIS channel [5] and the two-timescale channel estimation by utilizing the channel vary characteristic [6]. Moreover, deep learning (DL) has been proved feasible to improve the accuracy of channel estimation and reduce pilot overhead by learning the inherent features of massive communication data [7]-[9]. In the DL-based channel estimation scheme, single image super-resolution (SR) reconstruction technologies have been widely used. The theoretical foundation of the SR network-based channel estimation scheme is the correlations of channel matrix, e.g., the correlations of time-frequency domain in orthogonal frequency division multiplexing (OFDM) systems. In [10], super-resolution convolutional neural network (SRCNN) was applied to recover the complete time-frequency channel form partial channel of pilot subcarriers. However, the reconstruction performance of SRCNN was limited due to the simple network architecture. In [11], enhanced SR network (EDSR) was used to further improve the channel estimation accuracy by introducing the residual learning.

Since the metamaterial units of RIS are generally integrated closely, the channel at the neighboring units are highly correlated in spatial correlations domain. Hence, the design ideas in [10], [11] have been extend to the RIS-aided communication system. In [12], the low-dimensional cascaded channel matrix was obtained by opening partial RIS elements firstly, and then SRCNN was applied to recover the high-dimensional cascaded channel from the low-dimensional cascaded channel matrix. The work of [13] considered the part of cascaded channel estimation based on EDSR, where some active elements were equipped with RIS to acquire the initial channel information. In the above works, the channel extrapolation was realized in one upsampling step, e.g., the pre-upsampling in the input layer of SRCNN [10], [12] or the post-upsampling in the output layer of EDSR [11], [13], which leads to reconstruction difficulties of channel will increase for large upscale factor. In the conventional channel estimation schemes of RIS communication system, some dimension reduced channel estimation ideas have been proposed by exploiting the spatial correlations, e.g., element group [14], and similar ideas can be found in SR, e.g., Laplacian pyramid network [15], which motivated us to progressive reconstruct the cascaded channel matrix from small scale to large scale.

In this paper, we propose a multi-scale supervised learning-

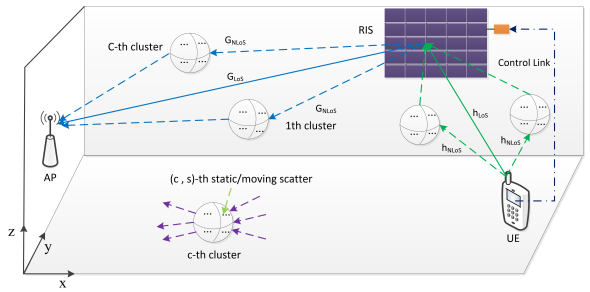


Fig. 1. The three-dimensional RIS-aided mmWave communication environment with random scattering elements.

based cascaded channel estimation scheme in RIS-aided communication system, where we design the Laplacian pyramid wide residual network (LapWRes) to progressive reconstruct the cascaded channel in a coarse-to-fine fashion, instead of one step reconstruction. The feature map is composed of high frequency and low-frequency components in neural network. With the increase of network layers, the representation of network will contain more high-frequency information. Hence, we introduce residual learning to fuse the high-frequency and low-frequency features by design two branches, i.e., feature extraction branch (FEB) and channel reconstruction branch (CRB). Meanwhile, wide activation residual blocks (WARBs) are used in the interior structure of FEB to improve the feature extraction ability of LapWRes [16].

## II. System Model and Problem Formulation

We consider a RIS assisted uplink mmWave communication system in Fig. 1., where a single-antenna user equipment (UE) simultaneously communicates the base station (BS) with $M = M_1 \times M_2$ uniform planar array (UPA) uniform planar array antennas via the RIS with $N = N_1 \times N_2$ reflection elements. Let $\mathbf{G}$ and $\mathbf{h}$ represent the RIS-BS channel and the UE-RIS channel, respectively. Following the channel modeling in [17], the clustered statistical MIMO channel model is used to capture the dynamic variations of the environmental objects, e.g., a large number of randomly distributed scattering elements between the terminals and the RIS. The RIS-BS channel $\mathbf{G} = \mathbf{G}_{\text{NLOS}} + \mathbf{G}_{\text{LOS}}$ can be represented as

