IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of current version 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.DOI

A Novel CSI Feedback Approach for Massive MIMO using LSTM-Attention CNN

QI LI[1], AIHUA ZHANG[1], PENGCHENG LIU[2], JIANJUN LI[1], and CHUNLEI LI[1]
[1]School of Electronic and Information Engineering, Zhongyuan University of Technology, Zhengzhou, CO 450007, China. (e-mail: zhah1229@sina.com)
[2]Department of Computer Science, University of York, York YO10 5DD, United Kingdom. (e-mail: pengcheng.liu@ieee.org)

Corresponding author: Aihua Zhang (e-mail: zhah1229@sina.com).

This study was supported by the National Natural Science Foundation of China under Grant (61501530, 61803396), Henan science and technology planning project under Grant 182102210522.

ABSTRACT In this paper, a novel mechanism is studied to improve the performance of the channel state information (CSI) feedback in massive multiple-input multiple-output (MIMO) systems. The proposed mechanism encompasses convolutional neural network (CNN)-based CSI compression and reconstruction structure. In this structure, the long-short term memory (LSTM) is adopted to learn temporal correlation of channels, and then, an attention mechanism is developed to perceive local information and automatically weight feature information. In addition, the CNN framework is further adjusted to reduce the number of training parameters and accelerate CSI recovery. The CNN structure with optimal training parameters can be achieved via offline iterative training and learning based on various training datasets. Comparative experimental studies demonstrate the effectiveness of the proposed approach that the trained CNN can obtain the higher feedback accuracy and better system performance in massive MIMO CSI online feedback reconstruction. Moreover, the proposed scheme in the less parameters-based neural network owns a higher performance with lower computational complexity compared to the conventional algorithms.

INDEX TERMS Massive MIMO, frequency division duplex (FDD), CSI feedback, long short-term memory (LSTM), Attention mechanism

I. INTRODUCTION

RECENTLY, massive MIMO is regarded as one of the key technologies for the next generation of wireless networks [1]. By equipping tens or hundreds of antennas at the base station (BS), a higher spectral efficiency and energy efficiency can be obtained [2]–[6]. The above potential benefits of massive MIMO technology depend on the exact CSI. In frequency division duplex (FDD) MIMO systems, the downlink CSI is firstly obtained at the user equipment (UE) via the downlink pilots, and then returned to the BS through uplink feedback links [7] [8]. However, with the application of a large number of antenna arrays at the BS, the uplink channel feedback overhead increases dramatically. Therefore, how to reduce the feedback overhead is a significant issue in a practical massive MIMO system.

In recent years, several algorithms have been proposed to reduce the feedback overhead [9]–[12]. The conventional vector quantization or codebook-based approaches can effectively reduce feedback overhead [9]. However, with the increase of the number of antennas, the feedback quantity increases dramatically, which leads to the a significant reduction in the efficiency of feedback. Based on compressive sensing (CS) theory, [10] proposed a distributed compressive channel state information at the transmitter (CSIT) estimation scheme to reduce the training as well as the feedback overhead in the CSIT estimation. Some other researches, including least absolute shrinkage and selection operator (LASSO) ℓ_1-solver [11] and approximate message passing (AMP) [12], have also been proposed to recover compressive CSI. Such solutions are based on the sparse priors which could hardly be restored because the practical channel is approximately sparse. In order to provide more accurate priors for reconstruction, the total variation augmented lagrangian (TVAL3) alternating direction algorithm [13] and block-matching and 3D filtering-AMP (BM3D-AMP) [14] are proposed via imposing hand-crafted priors to improve the channel reconstruction quality in some degree. However, the above algorithms struggle to improve

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2963896, IEEE Access

the accuracy of CSI reconstruction in a significant manner. Recently, deep learning (DL) theory has been successfully applied in wireless communications [15]–[19], particularly in channel feedback [17] [18]. In [17], the authors provided an auto encoder-decoder called CsiNet. In this algorithm, an encoder compresses the feature vector, and then a decoder decompresses the information and recover CSI. However, the time correlation characteristics of the time-varying channels are not fully explored to further improve the performance. [18] focuses on enhancing the performance of the channel recovery module, whereas cause the problem of excessive training parameters. In addition, the linear fully-connected network (FCN) is adopted in [17] [18] as the compression and decompression modules, however, it failed to thoroughly exploit the channel feature information, which in turn reduces the accuracy of channel reconstruction. Different from the previous works, this paper focuses on enhancing the ability of feature representation at the compression and decompression modules, and aims to improve the accuracy of channel feedback. The main contributions include:

- We propose a novel CSI feedback and recovery mechanism in the FDD massive MIMO system called LSTM-Attention CsiNet. Compression and decompression modules of the mechanism employ a LSTM-Attention to enhance the performance of channel recovery, respectively. The LSTM-Attention takes advantage of the memory characteristic of LSTM to learn temporal correlation of channel, further, the attention mechanism [20] is introduced to prioritize and weight the feature information automatically, which can allocate more attention to the important feature information.

