

# LSRN: A Recurrent Residual Learning Framework for Continuous Wireless Channel Estimation Using Super-Resolution Concept

SHUNQING ZHANG<sup>ID</sup>, (Senior Member, IEEE), YANGYU LIU<sup>ID</sup>, QI SHI<sup>ID</sup>,  
SHUGONG XU<sup>ID</sup>, (Fellow, IEEE), AND SHAN CAO<sup>ID</sup>

Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China

Corresponding author: Shugong Xu (shugong@shu.edu.cn)

**ABSTRACT** As only a few parts of wireless resources can be utilized for pilot transmission, channel estimation, especially the interpolation process, has often been recognized as a challenging ill-posed reconstruction problem. To deal with this task, we formulate it as a typical image super resolution problem, and propose a recurrent residual learning framework named LSRN. Our proposed scheme jointly utilizes the advantages of recurrent and residual structure in the machine learning area to approximate the non-linear interpolation relations between the reference signal and surrounding resource elements. In addition, we propose a low complexity implementation scheme called LSRN-L to address the stringent processing delay requirement in the channel estimation tasks. Through numerical examples as well as prototype verification, the proposed LSRN/LSRN-L can easily outperform the convolutional GI plus DFT based interpolation scheme by 10dB in terms of normalized mean square error. Meanwhile, the low complexity LSRN-L can maintain the processing delay within one millisecond.

**INDEX TERMS** Channel estimation, channel state information, super resolution, residual learning and recurrent learning.

## I. INTRODUCTION

Evolved mobile broadband (eMBB) transmission has been identified as one of the most important scenarios in the fifth generation (5G) communication systems [1]. To guarantee ultra-high throughput in the eMBB scenario, orthogonal frequency division multiplexing (OFDM) transmission with coherent detection is often regarded as the main approach in the frequency selective and time varying wireless environment, and the high-order modulation, such as 64 quadrature amplitude modulation (QAM), is widely adopted to boost the throughput [2]. In order to accurately detect the high-order modulated symbols, an efficient yet fast channel estimation is often regarded as the most important stage and has been investigated for more than twenty years [3].

As a few part of resources can be utilized for pilot transmission in modern wireless communications, channel estimation has been recognized as a challenging ill-posed reconstruction problem with only a small amount of observations (pilots) [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Fang Yang<sup>ID</sup>.

In order to solve this type of complicated problem within the channel coherence time, a standard channel estimation process usually contains a channel state recovery stage and a low complexity interpolation stage as illustrated in [5]. Given the additive white Gaussian noise (AWGN) assumption, the conventional channel state information (CSI) recovery schemes at the pilot locations usually apply the least square (LS) or the minimum mean squared error (MMSE) algorithm, which has been proved to be efficient in practice [6]. However, the low complexity interpolation mechanisms are under-investigated, where traditional schemes, such as linear interpolation (LI) [5], gaussian interpolation (GI) [7], or discrete fourier transform interpolation (DFTI) [8], are still used in recent communication systems.

Since the interpolation process to recover the whole channel frequency responses from some pilot observations can be regraded as a typical image super-resolution (SR) task [9], it is of great importance to review the history of classical SR methods and the potential applications in the channel estimation procedure. For example, by assuming the deterministic linear/non-linear relations between the low resolution (LR)

and the high resolution (HR) images, bilinear or bicubic interpolation schemes have been utilized to improve the quality of SR process in [10], and the reconstruction-based schemes, such as [11] and [12], have been proposed to further compensate the remaining high-frequency components. Instead of predicting the possible linear/non-linear relations between LR and HR image pairs through mathematical formulas, another approach is to directly learn the relations via some pre-known data-sets. Typical learning frameworks include dictionary learning [13], local linear regression [14], and random forest algorithm [15]. Although the above schemes have been applied to practical systems in [16]–[18], the resultant performance gain in terms of the estimation accuracy and the complexity reduction is not satisfactory.

Recently, with the development of the computing technology and the numerical calculating technology, deep neural networks [19] have been widely applied in the learning field, and achieve significant progresses in the general SR tasks. As summarized in [20], they can be fulfilled by several kinds of advanced network architectures, including recursive [21] and progressive reconstruction [22]. However, the above network designs often suffer from high computational complexity and are rarely deployed on low-cost computing devices, such as mobile terminals. Instead, linear [23]–[25] and residual networks [26] are more popular due to their simple network structures. For example, super-resolution convolutional neural network (SRCNN) and very deep super-resolution network (VDSR) have been proposed in [23] and [24] respectively, which up-sample the LR figures before the feature extraction process, while fast SRCNN [25] performs the up-sampling afterwards to reduce the processing delay of the entire SR task. All of them are based on linear neural network topology and able to achieve peak signal-to-noise ratio (PSNR) for more than 30 dB in the DIV2K public data set. In [26], enhanced deep super-resolution (EDSR) scheme has been proposed to apply the residual learning block and achieves 2-4 dB PSNR gain in the same data set. In addition to the above single image SR tasks, multi-frame SR tasks have been widely studied as well, including detail-revealing deep video SR [27] and temporally coherent generative adversarial networks video SR [28]. Since the interpolation process in the wireless channel estimation is quite similar to the image SR task, there are some initial research efforts on the SR based channel estimation, which provides superior performance gain in terms of the normalized mean square error (NMSE) according to [9] and [29]. Nevertheless, to propose a suitable SR based learning framework is never straight forward due to the following reasons.

- *Stringent Processing Delay* Different from the conventional image SR, the wireless channel estimation usually has stringent requirements on the processing delay. For example, in the commercial long term evolution (LTE) networks, the channel estimation needs to be performed on a sub-frame (10 milliseconds duration) basis and the corresponding delay budget is less than one millisecond in general. Hence, how to strike a balance between

the processing delay and the channel estimation performance using the SR learning framework, needs to be carefully investigated.

- *Domain Knowledge Exploration* The specific prior knowledge has been proved to be very effective for some tailored SR applications. For example, the spectral sparsity and the correlation information among consecutive frames have been explored for the hyper-spectral image SR and the video streaming SR tasks in [30] and [27], respectively, which shows significant performance gain in terms of PSNR. Since the wireless channel matrix has some slow-varying properties in the time domain as we shown later, how to explore this type of domain knowledge to further improve the estimation accuracy and reduce the processing delay based on the SR learning framework is still open.
- *High Generalization Ability* In the practical communication system, to adjust the learning framework based on different channel fading environments is quite challenging, and a more reasonable solution is to design a more robust network which only needs offline training. Therefore, the desired SR based learning framework should provide high generalization ability as well.

In our previous work [9], to address the above challenges, we have proposed a novel channel estimation scheme using the super resolution image recovery concept, which achieves significant performance improvement based on the numerical evaluations. However, one drawback of this scheme is the associated significant processing delay. Our current work is proposed to directly address this issue and the novelty of the current scheme can be summarized as follows. First, we propose to use the recurrent architecture together with the traditional residual learning task to balance the achieved NMSE performance and the processing delay, which is a novel concept based on our literature survey. The proposed recurrent residual learning framework, named *LSRN* (Long-Short term memory based Residual Network), which jointly exploits the advantages of recurrent and residual architectures on top of the traditional image SR tasks. Specifically, we apply the residual network model to approximate the non-linear interpolation relations of real-time CSI between the reference signals (RSs) and the neighboring resource elements (REs) to improve the estimation accuracy, and utilize the recurrent structure to learning the slow-varying time domain correlation among consecutive OFDM symbols. In order to control the processing delay, we simplify the original structures of residual blocks (ResBs), and reduce the number of cascaded blocks and convolutional filters simultaneously. We denote the corresponding low complexity implementation as *LSRN-L*. Second, we jointly use the classical channel model and the prototype measurement results to elaborate the effectiveness of the proposed scheme, which is much more valuable for practical implementations. We offline train the *LSRN-L* using the standard COST 2100 channel model [31] and deploy it online in the commercial WiFi system using OFDM transmissions. Based on some numerical

simulations and prototype verification, we show that the proposed LSRN-L can provide 10 dB to 11 dB NMSE improvement if compared with the conventional GI plus DFTI based interpolation schemes and preserve a processing delay within one millisecond.

The rest of this article is organized as follows. In Section II, we provide some preliminary information on the channel estimation and the machine learning empowered SR schemes. The proposed LSRN framework is discussed in Section III and the corresponding low-complexity implementation strategy LSRN-L is proposed in Section IV. The corresponding numerical and empirical examples are provided in Section V, and in Section VI, we conclude the whole paper.

## II. PRELIMINARIES

In this section, we briefly introduce the basic concept of interpolation in the channel estimation process and the existing SR schemes.