$$\mathbf{G} = \underbrace{\sqrt{G_e\left(\varphi_{\text{LOS}}^{G_r}\right) L_{\text{LOS}}^{G_r}}\mathbf{a}\left(\phi_{\text{LOS}}^{G_r}, \varphi_{\text{LOS}}^{G_r}\right)\mathbf{a}^T\left(\phi_{\text{LOS}}^{G_t}, \varphi_{\text{LOS}}^{G_t}\right)}_{\mathbf{G}_{\text{LOS}}} + \underbrace{\bar{\gamma}\sum_{c=1}^{\bar{C}}\sum_{s=1}^{\bar{S}_c}\bar{\beta}_{c,s}\sqrt{G_e\left(\varphi_{c,s}^{G_r}\right) L_{c,s}^{G_r}}\mathbf{a}\left(\phi_{c,s}^{G_r}, \varphi_{c,s}^{G_r}\right)\mathbf{a}^T\left(\phi_{c,s}^{G_t}, \varphi_{c,s}^{G_t}\right)}_{\mathbf{G}_{\text{NLOS}}}, \tag{1}$$

where $\bar{C}$ and $\bar{S}$ denote the total number of clusters and scatters in cluster $c$ between BS and RIS for non-line of sight (NLOS) component, respectively. $\gamma = \sqrt{\frac{1}{\sum_{c=1}^{C} S_c}}$ is a normalization factor in the clustered channel model. $\bar{\beta}_{c,s} \sim \mathcal{CN}(0,1)$ is the propagation path gain of the $(c,s)$-th scatter. $G_e\left(\theta_{c,s}^{G_r}\right) = 2(2q+1)\cos^{2q}\left(\theta_{c,s}^{G_r}\right)$ denotes the RIS elements pattern for the $(c,s)$-th scatter, where $q$ determines the gain of the element [18]. $L_{c,s}^{G_r}$ represents the path loss for the $(c,s)$-th scatter [19]

$$L_{c,s}^{G_r} = -20\log_{10}\left(\frac{4\pi}{\lambda}\right) - 10n\left(1 + b\left(\frac{f - f_0}{f_0}\right)\right)\log_{10}\left(d_{c,s}\right) - X_{\sigma_x}, \tag{2}$$

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

$\phi_{c,s}^{G_r}$ ($\theta_{c,s}^{G_r}$) and $\phi_{c,s}^{G_t}$ ($\theta_{c,s}^{G_t}$) denote the azimuth (elevation) angle of arrival at the BS, and the azimuth (elevation) angle of departure at the RIS for the $(c,s)$-th scatter, respectively. The UPA array response $\mathbf{a}\left(\phi,\theta\right)$ can be represented as

$$\mathbf{a}\left(\phi,\theta\right) = \left[1 \cdots e^{j2\pi d(x sin\theta + y\sin\phi\cos\theta)/\lambda} \cdots e^{j2\pi d((N_1-1)sin\theta + (N_2-1)\sin\phi\cos\theta)/\lambda}\right], \tag{3}$$

where $0 \le x \le N_1 - 1$ and $0 \le y \le N_2 - 1$. $d$ denotes the antenna spacing.

The UE-RIS channel $\mathbf{h}$ can be represented as

$$\mathbf{h} = \underbrace{\sqrt{G_e\left(\theta_{\text{LOS}}^{r}\right) L_{\text{LOS}}^{r}}\mathbf{a}\left(\phi_{\text{LOS}}^{r}, \theta_{\text{LOS}}^{r}\right)}_{\mathbf{h}_{\text{LOS}}} + \underbrace{\tilde{\gamma}\sum_{c=1}^{\widehat{C}}\sum_{s=1}^{\widehat{S}_c}\widehat{\beta}_{c,s}\sqrt{G_e\left(\theta_{c,s}^{r}\right) L_{c,s}^{r}}\mathbf{a}\left(\phi_{c,s}^{r}, \theta_{c,s}^{r}\right)}_{\mathbf{h}_{\text{NLOS}}}, \tag{4}$$

where $\widehat{C}$ and $\widehat{S}$ represent the total number of clusters and scatters per cluster between the RIS and the UE, respectively. $\widehat{\gamma} = \sqrt{\frac{1}{\sum_{c=1}^{\widehat{C}} \widehat{S}_c}}$. $\widehat{\beta}_{c,s} \sim \mathcal{CN}(0,1)$. $G_e\left(\theta_{c,s}^{r,k}\right)$ represents the RIS element gain, $L_{c,s}^{G_{r,k}}$ denotes the path loss. $\phi_{c,s}^{r,k}$ ($\theta_{c,s}^{r,k}$) is the azimuth (elevation) angle at the RIS.