- We also propose a lightweight LSTM-Attention CsiNet by adjusting connection mode between the LSTM-Attention and FCN. This lightweight network adopts FCN compress feature information to a lower vector and then input to LSTM-Attention, which effectively reduces the number of weights and biases from LSTM-Attention and accelerates the channel recovery.

II. THE SYSTEM MODEL

Consider a FDD massive MIMO system, where the BS is equipped with N_t transmit antennas with uniform linear array (ULA) and the user equipment (UE) is equipped with one single antenna. The transmit signals are operated by orthogonal frequency division multiplexing (OFDM) with $\widetilde{N}_c$ subcarriers and Rayleigh flat fading channel. The massive MIMO wireless communication system is shown in Figure. 1.

In the downlink of the FDD system, the BS pre-encodes the data-bearing signals and then transmits the signals through the wireless channel. The received signal of nth subcarrier at the receiver is

$$y_n = \widetilde{\boldsymbol{h}}_n^H \boldsymbol{v}_n x_n + z_n \tag{1}$$

where $\widetilde{\boldsymbol{h}}_n \in C^{N_t \times 1}$ is the channel vector at nth subcarrier, $\boldsymbol{v}_n \in C^{N_t \times 1}$ represents the precoding vector at the nth subcarrier, $x_n \in C$ denotes the modulated data symbol, and

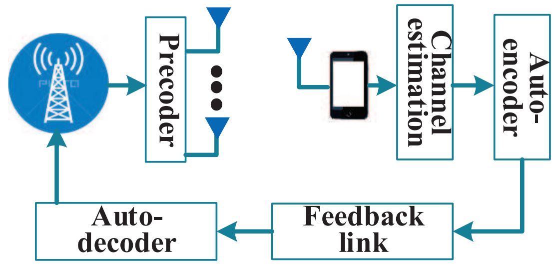


FIGURE 1. Diagram of channel feedback of massive MIMO system.

$z_n \in C$ denotes the additive Gaussian white noise with zero mean and unit variance. $\widetilde{\boldsymbol{H}} = [\widetilde{\boldsymbol{h}}_1 \widetilde{\boldsymbol{h}}_2 ... \widetilde{\boldsymbol{h}}_{\widetilde{N}_c}]^H \in C^{\widetilde{N}_c \times N_t}$ is the CSI in the spatial frequency domain.

After receiving the pilot signals, the user firstly calculates $\widetilde{\boldsymbol{H}}$ through channel estimation module, and then, sends back to BS through uplink channel. There are $\widetilde{N}_c N_t$ parameters will be sent in total, which will lead to large feedback overhead in a massive MIMO system. Here, we adopt compressive theory to reduce the parameters and then decrease the feedback overhead. In this work, the channel matrix $\widetilde{\boldsymbol{H}}$ can be transformed into approximately sparse matrix $\overline{\boldsymbol{H}}$ through Discrete Fourier Transform (DFT), which can be defined as

$$\overline{\boldsymbol{H}} = \boldsymbol{R}_d \widetilde{\boldsymbol{H}} \boldsymbol{R}_a^H \tag{2}$$

where $\boldsymbol{R}_d \in C^{\widetilde{N}_c \times \widetilde{N}_c}$ and $\boldsymbol{R}_a \in C^{N_t \times N_t}$ are all DFT matrices. As the time delay between multipath arrivals lies within a limited period, the processed $\overline{\boldsymbol{H}}$ is a sparse matrix [17], where the nonzero elements are the first N_c rows of $\overline{\boldsymbol{H}}$ contain values. We only preserve nonzero elements, denoting $\boldsymbol{H} \in C^{N_c \times N_t}$ as the truncated information matrix. Then, the auto-encoder compresses the processed CSI information $\boldsymbol{H}$ to codeword, and sends back to BS for CSI recovery through uplink channel. At the BS, the CSI can be obtained by the auto-decoder.

III. THE PROPOSED CSI FEEDBACK NEURAL NETWORKS

In this section, we combine LSTM-Attention CsiNet in conventional neural network for MIMO channel feedback, which includes two steps: encoding and decoding as shown in Figure. 2. Further, a Reduced LSTM-Attention CsiNet is also utilized to reduce training parameters and complexity of network in subsection B.

