IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 30, 2019, accepted July 12, 2019, date of publication July 16, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929091

UL-CSI Data Driven Deep Learning for Predicting DL-CSI in Cellular FDD Systems

JIE WANG[1], YING DING[2], SHUJIE BIAN[1], YANG PENG[1], MIAO LIU[1], (Member, IEEE), AND GUAN GUI[1], (Senior Member, IEEE)

[1]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2]College of Overseas Education, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Corresponding authors: Miao Liu (liumiao@njupt.edu.cn) and Guan Gui (guiguan@njupt.edu.cn)

This work was supported in part by the Jiangsu Specially Appointed Professor Program under Grant RK002STP16001, in part by the Summit of the Six Top Talents Program of Jiangsu under Grant XYDXX-010, in part by the Program for High-Level Entrepreneurial and Innovative Talents Introduction under Grant CZ0010617002, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX19-0903, and in part by the 1311 Talent Plan of Nanjing University of Posts and Telecommunications.

ABSTRACT Frequency division duplex (FDD) systems dominate current cellular networks due to its advantages of low latency and strong anti-interference ability. However, the computation and the feedback overheads for predicting the downlink channel state information (DL-CSI) are the major bottlenecks to further improve the cellular FDD systems performance. To deal with these problems, in this paper, a convolutional long short-term memory network (ConvLSTM-net)-based deep learning method is proposed for predicting the DL-CSI from the uplink channel state information (UL-CSI) directly. In detail, our proposed ConvLSTM-net consists of two modules: one is the feature extraction module that learns spatial and temporal correlations between the DL-CSI and the UL-CSI, and the other one is the prediction module that maps the extracted features to the reconstructions of the DL-CSI. To evaluate the outperformance of the ConvLSTM-net, a long short-term memory network (LSTM-net) and a convolutional neural networks (CNN)-based schemes are simulated for comparisons. The simulation experiments consist of two parts. One part is that the hyper parameters of the proposed ConvLSTM-net are analyzed to explore their effects on the prediction performance. Another part is that experiments are conducted in the time domain and frequency domain, respectively, for selecting a more proper domain to predict the DL-CSI accurately. From the experiment results above, it can be verified that the proposed ConvLSTM-net with proper hyper parameters outperforms the compared schemes at predicting DL-CSI according to UL-CSI in the cellular FDD systems, especially in the time domain.

INDEX TERMS Channel prediction, channel state information, downlink, deep learning, FDD, uplink.

I. INTRODUCTION

In wireless communication networks, accurate channel state information at transmitter sides is crucial for various technologies, such as precoding, beamforming and power allocation, to guarantee the communication quality and throughputs [1]. To achieve this goal, in time division duplex (TDD) based cellular networks, the downlink channel state information (DL-CSI) at transmitters is inferred from the uplink channel state information (UL-CSI) without any additional channel estimation or feedback at the receivers by exploiting the channel reciprocity [2]. However, the channel reciprocity property does not strictly hold in cellular frequency division duplex (FDD) systems since the uplink channel and the downlink channel are located in the different frequency bands [3]. Previously, to obtain the DL-CSI in cellular FDD systems, the receivers should firstly predict the DL-CSI and then sent it back to the transmitters [4]. Obviously, the DL-CSI prediction and feedback may cause high computation and feedback overheads at the receivers [5]. As a result, they are the major bottlenecks for the cellular FDD systems to improve the performance further. Particularly, the FDD systems dominate the current cellular networks due to the advantages of low latency and strong anti-interference ability [4], [6], [7]. Therefore, it is urgent to solve the problems of computation and feedback overheads in the FDD based cellular networks.

The associate editor coordinating the review of this manuscript and approving it for publication was Tomohiko Taniguchi.

To overcome the bottlenecks of the cellular FDD systems mentioned above, the existing schemes can be classified into two categories as below.

The first category is DL-CSI feedback methods [7]–[11]. These methods exploit the sparsity structure of the CSI in specific domains to obtain the moderately accurate measurements of the full DL-CSI from the limited feedbacks. Among existed DL-CSI feedback methods, [7]–[10] are based on compression sensing (CS) technologies. However, two essential problems are inherent in the CS-based methods. Firstly, the CSI matrix is not exactly sparse in any specific domain. Secondly, there still exist computational overhead of estimating DL-CSI and certain feedback overhead. Fortunately, with the development of the deep neural networks (DNN), machine learning, especially deep learning, has been applied in wireless communications [12], such as channel estimation [13], [14], channel feedback [11], signal detection [15], [16], beamforming [17], resource allocation [18], and modulated signal identification [19], [20], which brings hope to solve the problems of the CS-based methods above. To eliminate the limitations in the traditional CS-based DL-CSI feedback methods, [11] proposed a deep learning based CsiNet for DL-CSI feedback. The scheme learns the inherent characteristics of the CSI for compressing and recovering. Though the CsiNet achieves significant improvements on reducing the DL-CSI feedback overhead and enhancing DL-CSI recovery accuracy, it does not consider the computation overhead of the channel prediction at the receivers. Therefore, this method still cannot satisfy the scenarios with ultra-low latency requirements.