Let RIS reflecting vector $\Theta = [e^{j\theta_1}, e^{j\theta_2}, \cdots, e^{j\theta_N}]^T \in \mathbb{C}^N$, where $\theta_i (i = 1, 2, \cdots, N)$ denotes the phase shift at $i$-th RIS element. In the $q$-th $(q = 1, 2, \cdots, Q)$ pilot transmission slots, the received pilot signal $y_q$ can be represented as

$$y_q = \mathbf{G}\text{diag}(\Theta_q)\mathbf{h}s_q + w_q = \mathbf{G}\text{diag}(\mathbf{h})\Theta_q s_q + w_q, \tag{5}$$

where $w_q \sim \mathcal{CN}(0, {\sigma_n}^2 I_M)$ stands for Gaussian noise and let $\mathbf{H} = \mathbf{G}\text{diag}(\mathbf{h}) \in \mathbb{C}^{M\times N}$ denotes as the cascaded channel.

Since the performance gain of passive RIS is superior to traditional active relay technology only when there are a large number of reflection elements on RIS, the cascaded channel $\mathbf{H}$ is high-dimensional. Due to the constraint of full-rank condition for existing classic estimator, e.g., least square (LS) estimator, the required minimum pilot overhead is $N$, which causes intractable training overhead for the RIS-aided communication system.

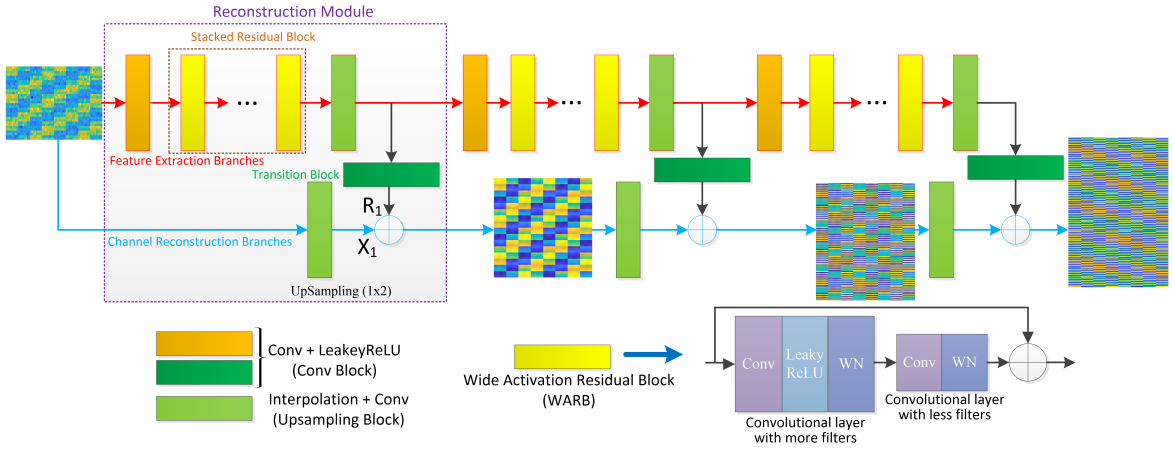


Fig. 2. The proposed Laplacian pyramid wide residual network (LapWRes) architecture.