### A. INTERPOLATION FOR CHANNEL ESTIMATION

Consider an OFDM transmission system with a single transmit and  $N_r$  receive antennas in the wireless fading environment.<sup>1</sup> With the AWGN assumption, the received symbols at the  $(f_{RS}, t_{RS})^{th}$  RE of the  $k^{th}$  resource block (RB),  $\mathbf{y}^k(f_{RS}, t_{RS}) \in \mathbb{C}^{N_r \times 1}$ , can be modeled through,  $\mathbf{y}^k(f_{RS}, t_{RS}) = \mathbf{h}^k(f_{RS}, t_{RS})x^k(f_{RS}, t_{RS}) + \mathbf{n}^k(f_{RS}, t_{RS})$ , where  $x^k(f_{RS}, t_{RS})$  and  $\mathbf{h}^k(f_{RS}, t_{RS}) \in \mathbb{C}^{N_r \times 1}$  are the predefined RS and the equivalent channel responses, respectively, and  $\mathbf{n}^k(f_{RS}, t_{RS}) \in \mathbb{C}^{N_r \times 1}$  denotes the additive complex Gaussian noise with mean  $\mathbf{0}_{N_r}$  and variance  $\sigma^2 \mathbf{I}_{N_r}$ . Based on the famous MMSE criteria, the estimated channel state,  $\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})$ , is given by

$$\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}) = \mathbf{R}_h \left[ \mathbf{R}_h + \frac{\sigma^2}{|x^k(f_{RS}, t_{RS})|^2} \mathbf{I}_{N_r} \right]^{-1} * \frac{x^{k,*}(f_{RS}, t_{RS})}{|x^k(f_{RS}, t_{RS})|^2} \mathbf{y}^k(f_{RS}, t_{RS}),$$

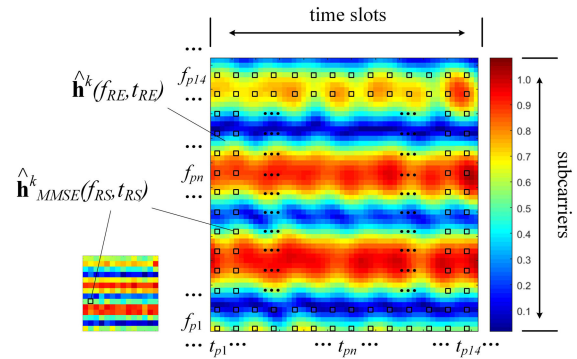
where  $\mathbf{R}_h = \mathbb{E}[\mathbf{h}^k(f_{RS}, t_{RS})\mathbf{h}^{k,H}(f_{RS}, t_{RS})]$  is the channel correlation matrix<sup>2</sup>

In the practical system, to minimize the channel estimation overhead, an interpolation process is usually performed on a RB basis (in accordance with the coherence time and coherence bandwidth) with  $N_t$  time slots and  $N_f$  sub-carriers as shown in Figure 1, which maps the estimated channel state of RSs into the entire RB. Denote  $\mathbf{h}^k$  to be the aggregated CSI of the  $k^{th}$  RB and the interpolation process can be described as,

$$\mathbf{h}^k = \mathcal{F} \left( \{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}, \forall (f_{RS}, t_{RS}) \in \Omega_{RS} \right), \quad (1)$$

<sup>1</sup>For illustration purpose, we denote the boldface upper-case and lower-case to be a matrix and a vector.  $(\cdot)^T$ ,  $(\cdot)^*$ ,  $(\cdot)^H$ , and  $(\cdot)^{-1}$  denote the matrix/vector transpose, conjugate, Hermitian, and inversion operations, respectively. The identity matrix with dimension  $N$  is denoted by  $\mathbf{I}_N$ , and we apply  $\mathbf{0}_N$  and  $\mathbf{1}_N$  to denote all zero and all one vectors with dimension  $N$ .

<sup>2</sup>In this paper, we assume the channel correlation matrix and the noise covariances are pre-known based on the previous long-term observations.



**FIGURE 1.** One RB contains  $N_f$  sub-carriers in frequency domain and  $N_t$  time slots in time domain. The black grid position indicates the pilot position, and the channel condition of the other grid is traditionally estimated by the interpolation mechanism.

where  $\Omega_{RS}$  represents the collection of all the possible RS positions in each RB,  $\mathcal{F}(\cdot)$  denotes the general interpolation function mapping from  $|\Omega_{RS}|$  RSs to  $N_t \times N_f$  REs.

In the existing literature, the interpolation processes rely on some heuristic algorithms, and a standard approach is to jointly utilize the DFTI process in the frequency domain and Gaussian approximation in the time domain as defined in [7], [8]. Based on (1), the above interpolation process can be rewritten as,

$$\begin{aligned} \hat{\mathbf{h}}_{GI+DFTI}^k &= \mathcal{F}_{GI+DFTI} \left( \{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}, \right. \\ &\quad \left. \forall (f_{RS}, t_{RS}) \in \Omega_{RS} \right) \\ &= g_{DFTI} \left( f_{GI}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}, \right. \\ &\quad \left. \forall (f_{RS}, t_{RS}) \in \Omega_{RS} \right), \quad (2) \end{aligned}$$

where  $f_{GI}(\cdot)$  and  $g_{DFTI}(\cdot)$  denote the Gaussian approximation and DFTI processes as illustrated in [32], respectively.

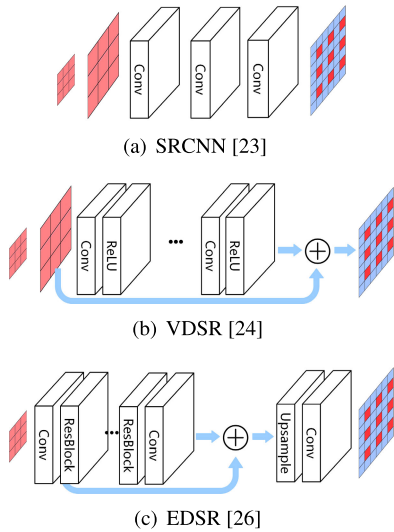
As shown before, the traditional method approximates the original function  $\mathcal{F}(\cdot)$  by cascading several deterministic linear functions together, e.g.,  $g_{DFTI}(\cdot)$  and  $f_{GI}(\cdot)$ , while the inherent relations between  $\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}$  and  $\mathbf{h}^k$  are in general nonlinear due to the additive random noises and the non-deterministic relations of CSIs from different REs.

### B. EXISTING SR SCHEMES

In order to solve the above ill-conditioned problem as defined in (1), a well known approach as mentioned before is to apply the single image SR as shown in Figure 2. Based on the existing literature, three types of CNN-based SR schemes are commonly adopted, which are summarized as below.

- **SRCNN** [23]: In the first stage, the original LR images are expanded to the target size, and then using three-layers neural networks to approximate the non-linear relation between LR images and HR images.

<sup>3</sup> $|\cdot|$  represents the cardinality of the inner set.

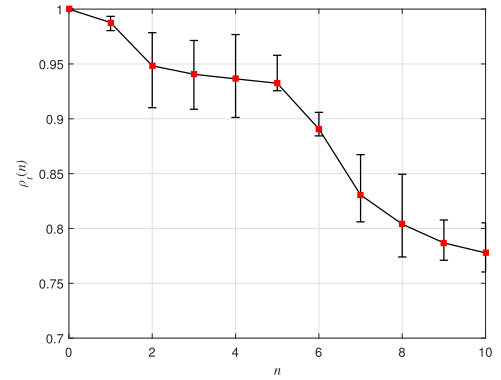


**FIGURE 2.** The network architecture of three typical SR neural networks, where subfigure (a), (b) and (c) demonstrates the abstracted network configuration of SRCNN [23], VDSR [24] and EDSR [26], respectively. Different from the traditional interpolation scheme, they approximate the non-linear interpolation via neural networks.

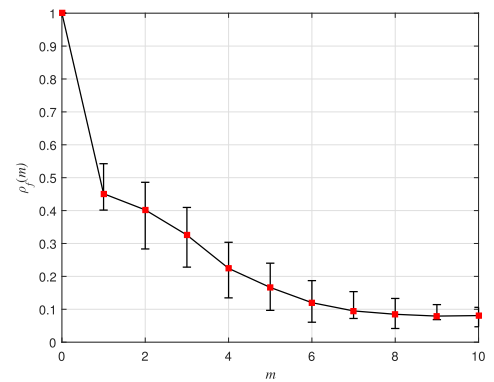
- VDSR [24]: The first stage of VDSR is similar as SRCNN. In feature extraction stage, VDSR apply the global residual learning network to progressively predict the SR images.
- EDSR [26]: Different from SRCNN and VDSR, EDSR directly input the LR image, and then apply the global and local residual learning network [33] to progressively predict the SR images from LR images.

Since the existing SR schemes are directly designed for image processing, there are still many challenges and opportunities to explore when applied to the channel estimation tasks. Despite the stringent processing delay and the generalization requirement as aforementioned, the time and frequency domain correlations among neighboring REs might be able to exploited as well. If we define  $\rho_t(n) = \mathbb{E} \left[ \frac{(\mathbf{h}^k(f_{RE}, t_{RE}))^* \mathbf{h}^k(f_{RE}, t_{RE} + n\Delta t)}{|\mathbf{h}^k(f_{RE}, t_{RE})|^* |\mathbf{h}(f_{RE}, t_{RE} + n\Delta t)|} \right]$  and  $\rho_f(m) = \mathbb{E} \left[ \frac{(\mathbf{h}^k(f_{RE}, t_{RE}))^* \mathbf{h}(f + m\Delta f, t_{RE})}{|\mathbf{h}^k(f_{RE}, t_{RE})|^* |\mathbf{h}(f_{RE} + m\Delta f, t_{RE})|} \right]$  to be the time and frequency domain correlations,<sup>4</sup> we can plot the relation between the correlation coefficients and the number of intervals in Figure 3. The red dots in the figure represent the average values, the maximum and minimum values are indicated by a dash line. As can be seen from Figure 3, the correlation of the channel in time domain is stronger than the correlation in frequency domain and the channel changes are subtle and gradual over a period of time. Using this type of domain knowledge, we can further improve the estimation accuracy and reduce the processing delay by improving the existing network modules and adding related modules. From Section II-B of the paper, we conclude that the change of channel is subtle and gradual over a period of time, and the time-domain correlation of channel is much stronger than

<sup>4</sup>We choose the time interval and frequency intervals according to the IEEE 802.11n standard, which is 10ms and 312.5KHz, respectively.



(a) time correlation coefficients



(b) frequency correlation coefficients

**FIGURE 3.** Schematic diagram of channel time/frequency correlation, subfigure(a) represents the temporal correlation in the time domain and subfigure(b) represents the frequency correlation in the frequency domain.

its frequency-domain correlation. Using this kind of domain knowledge, we think it is necessary to introduce recurrent learning for channel estimation.