The second category is channel correlations based methods. In detail, the transmitters acquire the DL-CSI by utilizing the UL-CSI. The relevant work is summarized as follows. The DL-CSI covariance matrix is estimated from the uplink (UL) pilots sent by the receivers to the transmitters in [3]. Based on the empirical results that there exists some strong correlations between the UL arrival and the downlink (DL) departure angles in the cellular FDD systems, a method called directional training was proposed in [21] to obtain the DL-CSI by sending a small number of training symbols. A method that derives the DL-CSI covariance matrix from the UL-CSI based on channel reciprocity in angle domain of the cellular FDD systems is proposed in [22]. The above methods are based on the slow variations of long-term statistical channel characteristics in different frequency bands, which mainly include the channel covariance matrix and the direction of arrival. Different from the above methods, [23] proposes a new perspective that physical paths traversed by signals in different frequency bands are same. By utilizing the observation, it extracts the frequency-independent physical paths of signals from the UL-CSI, and then learns parameter values of the physical paths. Finally, it transforms these information to DL-CSI at any other frequency bands.

In summary, all these methods mentioned above reveal that there exist certain correlations between the DL-CSI and the UL-CSI in the cellular FDD systems. However, these methods are either computationally exhaustive or based on the mathematical models, which may mismatch the actual channel characteristics due to the complex signal propagation environments.

Inspired by the advantages of deep learning technologies [24] and the channel correlations in the cellular FDD systems, the goal of the present study is to enable the cellular FDD systems to predict the DL-CSI from the UL-CSI directly. From the deep learning perspective, we treat deriving the DL-CSI from the UL-CSI as a spatiotemporal image forecasting problem and propose a convolutional long short-term memory network (ConvLSTM-net) to resolve this problem. The ConvLSTM-net learns frequency-independent information from the UL-CSI and then transfers it to the DL-CSI. Considering time-varying properties of the channel, we also treat predicting the DL-CSI from the UL-CSI as a time sequence forecasting problem and design a long short-term memory network (LSTM-net) to handle this problem. To evaluate the performance of the proposed networks, a convolutional neural network (CNN) [25] is simulated for comparisons. Firstly, since convolutional kernel size is a very important hyper parameter for convolutional operations, this paper explores the effects of different kernel sizes of the CNN and the CovnLSTM-net schemes. Secondly, to select a more effective domain for exploring channel reciprocity of the cellular FDD systems, this paper conducts experiments in frequency domain and time domain respectively. Simulation results demonstrate that the proposed ConvLSTM-net with proper hyper parameters outperforms the LSTM-net and CNN at predicting DL-CSI according to UL-CSI in the cellular FDD systems, especially in time domain.

Contributions of this paper can be concluded as follows: (1) Two new neural networks are proposed to derive the DL-CSI from the UL-CSI. (2) The effects of different kernel sizes of CNN and CovnLSTM-net are analyzed. (3) The proposed schemes are implemented in time domain and frequency domain respectively to select a more effective domain to predict the DL-CSI from the UL-CSI.

The rest of the paper is structured as follows. System model is introduced in Section II. Section III details the ConvLSTM-net and compared schemes. In section IV, experiments are simulated to evaluate the outperformance of the proposed ConvLSTM-net. Finally, Section V concludes the paper.

II. SYSTEM MODEL

We consider the base station (BS) and the user equipment (UE) with single antenna in a cellular FDD system. In current cellular FDD systems, DL-CSI is predicted at UE side and then is sent to BS for serving the information transmission, e.g. precoding, power allocation. This communication process incurs high communication and feedback overheads, as shown in Fig. 1 (a).

In present paper, we want to remove away the computation and feedback overheads by deriving the DL-CSI from the UL-CSI directly without any additional DL-CSI prediction

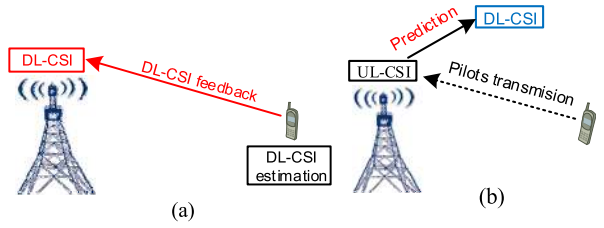


FIGURE 1. (a) The traditional communication process in the cellular FDD systems. (b) The proposed communication process in the cellular FDD systems.

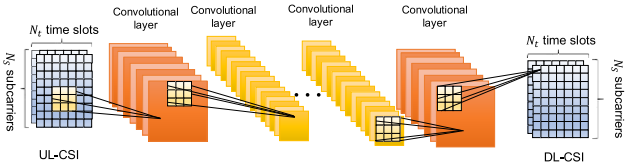


FIGURE 2. Framework of the CNN.

or feedback at UE sides. Under the proposed scheme above, the communication process is described as Fig.1 (b). Consider a scenario that the cellular FDD system is operated in orthogonal frequency division multiplexing (OFDM) model with N_s subcarriers and the channel is time-varying. Then a sample of N_t time slots CSI is a complex matrix with size of $N_s \times N_t$.