## III. Proposed Method

### A. Dataset Construction

By exploiting the spatial correlations, we extend the partial cascaded channel to complete cascaded channel matrix by designing the SR network. In the first phase of channel estimation, we select the $\mathcal{P} = \{1, k+1, \cdots, (P-1)\times k+1\}, \left(P = \left\lfloor \frac{N-1}{k} + 1 \right\rfloor\right)$ RIS elements with the interval $k = 2^S, (0 \le S \le \log_2 N)$ as the subset of whole RIS elements, and then estimate the partial cascaded channel by controlling the reflection vector of subset elements. To reduce the computing complexity of the first phase, the simple ON/OFF reflection protocol-based LS algorithm is used to obtain the partial channel $\widehat{\mathbf{H}}^{\mathcal{P}} \in \mathbb{C}^{M\times P}$. Suppose only $p, (1 \le p \le P)$-th elements is turned on at a time slot, the received signal at BS can be represented as $y_p = \mathbf{H}_p^{\mathcal{P}} s_p$, where $\mathbf{H}_p^{\mathcal{P}}$ represents the $p$-th column of cascaded channel $\mathbf{H}^{\mathcal{P}}$. Then the estimated channel $\widehat{\mathbf{H}}_p^{\mathcal{P}} = y_p s_p^{-1}$ by using LS algorithm. We serially turn on each elements in $\mathcal{P}$, and adopt the same method to obtain each column of $\widehat{\mathbf{H}}^{\mathcal{P}}$, i.e., $\widehat{\mathbf{H}}^{\mathcal{P}} = \left[\widehat{\mathbf{H}}_1^{\mathcal{P}}, \widehat{\mathbf{H}}_2^{\mathcal{P}}, \cdots, \widehat{\mathbf{H}}_P^{\mathcal{P}}\right]$.

Let $\bar{\mathbf{H}}^{\mathcal{P}} \in \mathbb{R}^{M\times P\times 2}$ as the input data of network, where the real parts and imaginary parts of complex channel matrix are separated, i.e., $\bar{\mathbf{H}}_{m,p,1}^{\mathcal{P}} = \mathrm{Re}(\widehat{\mathbf{H}}^{\mathcal{P}})$ and $\bar{\mathbf{H}}_{m,p,2}^{\mathcal{P}} = \mathrm{Im}(\widehat{\mathbf{H}}^{\mathcal{P}}), (1 \le m \le M)$. We design the label group $\bar{\mathbf{H}} = (\bar{\mathbf{H}}_1, \bar{\mathbf{H}}_2, \cdots, \bar{\mathbf{H}}_S)$ to achieve the progressive reconstruction of cascaded channel, where $\bar{\mathbf{H}}_S$ represents the complete cascaded channel matrix and $\bar{\mathbf{H}}_s \in \mathbb{R}^{M\times 2^s P\times 2} (1 \le s \le S)$.

### B. Laplacian Pyramid Networks

Fig. 2. shows the proposed Laplacian pyramid wide residual network (LapWRes) architecture with $S$ RMs, where we stage by stage upscale the lower-dimensional channel matrix by a scale of 2 in the reflection element-domain of RIS. Each RM can be divided into two branches: FEB and CRB, which imitate the Laplacian pyramid in image scaling.

*1) Laplacian pyramid:* Laplacian pyramid is the improvement of Gauss pyramid by introducing the residual coefficients. In the Gauss pyramid, the original resolution image at the bottom of pyramid is sequentially downsampled, which forms a set of images arranged from top to bottom in the shape of a pyramid according to the size of image resolution. However, this sampling operation will lose high-frequency information of images. Let $\mathcal{G}(I) = [I_0, I_1, \ldots, I_S]$ denotes a Gauss pyramid with $S$ levels, where $I_s, (0 < s < S)$ denotes the $s$-th level image of pyramid. The $s$-th level of Laplacian pyramid can be represented as

$$h_s = \mathcal{L}_s(I) = \mathcal{G}_s(I) - u\left(\mathcal{G}_{s+1}(I)\right) = I_s - u\left(I_{s+1}\right), \quad (6)$$

where $u\left(I_{s+1}\right)$ denotes the upsampling image of $I_{s+1}$, and residual coefficient $h_s$ contains high-frequency information.

If we regard the cascaded channel matrix as the image, $I_S = \bar{\mathbf{H}}^{\mathcal{P}}$ and $I_0 = \bar{\mathbf{H}}_S$ in the channel estimation. The $s$-th level cascaded channel matrix can be represented as

$$\bar{\mathbf{H}}_s = u\left(\bar{\mathbf{H}}_{s+1}\right) + h_s, \quad (7)$$

According to the structure of Laplacian pyramid, we design the LapWRes to realize the cascaded channel estimation, where the upsampling operator $u(\cdot)$ and the Laplacian coefficients $h_s$ is designed by neural network.