A. LSTM-ATTENTION CSINET

In this subsection, a CSI feedback structure based on LSTM-Attention, called LSTM-Attention CsiNet, is proposed to relieve the problem of CSI feedback overhead in FDD massive MIMO system. In this structure, $N_1 \times N_2 \times N_3$ are the length, width, and the numbers of feature maps. At UE, in order to reduce computational complexity, the real and imaginary parts of the truncated $\boldsymbol{H} \in C^{N_c \times N_t}$ is separated and concatenated into $\boldsymbol{H}' \in R^{N_c \times N_t \times 2}$. The processed $\boldsymbol{H}'$ can be fed into encoder, and then encoded to codewords.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2963896, IEEE Access

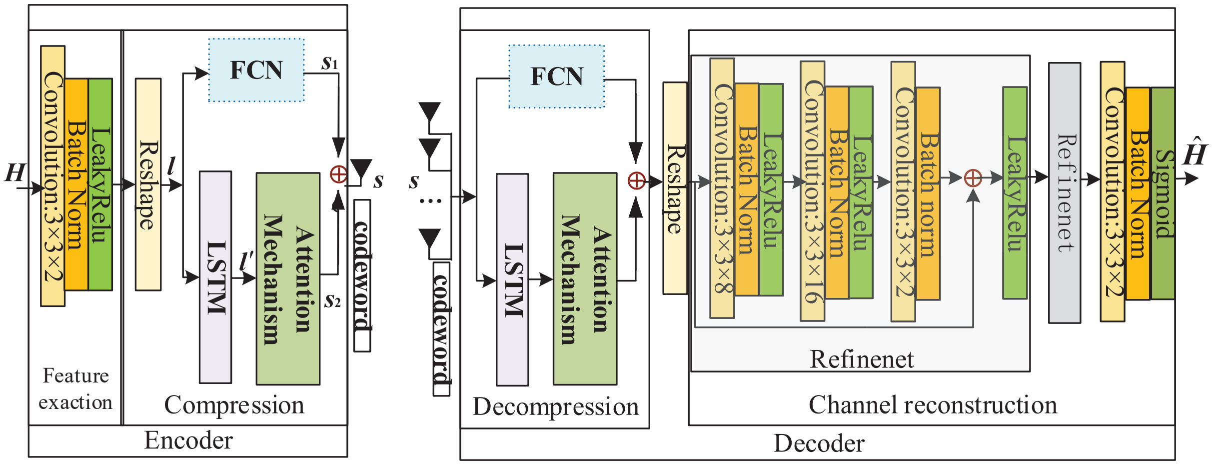


FIGURE 2. The structure of proposed LSTM-Attention based neural network which is used to be auto encoder-decoder.

One encoder consists of two parts in series: feature extraction and feature compression. The feature exaction is accomplished with a convolutional layer, batch normalization layer and leaky rectified linear unit (Leaky Relu). The convolutional operation is achieved with two 3×3 dimensional kernels. Then in the feature compression, we reshape the two feature matrices with $N_t \times N_c$ size into one feature vector as $\boldsymbol{l} \in R^{N\times 1}$, where $N = 2N_tN_c$, and then the $\boldsymbol{l} \in R^{N\times 1}$ is input into two paths:

A linear FCN: The FCN is a jump connection which transforms vector $\boldsymbol{l}$ into vector $\boldsymbol{s}_1 \in R^{M\times 1}$. This network can accelerate the convergence and solve the vanishing gradient problem [21].

A LSTM-Attention: The input of this module is $\boldsymbol{l}$ and the output is $\boldsymbol{s}_2 \in R^{M\times 1}$. LSTM-Attention can enhance performance of the feature compression by learning time correlation and calculating soft probability distribution fitting.

1) Long Short-term Memory(LSTM)

LSTM network is proposed to mitigate the problem of vanishing/exploding gradient by using a memory unit to keep the dependent information contained in sequential data for a long period of time. There is an intrinsic memory function in LSTM unit, which can preserve the information extracted before long-term. In this paper, LSTM is used to learn temporal correlation from previous time steps in feature compression and feature decompression modules. Every LSTM unit consists of forget gate, input gate and output gate, as depicted in Figure. 3. The results of forget gate and input gate act on cell state updates. t denotes the time step. The mathematical representation of the LSTM structure is as follows:

$$\boldsymbol{f}_t=\sigma(\boldsymbol{W}_{lf}\boldsymbol{l}_t+\boldsymbol{W}_{mf}\boldsymbol{m}_{t-1}+\boldsymbol{b}_f) \quad (3)$$

$$\boldsymbol{i}_t=\sigma(\boldsymbol{W}_{li}\boldsymbol{l}_t+\boldsymbol{W}_{mi}\boldsymbol{m}_{t-1}+\boldsymbol{b}_i) \quad (4)$$

$$\boldsymbol{o}_t=\sigma(\boldsymbol{W}_{lo}\boldsymbol{l}_t+\boldsymbol{W}_{mo}\boldsymbol{m}_{t-1}+\boldsymbol{b}_o) \quad (5)$$