III. DETAILED INTRODUCTION OF THE PROPOSED CONVLSTM-NET AND COMPARED SCHEMES

In this section, the principle and structure of the proposed ConvLSTM-net are introduced in detail. For the sake of analysis, a related work CNN [25] and the proposed LSTM-net are also described in detail as benchmarks.

A. THE CNN

This part reviews the CNN proposed by [25]. In this scheme, the UL-CSI matrix is fed into a CNN as a two-dimensional image with two channels, see Fig. 2. As shown in Fig. 2, convolution kernel can extract features of the UL-CSI hidden in the adjacent subcarriers and the adjacent time slots. The convolutional operation can be described as Eq. (1), where $h^{\ell+1}_{i_0,j_0,d}$ denotes the convolutional result of d *th* convolutional kernel in ℓ *th* layer at location (i_0, j_0) and $f_{i,j,d^\ell,d}$ denotes the weight of the convolutional kernel. Besides, D^ℓ is the number of channels of convolutional kernel, which is the same as the number of channels of the input image. When set the hyper parameter padding as "same", H equals the rows of the input image and W equals the columns of the input image.

$$h^{\ell+1}_{i_0,j_0,d} = \sum_{i=0}^{H} \sum_{j=0}^{W} \sum_{d^\ell}^{D^\ell} f_{i,j,d^\ell,d} \times h^{\ell}_{i_0+i,j_0+j,d^\ell} \quad (1)$$

With the stacking of several convolution layers, the depth features obtained by each layer gradually transforms from generalized features to high-level semantic representations. The CNN includes five hidden convolutional layers and uses *tanh* as activation function in all hidden layers. In order to fully utilize the edge information of the input data, symmetric padding is used in first two layers and zero padding is used in other layers [25].

B. THE PROPOSED LSTM-NET

The CNN only extract local temporal correlations between time slots (W time slots, see Fig. 2) of the UL-CSI. To exploit more information hidden in the whole time slots of the UL-CSI, we design a LSTM-net to resolve this problem. LSTM-net has been amazingly successful at capturing long-term patterns in time series data. In the proposed LSTM-net scheme, the UL-CSI matrix is split according to time slots and each time slot is sent to the LSTM-net as a time step. The proposed LSTM-net consists of four LSTM layers. An unrolled through time structure of a LSTM layer is shown in Fig. 3. In Fig. 3, from $h_{ul}(1)$ to $h_{ul}(t)$ are t time slots of the UL-CSI and from $h_{dl}(t+1)$ to $h_{dl}(2t)$ denote next t time slots of the DL-CSI. To describe the operating principle of LSTM-net, the architecture of a basic LSTM cell is shown in Fig. 4 [26].

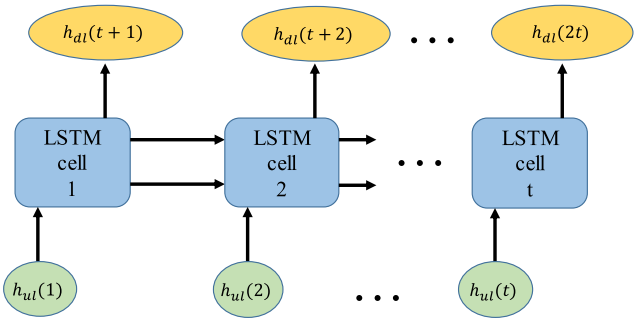


FIGURE 3. Unrolled through time structure of a LSTM layer.

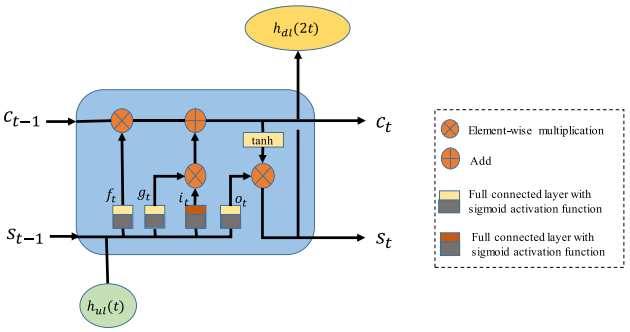


FIGURE 4. The inner structure of a basic LSTM cell.

In Fig. 4, c_{t-1} and s_{t-1} are the long-term state and short-term state of $t-1$ time step respectively. By training the LSTM cell, it can learn what to throw away through forgetting gate f_t, what to store in the long-term state through adding operation and what to read from it through output gate o_t. Equations (2) to (7) summarize the forward propagation process during the training process.

$$i_t = \sigma\left(W_{hi}^T h_{ul}(t) + W_{si}^T s_{t-1} + b_i\right) \quad (2)$$

$$f_t = \sigma\left(W_{hf}^T h_{ul}(t) + W_{sf}^T s_{t-1} + b_f\right) \quad (3)$$

$$o_t = \sigma\left(W_{ho}^T h_{ul}(t) + W_{so}^T s_{t-1} + b_o\right) \quad (4)$$

$$g_t = \tanh\left(W_{hg}^T h_{ul}(t) + W_{sg}^T s_{t-1} + b_g\right) \quad (5)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (6)$$

$$h_{dl}(t+7) = s_t = o_t \otimes \tanh(c_t) \quad (7)$$

where W_{hi}, W_{hf}, W_{ho}, W_{hg} are weight matrices between the input $h_{ul}(t)$ and four full connected layers of the LSTM cell, W_{si}, W_{sf}, W_{so}, W_{sg} are weight matrices between the short-term state s_{t-1} and four full connected layers of the LSTM cell, b_i, b_f, b_o, b_g are the bias terms for each of the four layers respectively.