*2) FEB:* In the FEB of $s$-th RM, we first use a convolutional block (CB), composed of convolutional layer and Leaky Rectified Linear Unit (LeakyReLU) activation functional layer, to adjust the channels of feature maps. Next, $W$ WARBs are stacked to extract the more representative features. Different from the residual blocks (RBs) in EDSR, we adopt WARBs to improve the transmission of information flow in the LapWRes [16]. The non-linear activation function can make neural network obtain the non-linear mapping ability, while non-linear transform will loss the partial original information of input data. Since the low-level features of input data is important for the SR tasks, we use the CB with more convolutional kernel to increase the channels of feature maps before the activation in the WDSR, which expands the low-level features and still keeps the non-linearity of LapWRes.

Since batch normalization (BN) will change the original data distribution, BN is not appropriate for the SR tasks, while

the training process without normalization layer is unstable for deep network. In the LapWRes, we adopt weight normalization (WN) to reparameterize the weight vector of network instead of normalizing the mini-batch data of each layer [20]. Compared with BN, the performance of WN is not related with the batch size and data, and the memory and computation overhead is lower. Moreover, the SR-based channel estimation is sensitive to the learning rate $lr$, e.g., $lr$ is set a small value $lr = 10^{-3}$ in [11]. The training loss of network without WN layer will explode for a larger $lr$, while the small learning rate is easy to lead to overfitting. The WN can provide a wider range of $lr$ in the training, which improve the estimation accuracy in the test phase. We use an upsampling block (UB) to scale the feature map to desired dimension of channel matrix, e.g., $\bar{\mathbf{H}}^{\mathcal{P}} \in \mathbb{R}^{M\times P\times 2} \rightarrow \bar{\mathbf{H}}_1 \in \mathbb{R}^{M\times 2p\times 2}$ in the first RM. In the UB, we adopt nearest interpolation and convolutional layer to increase the size of feature map, which can avoid the check artifacts in the upsampling.

*3) CRB:* A feature map $I$ composed of high-frequency feature $I_H$ and low-frequency feature $I_L$, that is $I = I_H + I_L$. When we use neural network to extract the feature of data, the feature map will represent more high-frequency information for deeper network layer, while $I_L$ is also important for the reconstruction of $I$. Consequently, the CRB is design to transfer the low-frequency information from original low-dimensional data, which is equal to the $u(\mathbf{.})$ in the Laplacian pyramid. In the CRB of $s$-th RB, the lower-dimensional channel matrix is directly upscale to $X_s$ by the UB, where the filters of UB are set to 2. Meanwhile, the output of FEB reduces the channels of feature maps to 2 channels $R_s$ by CB with 2 filters, which be denoted transition block (TB) in the Fig. 2. Hence, the output of $s$-th RM $\widehat{\mathbf{H}}_s = \mathbf{X}_s + \mathbf{R}_s$.

### C. Multi-Scale Supervised Learning

In each RM, we adopt multi-scale supervised learning to generate the cascaded channel matrix with different scale. For the SR and channel estimation task, $L_1$ loss function can obtain better performance compared with $L_2$ loss function [11], but $L_1$ loss is not robust in the zero value. We adopted Charbonnier loss function to optimize the whole network [14], which is a differentiable variant of $L_1$ loss.

$$\begin{aligned}\mathcal{L}(\bar{\mathbf{H}}, \widehat{\mathbf{H}}) &= \frac{1}{B}\sum_{i=1}^{B}\sum_{s=1}^{S}\rho\left(\bar{\mathbf{H}}_s^{(i)} - \widehat{\mathbf{H}}_s^{(i)}\right) \\ &= \frac{1}{B}\sum_{i=1}^{B}\sum_{s=1}^{S}\rho\left(\bar{\mathbf{H}}_s^{(i)} - \mathbf{X}_s^{(i)} - \mathbf{R}_s^{(i)}\right), \end{aligned} \tag{8}$$

where $\rho(x) = \sqrt{x^2 + \varepsilon^2}$ is the Charbonnier penalty function and $\varepsilon$ is a regularization parameter. $B$ is the number of training sample in each batch.

### D. Parameters and Computational Complexity Analysis

Let the number and size of the filter is $w_1$ and $k_1 \times k_1$ for the first convolutional layer in the RB, while the number and size of filter is $w_2$ and $k_2 \times k_2$ for the second convolutional layer, respectively. In the vanilla RB of EDSR, $w_1 = w_2$ and $k = k_1 = k_2$, so the parameters of a RB are $2{w_1}^2k^2$. In the WARB, we introduce an expansion factor $r$ before activation layer, i.e., $\widehat{w}_1 = r\widehat{w}_2$, so the parameters of a WARB are $2r\widehat{w}_2^2k^2$. Since it proves nothing except that more parameters lead to better performance, we keep the parameters is same for two RBs, i.e., $2{w_1}^2k^2 = 2r\widehat{w}_2^2k^2$, where the computational complexity is a constant scaling of parameters for the same input size. Consequently, $\widehat{w}_2$ should be slimmed for the same parameters and computational complexity of two RBs, i.e., $\widehat{w}_2 = \frac{w_2}{\sqrt{r}}$.