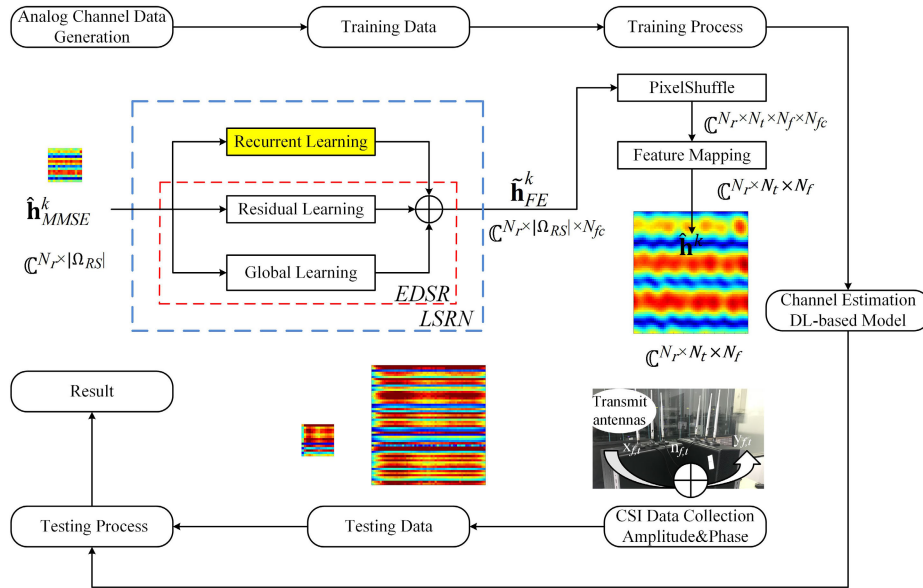
### III. LSRN FRAMEWORK

With the strong ability to model the non-linear function  $\mathcal{F}(\cdot)$ , the interpolation for channel estimation can be recognized as a single image SR reconstruction problem. In this section, we provide an overview of the proposed LSRN framework and then introduce block by block to demonstrate the effectiveness of the proposed scheme.

#### A. OVERVIEW

The primary target of using SR technology is to find a suitable interpolation function  $\mathcal{F}_{SR}$  by minimizing the potential MSE between the estimated CSI,  $\hat{\mathbf{h}}_{SR}^k$ , and the aggregated CSI,  $\mathbf{h}^k$ . Mathematically, the optimal SR reconstruction for CSI interpolation,  $\mathcal{F}_{SR}^*$ , can be modeled via the following problem.

$$\begin{aligned} \mathcal{F}_{SR}^*(\cdot) = \arg \min_{\mathcal{F}_{SR}(\cdot)} \lim_{N_{RB} \rightarrow \infty} \frac{1}{N_{RB}} \sum_{k=1}^{N_{RB}} \|\mathbf{h}^k - \hat{\mathbf{h}}_{SR}^k\|_F^2, \\ \text{subject to } \hat{\mathbf{h}}_{SR}^k = \mathcal{F}_{SR}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}, \\ \forall (f_{RS}, t_{RS}) \in \Omega_{RS}), \quad (3) \end{aligned}$$



**FIGURE 4.** Our system architecture. The whole process can be divided into training phase and testing phase.

where  $\|\mathbf{A}\|_F = \sqrt{\text{Tr}(\mathbf{A} \cdot \mathbf{A}^H)}$  denotes the Frobenius norm and  $N_{RB}$  denotes the total number of RBs. In the practical system, the true value of channel responses,  $\mathbf{h}^k$ , is difficult to observe in general due to the noisy environment, and a more feasible solution is to approximate the original  $\mathcal{F}_{SR}^*$  by solving the following problem.

$$\mathcal{F}_{SR}^*(\cdot) \approx \arg \min_{\mathcal{F}_{SR}(\cdot)} \lim_{N_{RB} \rightarrow \infty} \frac{1}{N_{RB}} \sum_{k=1}^{N_{RB}} \|\hat{\mathbf{h}}_{MMSE}^k - \hat{\mathbf{h}}_{SR}^k\|_F^2, \quad \text{subject to } \hat{\mathbf{h}}_{SR}^k = \mathcal{F}_{SR}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}, \forall (f_{RS}, t_{RS}) \in \Omega_{RS}). \quad (4)$$

Note that in the above formulation, the aggregated CSI,  $\mathbf{h}^k$ , is replaced by the estimated CSI using the conventional MMSE scheme, e.g.,  $\hat{\mathbf{h}}_{MMSE}^k$ , which is much easier to be implemented in practice. Since  $\mathcal{F}_{SR}$  is an ill-conditioned mapping from  $|\Omega_{RS}|$  RSs to  $N_t \times N_f$  REs, the above optimization problem is in general difficult to solve.

Inspired by the EDSR scheme as illustrated before, we divide the entire process into two steps, where the first step focuses on the feature extraction and fusion processes to obtain several  $N_{fc}$ -dimension features, and the second step scales up the  $N_{fc}$ -dimension features into  $N_t \times N_f$  REs by up-sampling and maps the high dimensional features back into the complex valued channel responses. As illustrated in Section II-B, the time domain correlations are much stronger than the frequency domain, which can be regarded as the domain knowledge for SR based wireless channel estimation tasks. To exploit this effect, we propose the LSRN framework as shown in Figure 4, where a recurrent learning block is applied to extract the time domain correlations. Mathematically, the proposed LSRN framework can be expressed as

follows.

$$\theta_{LSRN}^* \approx \arg \min_{\theta_{LSRN}} \lim_{N_{RB} \rightarrow \infty} \frac{\sum_{k=1}^{N_{RB}} \|\hat{\mathbf{h}}_{MMSE}^k - \hat{\mathbf{h}}_{LSRN}^k\|_F^2}{N_{RB}} \quad \text{subject to } \hat{\mathbf{h}}_{LSRN}^k = \mathcal{F}_{LSRN}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}; \theta_{LSRN}), \quad (5)$$

where  $\mathcal{F}_{LSRN}(\cdot; \theta_{LSRN})$  and  $\hat{\mathbf{h}}_{LSRN}^k$  denote the proposed LSRN with parameters  $\theta_{LSRN}$  and the corresponding estimated CSI, respectively.<sup>5</sup> Denote  $\mathcal{F}_{RL}(\cdot; \theta_{RL})$ ,  $\mathcal{F}_{LL}(\cdot; \theta_{LL})$ , and  $\mathcal{F}_{GL}(\cdot; \theta_{GL})$  to be the recurrent learning, the local residual learning, and the global residual learning blocks respectively, the extracted and fused  $N_{fc}$ -dimensional feature tensors,  $\{\hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS})\}$ , can be obtained via following formula,

$$\{\hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS})\} = \mathcal{F}_{RL}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}; \theta_{RL}) + \mathcal{F}_{LL}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}; \theta_{LL}) + \mathcal{F}_{GL}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}; \theta_{GL}). \quad (6)$$

Following the pixelShuffle network as proposed in [34],  $\mathcal{F}_{PS}(\cdot; \theta_{PS})$ , and the feature mapping operation,  $\mathcal{F}_{FM}(\cdot; \theta_{FM})$ , the estimated CSI by the proposed LSRN is finally given by

$$\hat{\mathbf{h}}_{LSRN}^k = \mathcal{F}_{FM}(\mathcal{F}_{PS}(\{\hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS})\}; \theta_{PS}); \theta_{FM}).$$

### B. RECURRENT LEARNING AIDED FEATURE EXTRACTION

Feature extraction has been proved as a key procedure in the image analysis tasks, including SR, de-blurring [35], and image semantic segmentation [36]. With the modern machine learning technology, the intrinsic features are automatically

<sup>5</sup>For illustration purpose, we omit  $\forall (f_{RS}, t_{RS}) \in \Omega_{RS}$  in the rest of this paper if no confusion is caused.

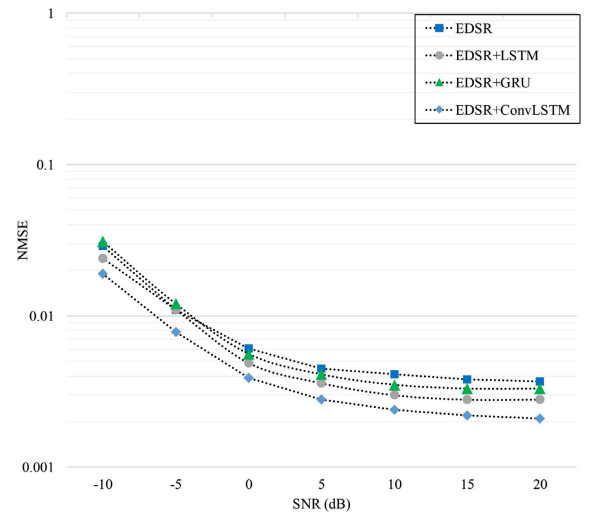
learned by feeding the “labelled” data-sets, and the residual learning approach, such as residual network (ResNet), is reported to achieve the state-of-the-art performance for feature extraction in the image recognition tasks [37]. The similar phenomenon has been observed in the image SR task as well, where the residual learning based EDSR networks achieves the state-of-the-art PSNR result. However, as a general-purpose SR network, the design of EDSR does not rely on any specific correlation models between LR and HR images, and the imbalanced correlation nature in the channel estimation tasks, as shown in Figure 1, is *NOT* considered. In order to take the advantage of this domain knowledge, we propose to use the recurrent learning on top of the original EDSR framework. Typical recurrent learning algorithms include long short term memory (LSTM) [38], gated recurrent unit (GRU) [39], and convolution LSTM (ConvLSTM) [40]. Since LSTM and GRU architectures are dedicated designed for one-dimensional fixed-length vector sequences, the two-dimensional correlation property of channel responses as shown in Section II-B can not be fully represented. To overcome this obstacle, we propose to use ConvLSTM structure, and the corresponding mathematical model to calculate  $\mathcal{F}_{RL}(\{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}); \theta_{RL}\})$  is shown below, where each extracted feature  $\hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS})$  is given by

$$\begin{aligned} & \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS}) \\ &= \sigma_1[\sigma_1(\theta_{RL}^3 + \theta_{RL}^2 \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 1) \\ &+ \theta_{RL}^1 \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})) * \sigma_2(\theta_{RL}^4 \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}) \\ &+ \theta_{RL}^5 \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 1) + \theta_{RL}^6) + C^{t_{RS}-1} * \sigma_2(\theta_{RL}^9 \\ &+ \theta_{RL}^8 \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 1) + \theta_{RL}^7 \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}))] * \sigma_2 \\ &(\theta_{RL}^{10} \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}) + \theta_{RL}^{11} \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 1) + \theta_{RL}^{12}), \end{aligned} \quad (7)$$