C. THE PROPOSED CONVLSTM-NET SCHEME

Although LSTM-net is powerful for learning temporal correlations, it has too much redundancy for spatial features. To resolve this problem, this paper formulates predicting the DL-CSI from the UL-CSI as a spatiotemporal image forecasting problem and designs a ConvLSTM-net. The ConvLSTM-net consists of several ConvLSTM layers that have been amazingly successful at capturing spatiotemporal correlations. Compared with the LSTM layer, the ConvLSTM layer replace full connection operations (see Fig. 4) with convolutional operations in both input-to-state and state-to-state transitions [27]. Besides, the inputs, long-term states, short-term states, forget gates, input gates and output gates of ConvLSTM layer are 3D tensors whose last two dimensions (rows and columns) are spatial dimensions. By stacking multiple ConvLSTM layers, the ConvLSTM-net has powerful representation ability and is suitable for giving predictions in complex dynamic systems like the DL-CSI forecasting problem in this paper.

Our designed ConvLSTM-net consists of two modules: one is feature extraction module and the other one is prediction module. In detail, the feature extraction module consists of five ConvLSTM layers. Following feature extraction module, a prediction module including one 3D convolutional layer (3D-CL) is concatenated to recover the DL-CSI. The architecture of the proposed ConvLSTM-net is shown in Fig. 5.

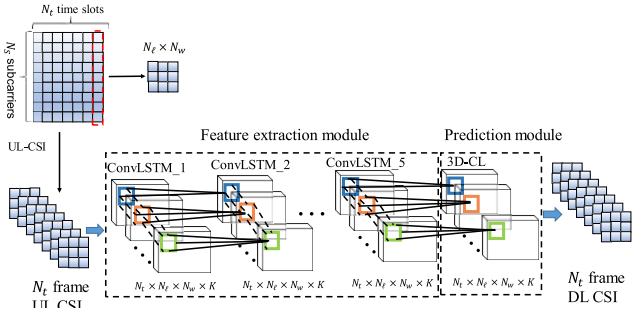


FIGURE 5. Framework of our proposed ConvLSTM-net.

Different from CNN, the UL-CSI data should be preprocessed before enter the ConvLSTM-net. Firstly, each time slot of UL-CSI is reshaped to a two dimension image with size of $N_\ell \times N_w$, where $N_s = N_\ell N_w$. In this paper, $N_\ell = N_w = 6$. Secondly, real values of the image are extracted as the first channel and imaginary values of the image are extracted as the second channel.

One sample of the input of the ConvLSTM-net is a 4D tensor, i.e., (frames, rows, cols, channels). In the 4D tensor, "frames" denotes the number of time steps of the input (here is the number of time slots of the UL-CSI), rows × cols × channels is the size of one time step (here is a 3D image with size of $N_\ell \times N_w \times 2$). As shown in Fig. 5, one sample in present problem are N_t frames images. Each image is presentation of one time slot of the UL-CSI matrix as stated above. Likewise, one sample of the output is also N_t frames images which are N_t time slots of the predicted DL-CSI.

In Fig.5, the values of $N_t \times N_\ell \times N_w \times K$ denote the frames, length, width, and channels of the feature maps, respectively. For all ConvLSTM layers, $K = 32$. For the final 3D-CL, $K = 2$. In the ConvLSTM-net, all layers use zero padding to make feature maps have the same size of the input. In feature extraction module, each ConvLSTM layer use tanh as activation function and followed by a batch normalization layer. The final convolutional layer uses linear activation function to recover the DL-CSI.

End-to-end training method is used to train all parameters of the ConvLSTM-net. We use $\Phi = \{\Phi_{ex}, \Phi_{pre}\}$ to denote parameters in feature extraction module and prediction module. And then the relationship of the UL-CSI and the reconstruction of the DL-CSI is defined as

$$\hat{H}_{DL} = f_{pre}\left(f_{ex}\left(H_{UL}; \Phi_{ex}\right); \Phi_{pre}\right) \quad (8)$$

IV. EXPERIMENTS AND DISCUSSIONS

In this section, we conduct experiments in time domain and frequency domain respectively to select a more effective domain for this problem.

To evaluate the proposed methods, extended Vehicular A (EVA) dataset including 40,000 independent samples (35,000 samples for train and 5,000 samples for test) is adopted [25]. To take into account Doppler frequency shift, EVA dataset is generated under mobile speed of 50km/h. One channel matrix sample has size of 72 × 14 (72 subcarriers in 14 time slots) and the first 36 subcarriers over the first 7 time slots is assigned as the UL-CSI, and the second 36 subcarriers over the second 7 time slots is allocated as the DL-CSI.