The time complexity of FEB, CRB and TB in the $s$-th RM is $\mathrm{O}\left(2^sMk^2p(Ww_1w_2)\right)$, $\mathrm{O}\left(2^{s+1}Mk^2p\right)$ and $\mathrm{O}\left(2^sMk^2pw_3\right)$. Consequently, the total time complexity of LapWRes is $\mathrm{O}\left(\sum_{s=1}^{S} 2^sMk^2p(Ww_1w_2 + w_3 + 2)\right)$. The space complexity is composed of parameters and computed feature map, which is $\mathrm{O}\left(\sum_{s=1}^{S} k^2(2Ww_1w_2) + 2^{s-1}Mp(W(w_1 + w_2)))\right)$.

## IV. Numerical Results

### A. Simulation Setting

In our simulation, $M = 8\times 8$, $N = 16\times 16$, $k = 8$, $S = 3$ and $\varepsilon = 10^{-6}$. The Poisson and uniform distribution are used to model the distribution of cluster $C \sim \max\{P(\lambda_p), 1\}$ and scatters of each cluster $S_c \sim \mathcal{U}[1, 30]$ [16], respectively. The mmWave communication frequency is set to 28 GHz and the parameter $\lambda_p = 1.8$ [20]. We consider the path loss model in Indoor Hotspot (InH) Indoor Office scenario, that is $n = 3.19$, $b = 0.06$, $\sigma = 8.29$ dB and $f_0 = 24.2$ GHz for InH Office-NLOS, while $n = 1.73$, $b = 0.06$ and $\sigma = 3.02$ dB for InH Office-LOS [18]. We generate 30000 paired channel samples to construct the dataset.

We train the networks for 100 epochs using the cosine learning rate decay with an initial learning rate $5\times 10^{-3}$ [21], and the networks are optimized using adaptive moment estimation optimizer in the training process [22]. The regularization parameter $\varepsilon = 10^{-6}$. The normalized mean squared error (NMSE) is used as the performance metric, i.e., $\mathrm{NMSE} = \mathbb{E}[||\widehat{\mathbf{H}} - \mathbf{H}||_F^2/||\mathbf{H}||_F^2]$. Let $r$ denotes the ratio of the number of the activated RIS elements to the total elements in the LS estimation, i.e., $r = \frac{P}{N} = \frac{1}{k}$, where $P$ pilots is used for LS-ON/OFF algorithm.

### B. Result and Analysis

We compared the proposed scheme with traditional algorithms, e.g., LS [23] and OMP [4], and other SR networks, e.g., SRCNN [12], EDSR [11] and LapSRN [14]. In the SRCNN and EDSR, the UB is design at the input and output layer of network, respectively. SRCNN1 is an improvement on vanilla SRCNN by adjusting the upsampling factors, where the single-step up-sampling in the SRCNN is modified to the asymptotic sampling with 2 times factor. For the fair comparison of different networks, the number of RBs is the same for EDSR and LapWRes.

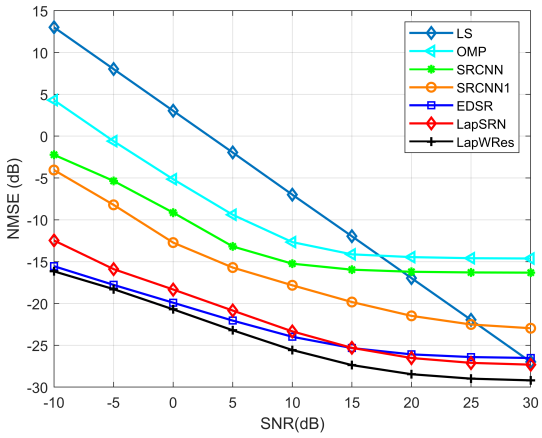


Fig. 3. NMSE performance for different algorithms.