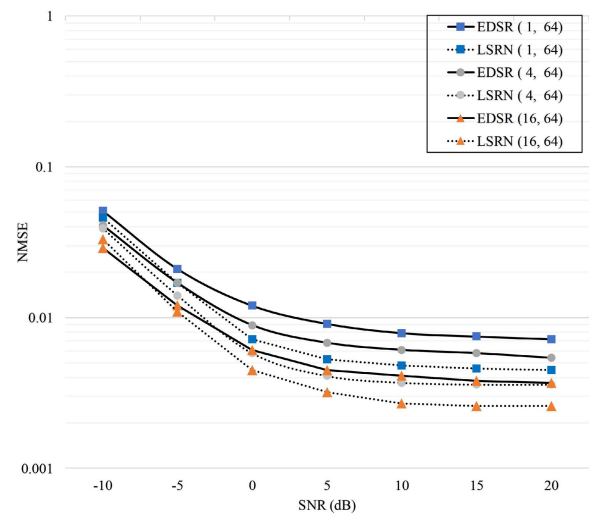
where  $\sigma_1(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  is hyperbolic tangent activation function and  $\sigma_2(x) = \frac{1}{1 + e^{-x}}$  is logistic sigmoid activation function, respectively.  $(t_{RS} - 1)$  represents the previous time slot before  $t_{RS}$ . The parameters of convolution  $\theta_{RL} = [\theta_{RL}^1, \dots, \theta_{RL}^{12}]$  are optimized during the training process. The cell state  $C^{t_{RS}-1}$  is iteratively updated by the following formula,

$$\begin{aligned} C^{t_{RS}-1} &= \sigma_1(\theta_{RL}^3 + \theta_{RL}^1 \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS} - 1) + \theta_{RL}^2 \\ &\hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 2)) * \sigma_2(\theta_{RL}^5 \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 2) + \theta_{RL}^6 \\ &+ \theta_{RL}^4 \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS} - 1)) + C^{t_{RS}-2} * \sigma_2(\theta_{RL}^9 \\ &+ \theta_{RL}^8 \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS} - 2) + \theta_{RL}^7 \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS} - 1)). \end{aligned} \quad (8)$$

To demonstrate the effectiveness of the proposed recurrent learning architecture, we perform the channel estimation tasks using the traditional EDSR network with three recurrent learning architectures, e.g. EDSR with LSTM, GRU, and ConvLSTM schemes, respectively. To make a fair comparison, the above schemes are trained and tested under the famous COST2100 channel models [31]. As shown



(a) Comparison of different recurrent learning structure



(b) Comparison of different network parameters.

**FIGURE 5.** NMSE comparison of different residual learning structure and parameters used in EDSR, subfigure(a) shows the different effects of LSTM [38], GRU [39] and ConvLSTM [40], subfigure(b) shows the different effects of different network parameters under COST 2100 model.

in Figure 5(a), the EDSR with ConvLSTM scheme outperforms the other three schemes under different received signal noise ratio (SNR) regions. In addition, to obtain a better understanding of the proposed scheme, we compare the channel estimation results with the traditional EDSR network under different number of residual blocks. As verified in Figure 5(b), the proposed scheme shows consistent NMSE improvement over all the tested network configurations.

### C. FEATURE RECONSTRUCTION

According to (6), the extracted and fused  $N_{fc}$ -dimensional feature tensors,  $\{\hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS})\}$ , can be expressed in detail as follows,

$$\begin{aligned} \hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS}) &= \hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS}) + \hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}) \\ &+ \mathcal{F}_{LL}(\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}); \theta_{LL}), \end{aligned}$$

where  $\hat{\mathbf{h}}_{RL}^k(f_{RS}, t_{RS})$  represents the features extracted by recurrent learning branch as defined in (7), and the global residual learning is simply a direct connection to the initial features, e.g.  $\mathcal{F}_{GL}(\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}); \theta_{GL}) = \{\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS})\}$ . For local residual learning, we use a 16-layer stacked residual block structure to model  $\mathcal{F}_{LL}(\hat{\mathbf{h}}_{MMSE}^k(f_{RS}, t_{RS}); \theta_{LL})$  which has been proved to be effective in SR cases [26]. After the feature extraction process, the pixelShuffle procedure up-scales the feature tensors  $\{\hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS})\}$  to the entire time-frequency resources. Typical reconstruction methods, as shown in Figure 6, gradually up-scale the feature tensors via the pixel re-arrangement or the deconvolution approach. Inspired by a direct up-scaling operation with the factor of 4, we extend this operation to more general cases with the factor of  $s^2$ . Specifically, we rearrange the feature tensor to the size of  $N_f \times N_t$ . If we denote

$$\hat{\mathbf{h}}_{FE}^k(f_{RS}, t_{RS}) = \begin{bmatrix} a_1(f_{RS}, t_{RS}) & \cdots & a_{N_{\hat{h}}}(f_{RS}, t_{RS}) \end{bmatrix},$$

the upscaled feature map  $\{\hat{\mathbf{h}}_{PS}^k(f_{RS}, t_{RS})\}$  can be obtained through,

$$\hat{\mathbf{h}}_{PS}^k(f, t) = \begin{cases} a_1(f_{RS}, t_{RS}), & f = (f_{RS} - 1) \times s + 1, \\ & t = (t_{RS} - 1) \times s + 1 \\ \vdots, & \vdots \\ a_s(f_{RS}, t_{RS}), & f = f_{RS} \times s, \\ & t = (t_{RS} - 1) \times s + 1 \\ a_{s+1}(f_{RS}, t_{RS}), & f = (f_{RS} - 1) \times s + 1, \\ & t = (t_{RS} - 1) \times s + 2 \\ \vdots, & \vdots \\ a_{2s}(f_{RS}, t_{RS}), & f = f_{RS} \times s, \\ & t = (t_{RS} - 1) \times s + 2 \\ a_{2s+1}(f_{RS}, t_{RS}), & f = (f_{RS} - 1) \times s + 1, \\ & t = (t_{RS} - 1) \times s + 3 \\ \vdots, & \vdots \\ a_{s^2}(f_{RS}, t_{RS}), & f = f_{RS} \times s, \\ & t = t_{RS} \times s \end{cases}$$

where  $s^2 = N_f \times N_t / |\Omega_s|$  is the period shuffle factor and the corresponding feature channel per pixel after up-scaling is  $N_{fc}/s^2$ .

We compared the experimental effects and time delay under these three reconstruction models as shown in Figure 6, the effects of the three reconstruction methods are relatively close. Due to the simpler structure, we choose pixelShuffle (4) which can effectively reduce the time delay.

The last step is a simple feature mapping process which maps the  $N_{fc}/s^2$  feature tensored into the final channel responses, and the corresponding realization  $\mathcal{F}_{FM}(\cdot; \theta_{FM})$  is performed by a two-layer convolutional neural network.

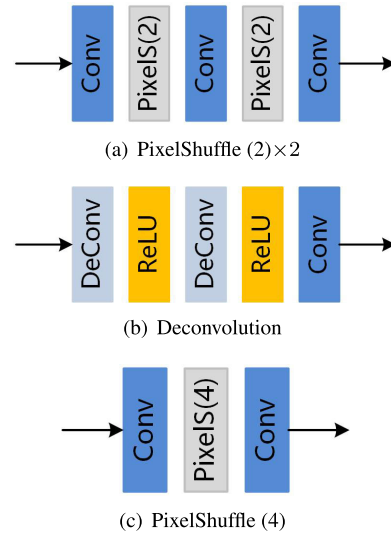


FIGURE 6. Comparison of some common image reconstruction structure (x4).

The detailed network configuration of the proposed LSRN structure is summarized in Table 1.<sup>6</sup>

#### IV. LOW COMPLEXITY IMPLEMENTATION

In the previous section, we designed a recurrent learning based LSRN scheme to recover the channel states. However, as the channel estimation is typically a delay critical task, the process delay of the proposed scheme needs to be carefully studied as well. As shown in Figure 7, more than 60% of the processing time is occupied by the residual learning process. Therefore, in this section, we focus on proposing a low complexity implementing strategy to satisfy the potential delay requirement.

##### A. SIMPLIFIED RESB STRUCTURE

As shown in Figure 4, the three parallel learning paths determine the processing delay of the feature extraction stage, and according to the existing literature [26], the local residual learning has the largest critical path, which dominates the entire delay chain. To solve this issue, a straight forward approach is to simplify the basic building block in the original ResNet architecture. Inspired from [41] and [26], we proposed to remove one activation layer and one convolutional layer as shown in Figure 9 and summarize the theoretical explanation in the following lemma.

*Lemma 1:* For a common residual learning network architecture with  $N$  cascaded residual blocks, two different residual block realization schemes, e.g. with  $M_1$  and  $M_2$  convolutional layers, respectively, can be trained to provide the similar reconstruction results, if the residual learning network is deep enough, e.g.  $N \rightarrow +\infty$ .

*Proof:* Please refer to Appendix for the proof. ■

<sup>6</sup>The inputs of the neural network are the amplitude and phase of the estimated channel responses at the pilot locations. Hence, the inputs to the two channels of the network are real values and dealt with in one network.

TABLE 1. Detailed configuration information about the LSRN structure.