For CNN, LSTM-net and ConvLSTM-net, the batch size and epochs are all set as 35 and 300 respectively. We use dynamic learning rate by monitoring validation loss. The learning rate is initialized with 0.01 and will be reduced to one tenth of the original value if the validation loss does not reduce after 40 epochs. We use Adam optimizer in the three networks. They are all trained offline. After training, these networks can be deployed online to input the UL-CSI and then get the responding DL-CSI directly. The implementations of these three neural networks are in Python and all experiments are run on NVIDIA GTX 1080 Ti.

We compare the proposed two neural networks (LSTM-net and ConvLSTM-net) with CNN [25] under normalized mean squared error (NMSE) metric, which is defined as

$$\mathrm{NMSE} = \mathrm{E}\left[\frac{\left\|h_{DL} - \hat{h}_{DL}\right\|_2^2}{\|h_{DL}\|_2^2}\right] \tag{9}$$

where h_{DL} is the actual DL-CSI and $\hat{h}_{DL}$ is the prediction of the DL-CSI. As a standard metric, NMSE can fairly evaluate the performance of a scheme since it is independent of the elements value of the channel matrix.

For CNN scheme, [25] uses Xavier method to initialize network parameters. For ConvLSTM-net, we use Glorot uniform method and orthogonal method to initialize convolutional kernel parameters and recurrent kernel parameters respectively. For all networks, we use mean square error (MSE) as the loss function which is defined as

$$L(\Phi) = \frac{1}{N}\sum\nolimits_N \left\|\hat{h}_{DL} - h_{DL}\right\|_2^2 \tag{10}$$

where operation $\|\cdot\|_2$ denotes Euclidean norm, and N is the number of training samples in one batch.

TABLE 1. Experiment results of the CNN and the ConvLSTM-net with different kernel sizes.

(a) NMSE (dB) in frequency domain.

Kernel size	CNN	ConvLSTM-net
1×1	-0.0222	-0.1453
3×3	-16.6154	-30.7425
5×5	-11.8977	-36.2217

(b) NMSE (dB) in time domain.

Kernel size	CNN	ConvLSTM-net
1×1	-0.3231	-0.4607
3×3	-15.3611	-31.8592
5×5	-7.9183	-37.8465

For different data formats, convolutional kernel size is an important hyper parameter for extracting features. Hence, we analyze how it effects the performance of the CNN and the ConvLSTM-net first. This paper trains the CNN and the ConvLSTM-net with different convolutional kernel sizes in frequency domain and time domain respectively. The corresponding NMSE results are summarized in Table 1 in which the best forecast results are marked in red. Notably, when we compute NMSE, the outputs of the ConvLSTM-net and the CNN are recovered to its original structure which is a complex matrix with dimension of 36×7. And CSI in time domain are transformed into frequency domain by Discrete Fourier Transform for justice. Fig. 6 visualizes the experiment results. The experiment results reveal that the CNN achieves best performance with kernel size of 3×3 and the ConvLSTM-net obtains best performance with kernel size

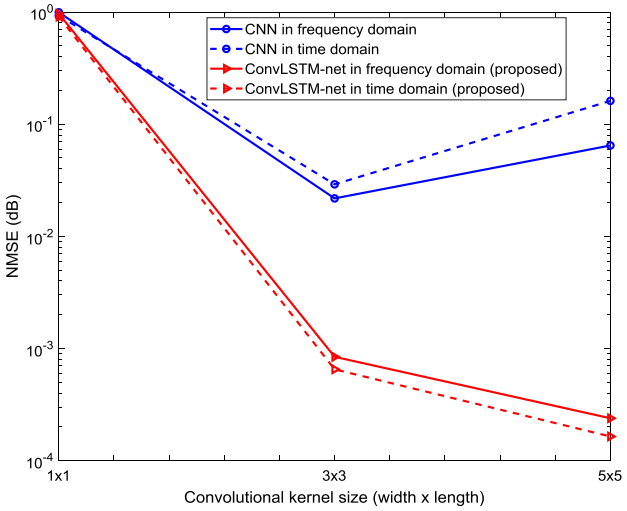


FIGURE 6. Performance comparisons of the CNN (conventional scheme) and the ConvLSTM-net (our proposed scheme) with different kernel size.

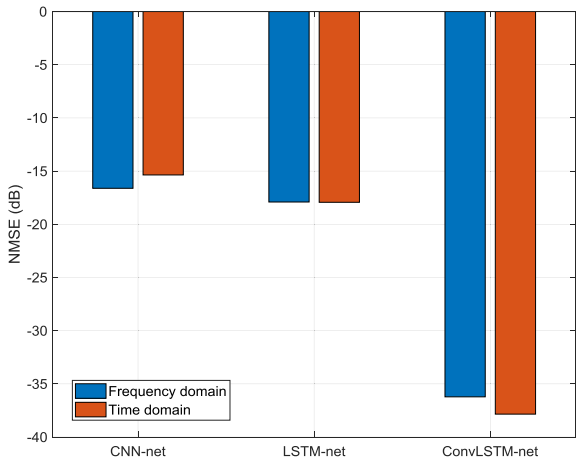


FIGURE 7. Performance comparisons of CNN (conventional scheme), LSTM-net (benchmark method) and ConvLSTM-net (our proposed scheme).