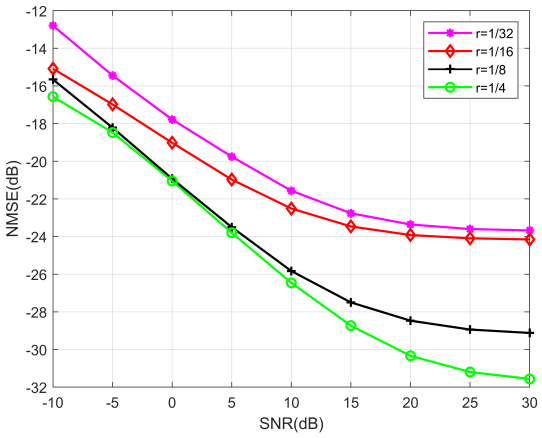


Fig. 5. NMSE performance for different pilot overhead.

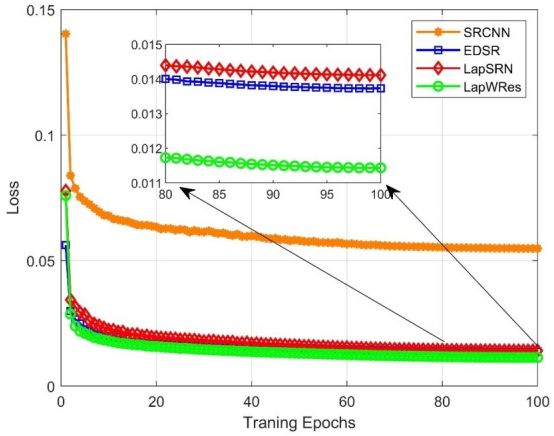


Fig. 4. Convergence performance for different network models.

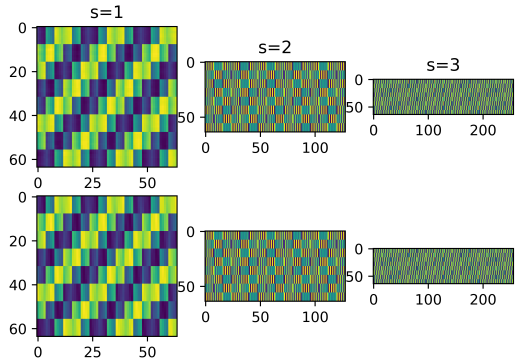


Fig. 6. Visualization of $S$ RMs' output when SNR=10dB. The three channel matrix in the first row represent predicted real parts of cascaded channel matrix, while the second row represent the corresponding label in different scales.

Fig. 3. shows the NMSE performance of LapWRes with less pilot overhead is superior to traditional LS and OMP algorithm. In the LS estimator and OMP algorithm, the required pilot overhead is set to $P_{\text{LS}} = N$ and $P_{\text{OMP}} = N/2$, respectively, while the required pilot overhead is $P = N/2^S = N/8$ for SR-based channel estimation networks. The SR network-based channel estimation performance is related with the method and location of upsampling. In the SRCNN, the single-step upsampling, i.e., $\bar{\mathbf{H}}^{\mathcal{P}} \in \mathbb{R}^{M\times P\times 2} \rightarrow \bar{\mathbf{H}}_S \in \mathbb{R}^{M\times N\times 2}$, will introduce serve interpolation errors in the input layer, which results in limited recovery effect of the subsequent network, and the high-dimensional input also increases the computational complexity of the network. SRCNN1 adopt the UB with a certain up-sampling factor to progressive upscale to the complete dimension of $M \times N \times 2$, which reduces the interpolation error of input data to a certain extent. In the EDSR, a large number of RBs are stacked before upsampling, and then the extracted efficient features are used to the last reconstruction. This post-upsampling architecture can reduce the computational complexity and improve the reconstruction performance. However, the post-upsampling layer is difficult to recover the high-resolution cascaded channel matrix directly when the upsampling factor is large. In the proposed LapWRes, we further optimize the location of up-sampling, where the UB is embedded in the network from low dimension to high dimension to realize progressive reconstruction of cascaded channel matrix. The existing LapSRN simply stack convolutional layer in the FEB, while we introduce WARBs into the FEB to increase the network capacity in the LapWRes.