Input size = 14 × 14, input channel = 2				
Function	Layer Name	Output Size	Output Channel	Kernel Size, Stride
$\mathcal{F}_{GL}(\cdot; \theta_{GL})$	Convolution	14 × 14	64	3 × 3 conv, stride 1
	Convolution	14 × 14	64	3 × 3 conv, stride 1
$\mathcal{F}_{LL}(\cdot; \theta_{LL})$	ResB(1)	14 × 14	64	$\begin{pmatrix} 3 \times 3 \text{ conv} \\ \text{ReLU} \\ 3 \times 3 \text{ conv} \end{pmatrix}$
	⋮	⋮	⋮	
	ResB(n)	14 × 14	64	
	Convolution	14 × 14	64	3 × 3 conv, stride 1
$\mathcal{F}_{RL}(\cdot; \theta_{RL})$	ConvLSTM	14 × 14	64	3 × 3 conv, stride = 1
$\mathcal{F}_{PS}(\cdot; \theta_{PS})$	Convolution	14 × 14	64 × 4 × 4	3 × 3 conv, stride = 1
	PixelShuffle	56 × 56	64	-
$\mathcal{F}_{FM}(\cdot; \theta_{FM})$	Convolution	56 × 56	2	3 × 3 conv, stride = 1

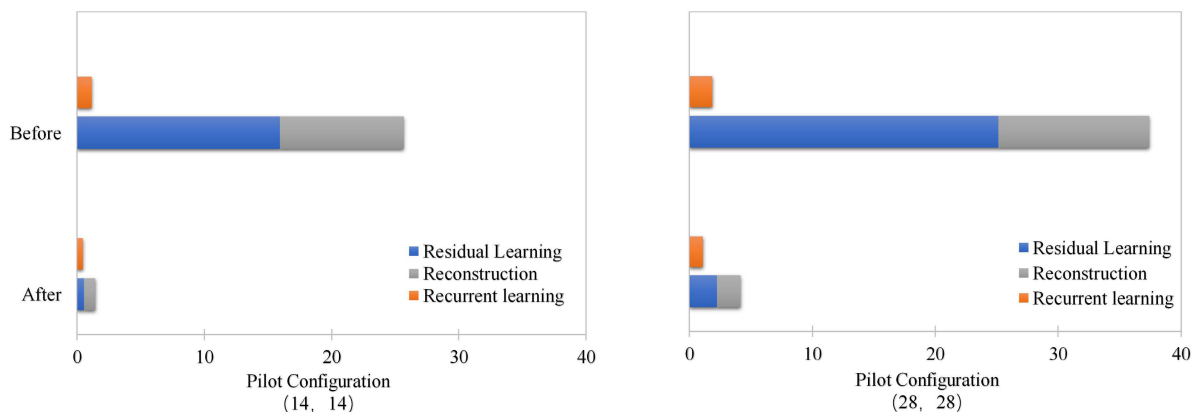


FIGURE 7. Processing delay of LSRN before and after network simplification.

Based on Lemma 1, we can simplify the number of convolutional layers in each ResB when the residual learning network architecture is deep enough. However, in the practical systems, we need to balance the processing delay and the implementation complexity with the achievable performance. In our cases, the number of ResB,  $N$ , is chosen to be 16 and we reduce the number of convolutional layers on top of ResB\_Lim structure. As shown in Figure 8, although the number of ResB,  $N$ , is equal to 16, we can still simplify the number of convolutional layers of each ResB with marginal NMSE losses (e.g. 7.14%) and the overall processing delay is reduced average from 27ms to 20ms, which corresponds to 25.9%.

**B. REDUCED NUMBER OF RESB AND CONVOLUTIONAL KERNELS**

Another scheme to reduce the complexity of the local residual learning path is to reduce the number of ResBs and convolutional kernels. In the conventional EDSR scheme, to reduce the number of ResBs and kernels may greatly affect the image SR performance. However, in the specific task, to reduce the number of ResBs and convolutional kernels, the corresponding performance degradation is controllable due to

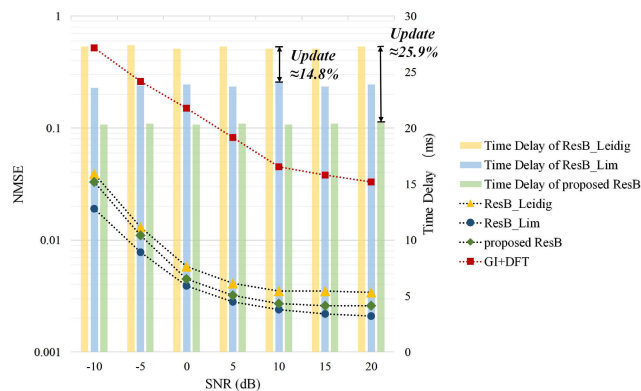
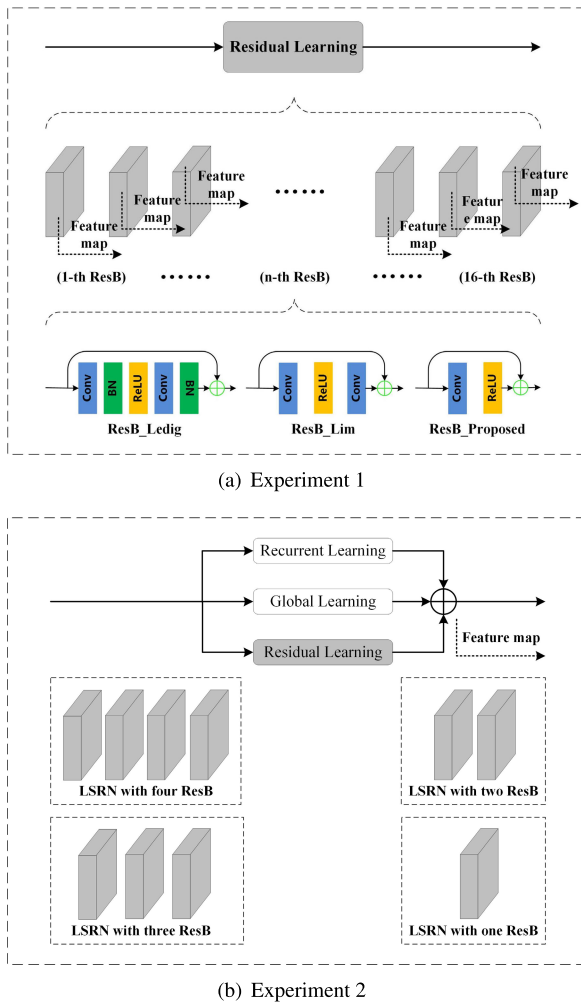


FIGURE 8. NMSE and processing delay comparison of three different ResBs, the blue, green, and yellow histogram and scatter chart represent ResB\_Ledig [42], ResB\_Lim [26], and ours, respectively.

the following reasons. First, since the correlations between different REs is much stronger than the pixels in the convolutional image SR task, the differences of extracted feature maps among different ResBs are much smaller, which makes the possibility to reduce the number of ResBs with marginal performance. Second, with the proposed LSRN architecture, part of the feature losses due to the smaller number of ResBs





**FIGURE 9.** Two experimental architectures that generate feature maps of different ResB layers, subfigure(a) shows the feature map of each layer in the residual learning chain with 16 ResBs, and subfigure(b) shows the feature map of different combined network of residual learning and the recurrent learning.

and convolutional kernels can be compensated by the recurrent learning branch.

To demonstrate the above two effects, we perform the following two experience, where in the first experiment, we extracted the feature map of each layer in the residual learning chain with 16 ResBs as shown in Figure 9(a), and in the second experiment, we combine the recurrent learning and the global learning chains to obtain the overall feature map variations as shown in Figure 9(b).

The extracted feature map of the residual learning chain is depicted in Figure 10(a), where the entire feature values begin to converge after the fourth ResB. Based on this observation, we collect 10000 samples from the COST 2100 channel model and pass them into the residual learning chain.

The average NMSE, define as,  $NMSE = \frac{\sum_{i=1}^{10000} [\hat{h}_{16}^i - \hat{h}_4^i]^2}{\sum_{i=1}^{10000} [\hat{h}_{16}^i]^2}$ , is around 0.0104, which provides another possibility to reduce the processing complexity bu inducing the number of ResBs from 16 to 4. As mentioned before, we combine

the residual learning and the recurrent learning together as the second approach to reduce the processing delay and the corresponding simulation results are shown in Figure 10(b). In this experiment, we directly use 4 ResB rather than the entire 16 ResBs to see the combined effects. By averaging over 10000 examples, we compare the difference in terms of NMSE between the proposed LSRN architecture with four ResBs and with the other three possibilities. Since the NMSE values are quite close for three different cases, we choose the proposed LSRN architecture with one ResB only to reduce the potential complexity.<sup>7</sup> As shown in Figure 11, the overall processing delay for  $14 \times 14$  pilot configuration can be reduced average from 20ms to about 0.9ms, which corresponds to 95.5%.

As discussed in [43], for both classification and generation tasks, even-sized kernels will result in severe performance degradation in general. For odd-sized kernels, increasing the size of kernels cannot provide a monotonic improvement in terms of Top-1 accuracy for ImageNet dataset [44]. Therefore, in the following numerical evaluations, we choose different kernel sizes (with odd numbers) and compare the NMSE result for LSRN-L schemes. As shown in the following Figure 12, we can observe the similar behavior for NMSE performance and the neural networks with kernel size  $5 \times 5$  achieve the best NMSE performance. However, to balance the achieved NMSE and the processing delay, we choose the kernel size to be  $3 \times 3$ , since it achieves the similar NMSE performance in the high SNR regime, and reduces the processing delay by more than 20%, if compared with the kernel size  $5 \times 5$  case.

In short, we propose two different approaches to simplify the network architecture, e.g. to simplify the structure of ResB and to reduced the number of ResBs by measuring the feature variations. Through the priliminary simulation results as shown in Figure 7, the overall processing delay for  $14 \times 14$  pilot configuration can be reduced from 27ms to about 0.9ms, which corresponds to 25.9% and 95.5%, respectively.