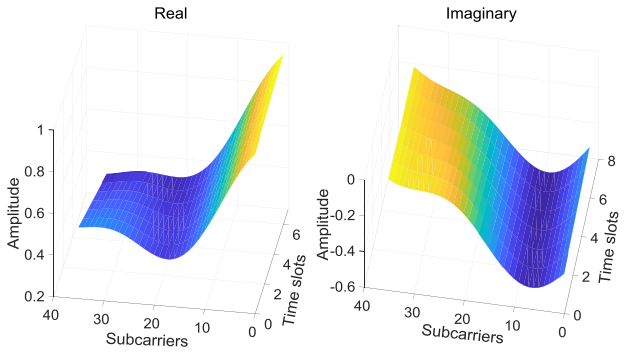


FIGURE 8. The original DL-CSI.

of 5×5. When kernel size is 1×1, convolutional operations learning nothing between adjacent nodes since the convolutional layer is equal to full a connected layer under this condition. Experiment results also reveal that the CNN and the

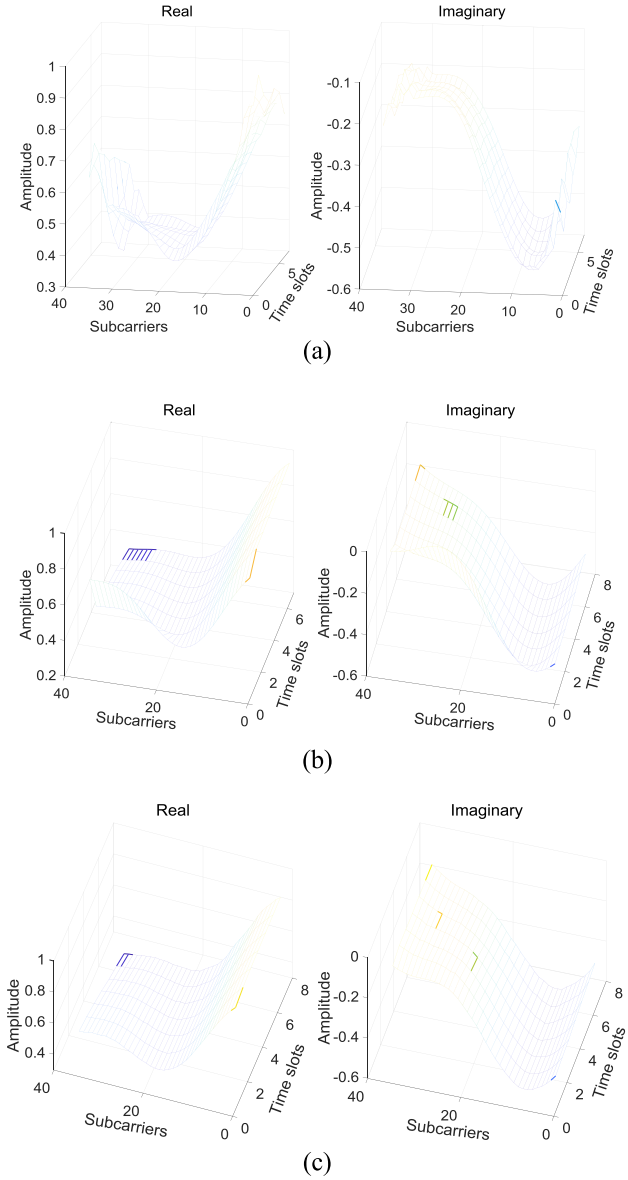


FIGURE 9. Prediction of the DL-CSI in frequency domain. (a) Prediction using the CNN; (b) Prediction using the LSTM-net; (c) Prediction using the ConvLSTM-net.