Fig. 4. shows the convergence of different network models in the training process, where the average Charbonnier loss between generated cascaded channel matrix and ground truth is used as the measurement. Since the residual learning can avoid the gradient disappearance and improve the transmission of information flow [24], the EDSR can obtain faster convergence than vanilla LapSRN by stacking a number of RBs. In the proposed LapWRes, we merge the WARBs and multi-scale supervised learning to accelerate the network convergence, so the convergence level of LapWRes is superior to other SR networks.

Fig. 5. shows the NMSE performance of LapWRes under different $r$. The baseline LapWRes composes of 3 RMs and the upscale factor of each RM is 2, which matches $r = \frac{1}{8}$. If $r < \frac{1}{8}$, we increase the UBs in the first RM, e.g., adding 1 UB when $r = \frac{1}{16}$. Conversely, we delete partial RMs to match larger $r$, e.g., deleting 1 UB when $r = \frac{1}{4}$. With the decrease of $r$, the required upsampling dimension will be larger, so the reconstruction difficult will be increased. However, LapWRes can achieve satisfactory channel estimation accuracy even with little pilot overhead, e.g., $p = Nr = 8$. In Fig. 6., we visualize the output channel matrix of each RM in LapWRes, which shows each RM can learning the data distribution of channel in different scales.

## V. Conclusion

In this paper, we leverage the SR-based channel extrapolation idea to realize the cascaded channel reconstruction. Different with the one-step reconstruction used in previous works, we propose a progressive reconstruction strategy by utilizing the multi-scale supervised learning. The proposed LapWRes adopt dual branch architecture to extract the high frequency and low frequency information of cascaded channel matrix, respectively, and then the residual learning is used to realize information fusion. Numerical results show that the proposed channel estimation model with limited pilot overhead outperforms other estimation schemes. In the future works, we will extend the proposed model to higher-dimensional channel estimation scenarios, e.g., cooperative communication of multiple RISs.

## Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 62101205, in part by the Key Research and Development Program of Hubei Province under Grant 2021BAA170, in part by the Natural Science Foundation of Hubei Province under Grant 2021CFB248, and in part by the Fundamental Research Funds for the Central Universities of China under grant CCNU20QN004.

## References

[1] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 3rd Quart., 2018.

[2] 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.

[3] B. Zheng, C. You, W. Mei and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1035-1071, Secondquarter 2022.

[4] X. Chen, J. Shi, Z. Yang, and L. Wu, "Low-complexity channel estimation for intelligent reflecting surface-enhanced massive MIMO," *IEEE Wireless Commun. Lett.,* vol. 10, no. 5, pp. 996–1000, May 2021.

[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.

[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] 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. 92239228, Jun. 2020.

[8] S. Zhang, S. Zhang, F. Gao, J. Ma and O. A. Dobre, "Deep learning-based RIS channel extrapolation with element-grouping," *IEEE Wireless Commun. Lett.,* vol. 10, no. 12, pp. 2644-2648, Dec. 2021.

[9] W. Xie, J. Xiao, P. Zhu, C. Yu and L. Yang, "Deep compressed sensing-based cascaded channel estimation for RIS-aided communication systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 846-850, April 2022.

[10] 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.

[11] L. Li, H. Chen, H. -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.

[12] Y. Jin, J. Zhang, X. Zhang, H. Xiao, and B. Ai, D. W. K. Ng, "Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks," *IEEE Trans. Veh. Technol.,* vol. 69, no. 8, pp. 9223–9228, Jun. 2020.

[13] 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.

[14] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* 2017, pp. 624–632.

[15] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, T. Huang,"Wide activation for efficient and accurate image super-resolution," *arXiv preprint arXiv:1808.08718*, 2018.

[16] 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.

[17] P. Nayeri, F. Yang, and A. Z. Elsherbeni, "Reflectarray antennas: theory, designs, and applications. " *USA: Wiley,* 2018.

[18] "5G channel model for bands up to 100 GHz,"[Online]. Available: http://www.5gworkshops.com/5GCMSIG White.

[19] T. Salimans and D. P. Kingma. "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," in *Proc. Neural Informat. Process. Syst. (NIPS),* 2016, pp. 901–909.

[20] 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.

[21] I. Loshchilov, F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Mach. Learn. (ICLR)*, May 2017.

[22] D.P. Kingma, J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Mach. Learn. (ICLR)*, Jan. 2014.

[23] 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)*, May 2019, pp. 4659–4663.

[24] K He, X Zhang, S Ren and J Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770-778.