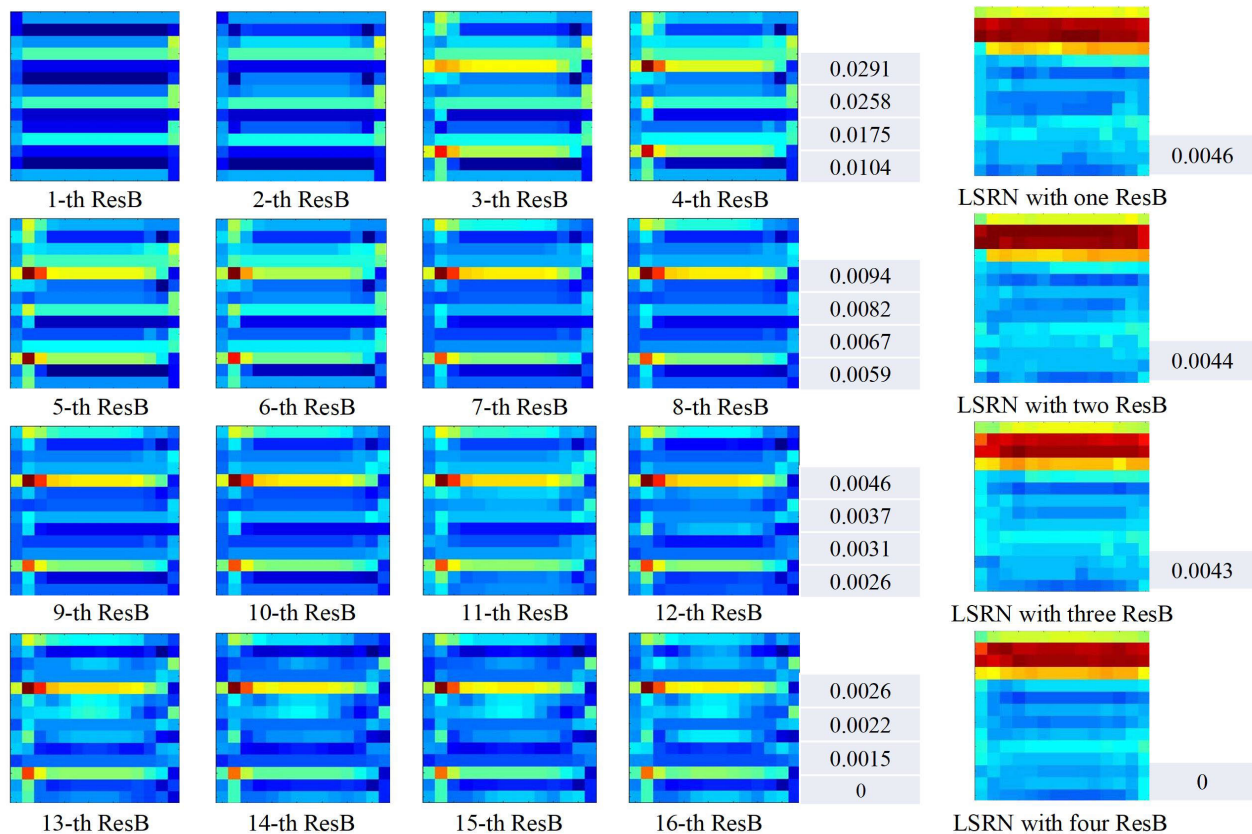
## V. EXPERIMENTS & RESULTS

In this section, we introduce the generation of data sets and simulation environments, for both model generated case and prototype sampled case. In what follows, we present the corresponding evaluation results for the proposed recurrent learning framework.

### A. DATA SETS AND IMPLEMENTATION DETAILS

In the following evaluation, we apply two different approaches to generate the real-time channel state, namely directly “Model Generated” and “Prototype Sampled”. The “Model Generated” data set is directly collected from the famous channel model, COST2100 [31], while the “Prototype Sampled” data set is estimated from the practical

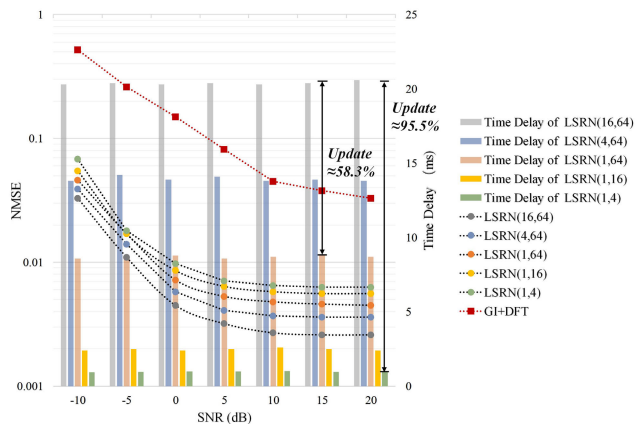
<sup>7</sup>We reduce the number of kernels accordingly, e.g. from 64 to 4, to accommodate the changes of ResBs.



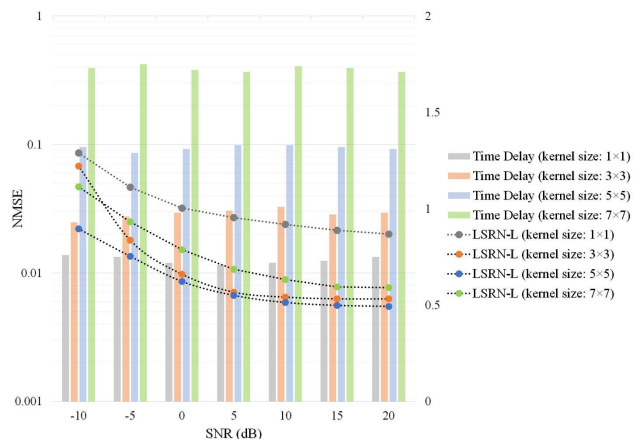
(a) Feature map of each layer in the residual learning chain with 16 ResBs

(b) Overall feature map variations

**FIGURE 10.** Visualization results of of different experimental architectures, subfigure(a) shows the feature map extracted from each layer in the residual learning chain with 16 ResBs and subfigure(b) shows the feature map extracted from the overall LSRN network with different number of ResBs.



**FIGURE 11.** NMSE and processing delay comparison of different network structure with different number of ResBs and convolutional kernels.



**FIGURE 12.** NMSE comparison of different kernel sizes of the proposed scheme in analog channel environment under different pilot configurations with SNR ranges from -10 dB to 20 dB.

WiFi prototype system as shown in Figure 4. The detailed configuration of the prototype system is listed in Table 2.

After we generate the original high resolution CSI  $\mathbf{h}^k(f_{RE}, t_{RE})$ , we can follow the conventional MMSE channel estimation method to generate the low resolution CSI at the RS positions  $\hat{\mathbf{h}}^k_{MMSE}(f_{RS}, t_{RS})$ . Based on different configurations of RSs and the LOS/NLOS, we can construct

different data sets for training and testing as shown in the following Table 3. The sizes of training and testing data sets are chosen to be  $5.6 \times 10^5$  and  $1.12 \times 10^5$ , respectively, and the other related learning parameters are listed in Table 2 as well.

**TABLE 2. Configuration of the WiFi Prototype system and the learning parameters.**

WiFi Prototype		Learning Configuration	
Parameters	Value	Parameters	Value
No. of Tx Antennas	3	Optimizer	Adam
No. of Rx Antennas	1	Learning Rate	0.0001
Carrier Frequency	5.3GHz	Loss Function	l2 Norm
Bandwidth	20MHz		
Protocol	IEEE 802.11n		

**B. COMPARISON WITH STATE-OF-THE-ART**

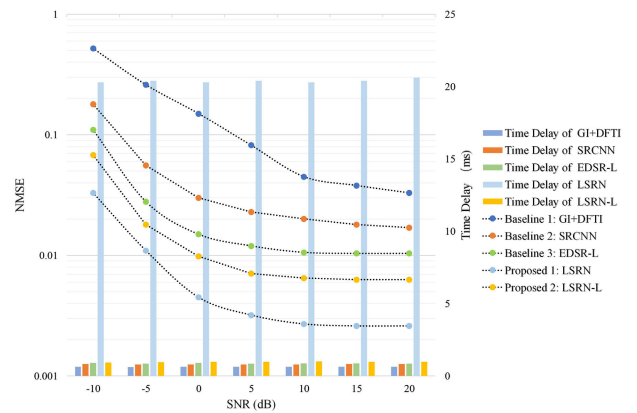
In the following experiments, we compare the proposed LSRN scheme and the corresponding low complexity implementation (denoted as LSRN-L) with the following baseline systems. *Baseline 1* is the conventional GI plus DFTI scheme as elaborated in Section II-A. *Baseline 2* and *Baseline 3* apply the image SR technique for the interpolation process using SRCNN and EDSR networks as explained in Section II-B, respectively.

In the first experiment, we compare the channel estimation performance in terms of NMSE as well as the processing delay versus SNR results under different pilot configuration, where we rely on COST 2100 model to generate the training and testing data sets. All the neural network based solutions are pre-trained using the data sets  $\mathcal{A}$  and tested under the data sets  $\tilde{\mathcal{A}}$ . The corresponding numerical results are depicted in Figure 13. As we can see from this figure, in the condition of  $14 \times 14$  pilot configuration, the proposed LSRN and LSRN-L schemes outperforms the baseline schemes in terms of NMSE among 14 dB to 15 dB and 10 dB to 11 dB, respectively. Meanwhile, in terms of the processing delay, LSRN-L performs much better than LSRN scheme and eventually achieves less than 1 ms delay budget.