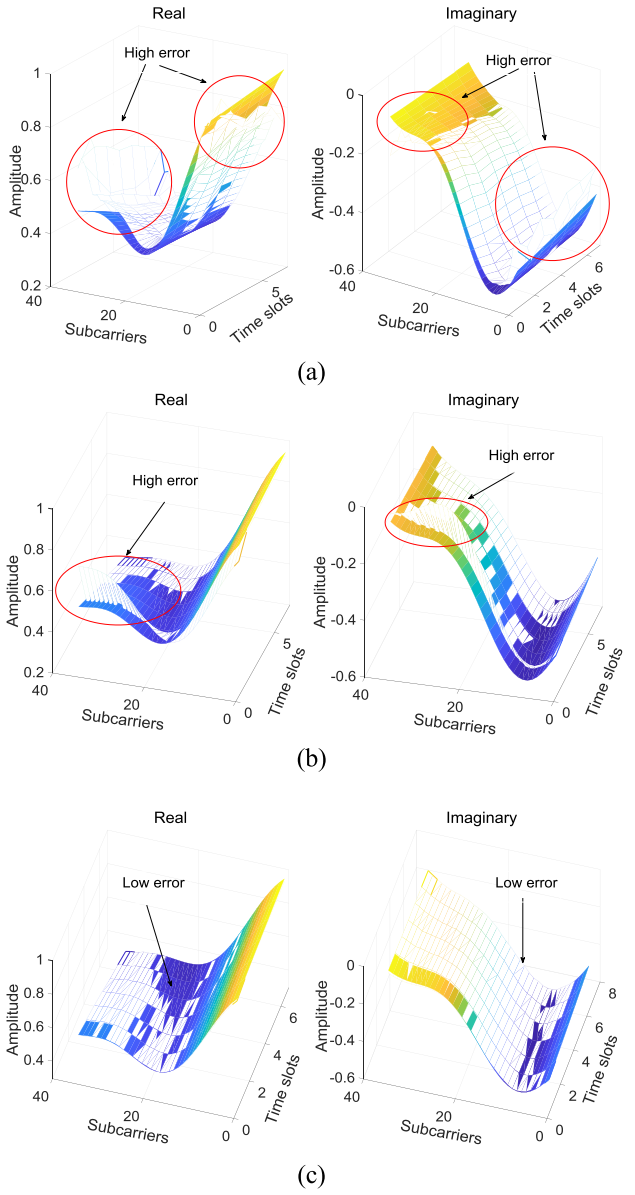


FIGURE 10. Comparison between the prediction and the original DL-CSI in frequency domain. (a) Comparative results of the CNN; (b) Comparative results of the LSTM-net; (c) Comparative results of the ConvLSTM-net.

ConvLSTM-net whether in frequency domain or time domain are all non-effective under 1×1 kernel size which denotes that almost correlations exist in adjacent subcarriers. With kernel size increases, convolutional operations learn features among bigger areas. For CNN, the input is an image with size of 36×7 (36 subcarriers and 7 time slots) and convolutional operations are conducted on this image. Compared to CNN, the input of ConvLSTM-net are 7 frames (7 time slots) images with size of 6×6 (36 subcarriers are reshaped into 6×6) and convolutional operations are operated on each 6×6 image. From Table. 1 and Fig. 6, we can see that whether in frequency domain and time domain, CNN with kernel size of 3×3 is superior to CNN with kernel size of 5×5 significantly (NMSE decreased about 6 dB). However, whether in frequency domain and time domain, the ConvLSTM-net with kernel size of 5×5 precedes the ConvLSTM-net with kernel size of 3×3 dramatically (NMSE decreased about 6 dB). Analysis from data formats of the two networks, the results demonstrate that larger kernel size can extract more features between subcarriers, however, time correlations between time slots can't be extracted perfectly by convolutional operations.

Compared to frequency domain, NMSE of the CNN with kernel size of 3×3 in time domain increased about 1.3 dB. However, compared to frequency domain, NMSE of the ConvLSTM-net with kernel size of 3×3 in time domain decreased about 1 dB and NMSE of the ConvLSTM-net with size of 5×5 in time domain decreased about 1.5 dB. The results demonstrate that, compared to frequency domain, CSI in time domain contains a bit more time information that can

be extracted by the ConvLSTM-net and a bit less space information that can be extracted by CNN. Whether in frequency domain or in time domain, the proposed ConvLSTM-net outperformance the CNN dramatically with about 20 dB gain.

In the following, we compare NMSE results of the ConvLSTM-net with the CNN and the LSTM-net, as shown in Fig. 7. From Fig. 7, we can see that the LSTM-net outperforms the CNN with about 1dB gain and the ConvLSTM-net achieves the best performance. For the LSTM-net, the prediction accuracy in time domain is little better than that in frequency domain, which also demonstrates that CSI in time domain contains a bit more time information.

To further analyze the performance of the ConvLSTM-net, we visualize the performance of the CNN, the LSTM-net and the ConvLSTM-net respectively. This paper plot one sample of original DL-CSI, predictions of the DL-CSI and the comparison between them in Figs. 8∼10 (conduct experiments in frequency domain only for limitation of space). In these figures, X-axis denotes the number of time slots of the channel matrix (i.e., columns), Y-axis denotes the number of subcarriers of the channel matrix (i.e., rows) and Z-axis denotes values of the channel matrix. From Fig. 9 and Fig. 10, we can see that the prediction result of the CNN has high errors at the edge subcarriers because of the principle of convolutional layers and the LSTM-net has high errors at the edge time slots because of the principle of LSTM layers. However, the proposed ConvLSTM-net method can achieve low NMSE at all subcarriers and all time slots due to its stronger ability of exploiting temporal correlations among adjacent time slots compared with the CNN and stronger power of exploiting spatial correlations among subcarriers compared with the LSTM-net.