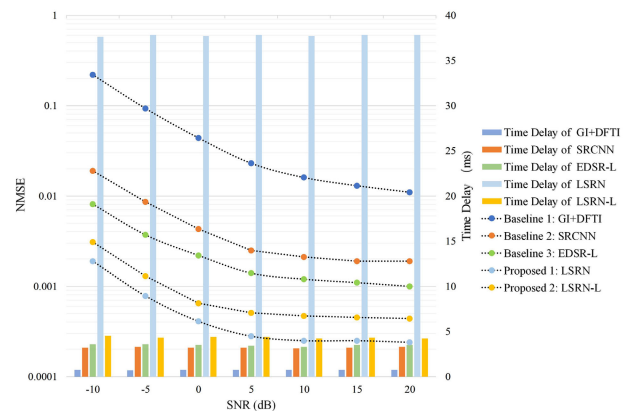
In the second experiment, we compare the channel estimation performance between the tradition GI plus DFTI scheme and our low complexity LSRN-L scheme. As for LSRN network, we respectively use model generated training data set  $\mathcal{A}$ , sampled training data set  $\mathcal{B}$  and  $\mathcal{C}$  for training, and test under the sampled testing data sets  $\tilde{\mathcal{B}}$  and  $\tilde{\mathcal{C}}$ , respectively. As the corresponding numerical results shown in Figure 14,

**TABLE 3. Related training and testing data sets in our experiments.**

Generation Method	Dimension of HR CSI ( $\mathbf{h}^k(f_{RE}, t_{RE})$ )	Dimension of LR CSI ( $\mathbf{h}_{MMSE}^k(f_{RS}, t_{RS})$ )	Training	Testing
Model Generated (COST 2100 LOS)	$56 \times 56$	$14 \times 14$	$\mathcal{A}_{14 \times 14}$	$\tilde{\mathcal{A}}_{14 \times 14}$
		$28 \times 28$	$\mathcal{A}_{28 \times 28}$	$\tilde{\mathcal{A}}_{28 \times 28}$
		$7 \times 28$	$\mathcal{A}_{7 \times 28}$	$\tilde{\mathcal{A}}_{7 \times 28}$
		$28 \times 7$	$\mathcal{A}_{28 \times 7}$	$\tilde{\mathcal{A}}_{28 \times 7}$
Sampled Data Set (WiFi LOS)	$56 \times 56$	$14 \times 14$	$\mathcal{B}_{14 \times 14}$	$\tilde{\mathcal{B}}_{14 \times 14}$
		$28 \times 28$	$\mathcal{B}_{28 \times 28}$	$\tilde{\mathcal{B}}_{28 \times 28}$
Sampled Data Set (WiFi NLOS)	$56 \times 56$	$14 \times 14$	$\mathcal{C}_{14 \times 14}$	$\tilde{\mathcal{C}}_{14 \times 14}$
		$28 \times 28$	$\mathcal{C}_{28 \times 28}$	$\tilde{\mathcal{C}}_{28 \times 28}$



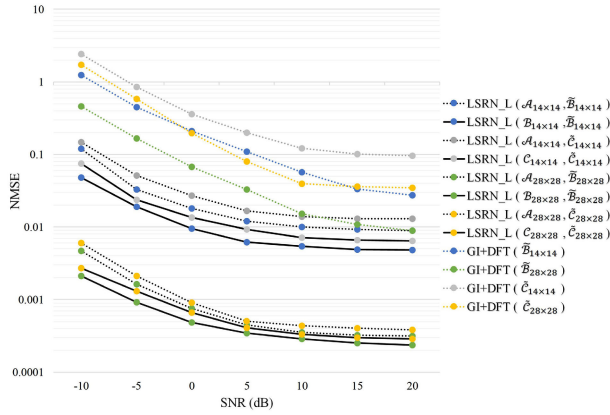
(a) Pilot Configuration:  $14 \times 14$



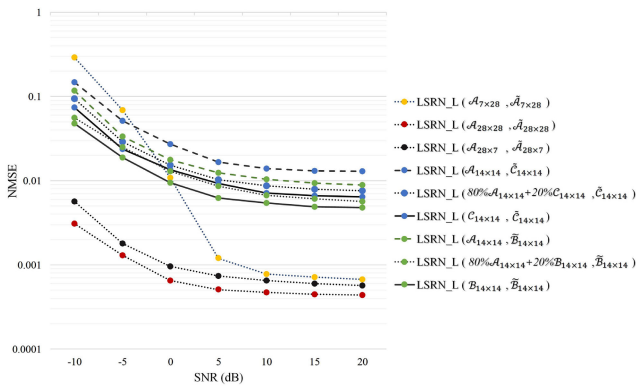
(b) Pilot Configuration:  $28 \times 28$

**FIGURE 13. NMSE comparison between deep-learning based and traditional channel estimation schemes in analog channel environment under different pilot configurations with SNR ranges from  $-10$  dB to  $20$  dB.**

we can see that our LSRN-L scheme outperforms the baseline scheme GI plus DFTI in terms of NMSE in all circumstances. By comparing the test results of different data set training models, we can conclude that the network trained by model generated data can also be applicable to the channel estimation scenarios under the WiFi environment LOS and NLOS, which represents our network is robust to the actual application scenarios. Last but not least, this experiment



**FIGURE 14.** NMSE comparison between traditional and LSRN based channel estimation schemes in LOS and NLOS environments with SNR range from  $-10$  dB to  $20$  dB.



**FIGURE 15.** NMSE comparison between channel estimation schemes under three different sample training situations and different pilots arrangements with SNR range from  $-10$  dB to  $20$  dB.

shows that the complex correlation characteristics between pilots can be learned through network training, so that the network trained by model generate data training can adapt to different channel fading environments in actual communication systems.

### C. EXTENSIVE STUDY

From the results of the previous experiment, we can see that the theoretical model can already deal with the channel estimation problem of the real communication system. In order to observe whether a small amount of real sampled data is helpful to the channel estimation of the real communication system, we improve the training data set by mixing data sets  $\mathcal{B}$  or  $\mathcal{C}$  to data sets  $\mathcal{A}$  in order to simulate the real communication system. According to the experimental results shown in Figure 15, by using the mixed training data sets include 20% sampled data, we can obtain the majority gain of the result, which is around 70%. In the other experiment, with a fixed number of pilots, we discuss the influence of different pilot designs on the experimental results. Two another different pilot design modes are considered, they are training data sets  $\mathcal{A}_{7 \times 28}$ ,  $\mathcal{A}_{28 \times 7}$  and testing data sets  $\tilde{\mathcal{A}}_{7 \times 28}$ ,  $\tilde{\mathcal{A}}_{28 \times 7}$ .

The design idea of pilot arrangements comes from the different correlation of pilots in time and frequency domain. Data set  $\mathcal{A}_{7 \times 28}$  and  $\tilde{\mathcal{A}}_{7 \times 28}$  are generated by concentrates more pilots on different time slots than on different sub-carriers. Data set  $\mathcal{A}_{28 \times 7}$  and  $\tilde{\mathcal{A}}_{28 \times 7}$  are generated by concentrates more pilots on different sub-carriers than on different time slots. The experimental results are shown in figure 15, we can see that the estimation results of the  $7 \times 28$  pilot configuration are quite different under different SNRs, and the overall estimation effect is not as good as that of the  $28 \times 7$  pilot configuration. What's more, the  $28 \times 7$  pilot configuration with fewer pilots obtains the close estimation effect with the  $28 \times 28$  pilot configuration, in term of NMSE is among 1 dB to 2 dB. The above experimental results provide us a possibility to explore more pilot design methods to improve our network in the future.

## VI. CONCLUSION

In this paper, we propose a LSRN architecture to jointly utilize the recurrent and residual learning capabilities for the channel estimation tasks in the wireless transmission. By exploiting the slow-varying time domain correlation and the non-linear interpolation relations among different RSs and REs, the proposed low complexity LSRN can provide among 10db to 11dB NMSE improvement if compared with the conventional GI plus DFTI based scheme, and eventually consumes less than one millisecond processing delay.

## APPENDIX

### PROOF OF LEMMA 1

Given function  $\mathcal{F}_{LL}^{n,M}(\hat{\mathbf{h}}^{k,n}, \Theta_{LL}^M)$  is the  $n$ -th local residual learning function of the network with  $N$  ResBs, each of which includes  $M$  convolutional layers. And the process of  $l$ -th ResB can be expressed as  $\hat{\mathbf{h}}^{k,l+1} = \hat{\mathbf{h}}^{k,l} + \mathcal{F}_{LL}^{n,l,M}(\hat{\mathbf{h}}^{k,l}, \Theta_{LL}^M)$ , where  $\hat{\mathbf{h}}^{k,l}$  and  $\hat{\mathbf{h}}^{k,l+1}$  are the input and output vectors of the  $l$ -th ResB layers considered.  $\Theta_{LL}^M = \{\theta_{LL,M}^1, \theta_{LL,M}^2, \dots, \theta_{LL,M}^M\}$  represents the weight sets of the  $l$ -th ResB. Specifically,  $\mathcal{F}_{LL}^{n,M}(\hat{\mathbf{h}}^{k,n}, \Theta_{LL}^M)$  can be expressed in detail by the following formula,

$$\mathcal{F}_{LL}^{n,M}(\hat{\mathbf{h}}^{k,n}, \Theta_{LL}^M) = \begin{cases} \sigma \theta_{LL,M}^{1,n} \hat{\mathbf{h}}^{k,n}, & M = 1 \\ \theta_{LL,M}^{M,n} \sigma \theta_{LL,M=2}^{1,n} \hat{\mathbf{h}}^{k,n}, & M = 2 \\ \theta_{LL,M}^{M,n} \sigma \mathcal{F}_{LL}^{n,M-1}(\hat{\mathbf{h}}^{k,n}, \Theta_{LL}^M), & M > 2. \end{cases}$$

where  $\sigma$  denotes ReLU [45] and the biases are omitted for simplifying notations. And for any  $N$ -layer residual network, the output of the residual network  $\hat{\mathbf{h}}$  can be represented as follow,

$$\begin{aligned} \hat{\mathbf{h}} &= \hat{\mathbf{h}}^{k,1} + \sum_{n=1}^N \mathcal{F}_{LL}^{n,M}(\hat{\mathbf{h}}^{k,n}, \Theta_{LL}^M) \\ &= \hat{\mathbf{h}}^{k,1} + \sum_{n=1}^N \mathcal{F}_{LL}^{n,M}(\hat{\mathbf{h}}^{k,n}, \Theta_{LL}^M) \\ &= \prod_{n=1}^N (1 + \theta_{LL,M}^{M,n} \sigma \dots \theta_{LL,M}^{2,n} \sigma \theta_{LL,M}^{1,n}) \hat{\mathbf{h}}^{k,1} \end{aligned} \quad (9)$$