V. CONCLUSION

In this paper, we have proposed a new ConvLSTM-net to enable the cellular FDD systems to predict their DL-CSI from the UL-CSI directly without any additional computation and feedback overheads. Besides, a LSTM-net has also been proposed as a benchmark. To evaluate the performance of the ConvLSTM-net, we have conducted comparative experiments in time domain and frequency domain respectively. In addition, we have explored the performance of the proposed networks with different kernel sizes. The corresponding experiment results have revealed that the proper kernel size is different for different networks due to their diverse data formats. And the ConvLSTM-net obtains the best performance among three networks due to its strong ability of exploiting temporal correlations as well as spatial correlations in data samples, especially in time domain.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[2] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[3] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 121–135, Jan. 2019.

[4] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178–2181, Dec. 2015.

[5] W. Shen, L. Dai, G. Gui, Z. Wang, R. W. Heath, Jr., and F. Adachi, "AoD-adaptive subspace codebook for channel feedback in FDD massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–5.

[6] P. W. C. Chan, E. S. Lo, R. R. Wang, E. K. S. Au, V. K. N. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. Ben Letaief, "The evolution path of 4G networks: FDD or TDD?" *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 42–50, Dec. 2006.

[7] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based Spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2016.

[8] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2012, pp. 492–497.

[9] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[10] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.

[11] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[12] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, 2017.

[13] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[14] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[15] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[16] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 584–588.

[17] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2018.

[18] M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2018.

[19] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[20] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[21] X. Zhang, L. Zhong, and A. Sabharwal, "Directional training for FDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5183–5197, Aug. 2018.

[22] L. Miretti, R. L. G. Cavalcante, and S. Stanczak, "FDD massive MIMO channel spatial covariance conversion using projection methods," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2018, pp. 3609–3613.

[23] D. Vasisht, S. Kumar, H. Rahul, and D. Katabi, "Eliminating channel feedback in next-generation cellular networks," in *Proc. ACM SIGCOMM Conf.*, Aug. 2016, pp. 398–411.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[25] M. S. Safari and V. Pourahmadi, "Deep UL2DL: Channel knowledge transfer from uplink to downlink," 2018, *arXiv:1812.07518*. [Online]. Available: https://arxiv.org/abs/1812.07518

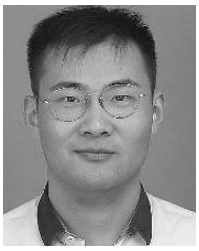
[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[27] X. Shi, Z. Chen, H. Wang, W.-K. Wong, W.-C. Woo, and D.-Y. Yeung, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

JIE WANG received the B.S. degree from the College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China, in 2015, where she is currently pursuing the Ph.D. degree. Her research interests include deep learning, channel estimation, and resource allocation and its application in wireless communications.

YING DING was born in Yancheng, China, in 1998. She is currently pursuing the B.E. degree in computer science from the Nanjing University of Posts and Telecommunications, Nanjing, China, where she has been a member with the Focus Lab, since 2018. Her current research interests include deep learning and computer vision.

SHUJIE BIAN was born in Yangzhou, China, in 1999. She is currently pursuing the B.S.E.E. degree in electronic information engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China. Her current research interests include signal processing and pattern recognition.

YANG PENG, photograph and biography not available at the time of publication.



MIAO LIU (M'18) received the B.Sc. degree in communication engineering from the University of Electric Science and Technology of China, Chengdu, China, in 2003, and the M.Sc. and Ph.D. degrees in communication engineering from Southeast University, Nanjing, China, in 2006 and 2018, respectively. He is currently an Assistant Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China.

GUAN GUI (M'11–SM'17) received the Dr. Eng. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2012.

From 2009 to 2012, with the financial support from the China Scholarship Council (CSC) and the Global Center of Education (ECOE) of Tohoku University, he joined the Wireless Signal Processing and Network Laboratory (Prof. Fumiyuki Adachi Laboratory), Department of Communications Engineering, Graduate School of Engineering, Tohoku University, as a Research Assistant and a Postdoctoral Research Fellow. From 2012 to 2014, he was supported by the Japan Society for the Promotion of Science (JSPS) Fellowship as a Postdoctoral Research Fellow with the Wireless Signal Processing and Network Laboratory. From 2014 to 2015, he was an Assistant Professor with the Department of Electronics and Information System, Akita Prefectural University. Since 2015, he has been a Professor with the Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China. He is currently involved in the research of deep learning, compressive sensing, and advanced wireless techniques. He has published more than 200 international peer-reviewed journal/conference papers.

Dr. Gui received Member and Global Activities Contributions Award in the IEEE ComSoc and seven best paper awards, i.e., ICEICT 2019, ADHIP 2018, CSPS 2018, ICNC 2018, ICC 2017, ICC 2014, and VTC 2014-Spring. He was also selected for Jiangsu Specially-Appointed Professor, in 2016, Jiangsu High-level Innovation and Entrepreneurial Talent, in 2016, Jiangsu Six Top Talent, in 2018, and Nanjing Youth Award, in 2018. He was an Editor of *Security and Communication Networks*, from 2012 to 2016. He has been the Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, since 2017, the Editor of *KSII Transactions on Internet and Information Systems*, since 2017, the Editor of IEEE ACCESS, since 2018, the Editor-in-Chief of *EAI Transactions on Artificial Intelligence*, since 2018, and the Editor of *Journal of Communications*, since 2019.