Considering two architectures composed of two different ResBs, one of which includes  $M_2$  convolution layers and the other of which includes  $M_1$  convolution layers. The output of them is  $\hat{\mathbf{h}}_{M_2} = \prod_{n=1}^N (1 + \theta_{LL,M_2}^{M_2,n} \dots \sigma \theta_{LL,M_2}^{1,n}) \hat{\mathbf{h}}^{k,1}$  and  $\hat{\mathbf{h}}_{M_1} = \prod_{n=1}^N (1 + \theta_{LL,M_1}^{M_1,n} \dots \sigma \theta_{LL,M_1}^{1,n}) \hat{\mathbf{h}}^{k,1}$ , respectively. When both networks are trained to convergence, the training error of the two networks will decrease gradually as  $N$  increases gradually. When  $N \rightarrow +\infty$ , they can fit into the same function, parameters of the two networks meet the following equation,

$$\prod_{n=1}^{N \rightarrow +\infty} \left( (1 + \theta_{LL,M_2}^{M_2,n} \dots \sigma \theta_{LL,M_2}^{1,n}) - (1 + \theta_{LL,M_1}^{M_1,n} \dots \sigma \theta_{LL,M_1}^{1,n}) \right) \rightarrow 0. \quad (10)$$

## REFERENCES

- [1] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5G ultra-reliable and low-latency systems design," in *Proc. EuCNC*, Jun. 2017, pp. 1–5.
- [2] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications: Theory and Applications of OFDM*. Norwell, MA, USA: Kluwer, 1999.
- [3] J. Beek, O. Edfors, S. Magnus, S. Wilson, P. Brjesson, S. Kate, W. Per, and O. Brjesson, "On channel estimation in ofdm systems," in *Proc. Veh. Technol. Conf.*, 1995, pp. 815–819.
- [4] X. Liu, Y. Shi, J. Zhang, and K. B. Letaief, "Massive CSI acquisition in dense cloud-RAN with spatial and temporal prior information," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [5] A. H. Khanshan and H. Amindavar, "Performance evaluation of two-dimensional interpolations on OFDM channel estimation," in *Proc. Australas. Telecommun. Netw. Appl. Conf.*, Dec. 2007, pp. 1–5.
- [6] M. B. Sutar and V. S. Patil, "LS and MMSE estimation with different fading channels for OFDM system," in *Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, vol. 1, Apr. 2017, pp. 740–745.
- [7] W. Chen and Q. Zhu, "Research on interpolation methods for channel estimation in the MIMO-OFDM systems," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2007, pp. 306–309.
- [8] F. Weng, C. Yin, and T. Luo, "Channel estimation for the downlink of 3GPP-LTE systems," in *Proc. 2nd IEEE Int. Conf. Netw. Infrastruct. Digit. Content*, Sep. 2010, pp. 1042–1046.
- [9] Q. Shi, Y. Liu, S. Zhang, S. Xu, S. Cao, and V. Lau, "Channel estimation for WiFi prototype systems with super-resolution image recovery," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
- [10] Y. Ju Lee and J. Yoon, "Nonlinear image upsampling method based on radial basis function interpolation," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2682–2692, Oct. 2010.
- [11] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 479–490, Jan. 2007.
- [12] X. Liang and Z. Gan, "Improved non-local iterative back-projection method for image super-resolution," in *Proc. 6th Int. Conf. Image Graph.*, Aug. 2011, pp. 176–181.
- [13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [14] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. ACCV*, 2014, pp. 111–126.
- [15] J. Salvador and E. Perez-Pellitero, "Naive Bayes super-resolution forest," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 325–333.
- [16] Y. Yang, G. Zhu, Y. Chen, and R. Tassing, "Local linear regression based channel estimation for OFDM systems," in *Proc. 4th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Oct. 2008, pp. 1–5.
- [17] Y. Ding and B. D. Rao, "Channel estimation using joint dictionary learning in FDD massive MIMO systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 185–189.
- [18] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.
- [19] N. Ketkar, "Convolutional neural networks," in *Deep Learning With Python*. New York, NY, USA: Apress, 2017.
- [20] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," *CoRR*, vol. abs/1904.07523, Apr. 2019.
- [21] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4549–4557.
- [22] 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)*, Jul. 2017, pp. 5835–5843.
- [23] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. ECCV*, vol. 8692, Sep. 2014, pp. 184–199.
- [24] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [25] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, Switzerland: Springer, 2016, pp. 391–407.
- [26] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
- [27] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia, "Detail-revealing deep video super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4482–4490.
- [28] M. Chu, Y. Xie, N. Thuerey, and L. Leal-Taixé, "Learning temporal coherence via self-supervision for GAN-based video generation," *CoRR*, vol. abs/1811.09393, 2018.
- [29] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1228–1231, Aug. 2019.
- [30] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [31] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quirin, F. Tufvesson, and P. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, Dec. 2012.
- [32] J. Kyoung Moon and S. In Choi, "Performance of channel estimation methods of OFDM systems in a multipath fading channels," *IEEE Trans. Consum. Electron.*, vol. 46, no. 1, pp. 161–170, Feb. 2000.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [34] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *CoRR*, vol. abs/1609.05158, Sep. 2016.
- [35] K. Purohit and A. N. Rajagopalan, "Region-adaptive dense network for efficient motion deblurring," *CoRR*, vol. abs/1903.11394, Sep. 2019.
- [36] X. Hu, K. Yang, L. Fei, and K. Wang, "ACNet: Attention based network to exploit complementary features for RGBD semantic segmentation," *CoRR*, vol. abs/1905.10089, May 2019.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, Dec. 2015.
- [38] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *CoRR*, vol. abs/1409.3215, Dec. 2014.
- [39] K. Cho, B. van Merriënboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *CoRR*, vol. abs/1406.1078, Sep. 2014.
- [40] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *CoRR*, vol. abs/1506.04214, Sep. 2015.
- [41] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
- [42] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. S. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, Sep. 2016.
- [43] S. Wu, G. Wang, P. Tang, F. Chen, and L. Shi, "Convolution with even-sized kernels and symmetric padding," *CoRR*, vol. abs/1903.08385, May 2019.

- [44] M. Tan and Q. V. Le, "MixConv: Mixed depthwise convolutional kernels," *CoRR*, vol. abs/1907.09595, Dec. 2019.
- [45] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.



**SHUNQING ZHANG** (Senior Member, IEEE) received the B.S. degree from the Department of Microelectronics, Fudan University, Shanghai, China, in 2005, and the Ph.D. degree from the Department of Electrical and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, in 2009.

He was with the Communication Technologies Laboratory, Huawei Technologies, as a Research Engineer and then a Senior Research Engineer, from 2009 to 2014, and a Senior Research Scientist of the Intel Collaborative Research Institute on Mobile Networking and Computing, Intel Labs, from 2015 to 2017. Since 2017, he has been a Full Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai. He has published over 60 peer-reviewed journal articles and conference papers and over 50 granted patents. His current research interests include energy efficient 5G/5G+ communication networks, hybrid computing platform, and joint radio frequency and baseband design. He has received the National Young 1000-Talents Program and won the Paper Award for Advances in Communications from the IEEE Communications Society, in 2017.



**YANGYU LIU** received the B.E. degree in communication engineering from Shanghai University, China, in 2017, where she is currently pursuing the master's degree in information and communication engineering. Her research interests include image reconstruction, image recognition, and deep learning.



**QI SHI** received the B.E. degree in communication engineering from Shanghai University, China, in 2019, where she is currently pursuing the M.E. degree in communication and information system. Her research interests include channel estimation and deep learning.



**SHUGONG XU** (Fellow, IEEE) received the degree from Wuhan University, China, in 1990, and the master's degree in pattern recognition and intelligent control and the Ph.D. degree in EE from the Huazhong University of Science and Technology (HUST), China, in 1993 and 1996, respectively. He is currently a Professor with Shanghai University and the Head of the Shanghai Institute for Advanced Communication and Data Science (SICS). He was the Center Director and the Intel

Principal Investigator of the Intel Collaborative Research Institute for Mobile Networking and Computing (ICRI-MNC), prior to December 2016 when he joined Shanghai University. Before joining Intel in September 2013, he was a Research Director and a Principal Scientist with the Communication Technologies Laboratory, Huawei Technologies. Among his responsibilities at Huawei, he founded and directed Huawei's Green Radio Research Program, Green Radio Excellence in Architecture and Technologies (GREAT). He was also the Chief Scientist and a PI for the China National 863 Project on End-to-End Energy Efficient Networks. He was one of the co-founders of the Green Touch Consortium together with Bell Labs and a Co-Chair of the Technical Committee for three terms in this international consortium. Prior to joining Huawei in 2008, he was with Sharp Laboratories of America, as a Senior Research Scientist. Before that, he conducted research as a Research Fellow with The City College of New York, Michigan State University, and Tsinghua University. He has published over 100 peer-reviewed research articles in top international conferences and journals. One of his most referenced articles has over 1400 Google Scholar citations, in which the findings were among the major triggers for the research and standardization of the IEEE 802.11S. He has over 20 U.S. patents granted. Some of these technologies have been adopted in international standards, including the IEEE 802.11, 3GPP LTE, and DLNA. His current research interests include wireless communication systems and machine learning. He was elevated to IEEE Fellow, in 2015, for contributions to the improvement of wireless networks efficiency. He was awarded the National Innovation Leadership Talent by the China Government, in 2013. He is also the Winner of the 2017 Award for Advances in Communication from the IEEE Communications Society.



**SHAN CAO** received the B.S. and Ph.D. degrees in microelectronics from Tsinghua University, China, in 2009 and 2015, respectively. She has held a postdoctoral position at the School of Information and Electronics, Beijing Institute of Technology, from 2015 to 2017. She is currently an Assistant Professor with Shanghai University. Her current research interests include wireless communication systems, channel encoding and decoding, machine learning acceleration, and its ASIC design.

...