$$\boldsymbol{a}_t= tanh(\boldsymbol{W}_{la}\boldsymbol{l}_t+\boldsymbol{W}_{ma}\boldsymbol{m}_{t-1}+\boldsymbol{b}_a) \quad (6)$$

$$\boldsymbol{c}_t=\boldsymbol{c}_{t-1} \otimes \boldsymbol{f}_t+\boldsymbol{i}_t \otimes \boldsymbol{a}_t \quad (7)$$

$$\boldsymbol{m}_t=\boldsymbol{o}_t \otimes \tanh \boldsymbol{c}_t \quad (8)$$

where σ is the logistic sigmoid function, $\boldsymbol{f}_t$, $\boldsymbol{i}_t$, $\boldsymbol{o}_t$ are forget gate, input gate, and output gate, respectively. $\boldsymbol{c}_t$, $\boldsymbol{a}_t$ are memory cell and hidden vector, respectively. $\boldsymbol{W}_{l*} = \{\boldsymbol{W}_{lf}, \boldsymbol{W}_{li}, \boldsymbol{W}_{la}, \boldsymbol{W}_{lo}\}$ and $\boldsymbol{W}_{m*} = \{\boldsymbol{W}_{mf}, \boldsymbol{W}_{mi}, \boldsymbol{W}_{ma}, \boldsymbol{W}_{mo}\}$ are recurrent weights of the corresponding gates, $\boldsymbol{b}_f$, $\boldsymbol{b}_i$, $\boldsymbol{b}_o$, $\boldsymbol{b}_a$ are output biases. $\otimes$ denotes the Hadamard product.

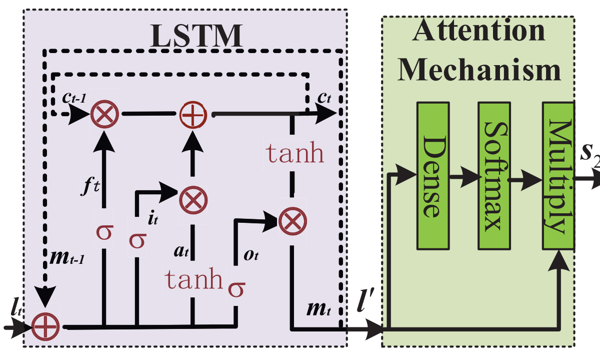


FIGURE 3. LSTM-Attention.

In the LSTM algorithm, the input sequences, regardless of their length, are encoded into a fixed-length vector representation, which has limited features in terms of decoding and will therefore reduce the representational ability [22]. Therefore, an attention mechanism is adopted to enhance functionality of learning time pertinence after the above LSTM units.

2) Attention mechanism

As described in Figure. 3, the nuclear processes of the attention mechanism are the dense layer and softmax activation function. The output signal $\boldsymbol{l}'=[\boldsymbol{m}_1\boldsymbol{m}_2\cdots\boldsymbol{m}_T]$ of LSTM units is regarded as the input of the attention mechanism. The output sequence $\boldsymbol{s}_2$ of attention unit can be computed with the linear weighted sum of $\boldsymbol{l}'$, $\boldsymbol{s}_2$ can be denoted as:

$$(\boldsymbol{s}_2)_i = \sum_{j}^{T} \boldsymbol{\lambda}_{ij} \boldsymbol{m}_j \tag{9}$$

where the weight factor $\boldsymbol{\lambda}_{ij}$ is soft probability distribution fitting ability of each annotation of $\boldsymbol{m}_j$. And $\boldsymbol{\lambda}_{ij}$ is obtained from the weight normalization which can be accomplished with the help of softmax activation function, and $\boldsymbol{\lambda}_{ij}$ can be expressed by:

$$\boldsymbol{\lambda}_{ij} = \frac{\exp(\boldsymbol{\beta}_{ij})}{\sum\limits_{k=1}^{T} \exp(\boldsymbol{\beta}_{ik})} \tag{10}$$

where $\boldsymbol{\beta}_{ij}$=$\boldsymbol{a}(\boldsymbol{\varphi}_{i-1}$,$\boldsymbol{m}_j)$ =$\boldsymbol{W}_{i-1}^{mech}\boldsymbol{m}_j$ is an alignment model operated by the dense layer, $\boldsymbol{W}_{i-1}^{mech}$ is the weights of $(i-1)$th annotation. $\boldsymbol{\beta}_{ij}$ denotes the weight getting from similarity calculation and it represents the similarity correlational between $\boldsymbol{\varphi}_{i-1}$ and $\boldsymbol{m}_j$.

The attention mechanism can be deployed to CSI feedback network and combined with others neural network module. This mechanism prioritizes the promising series through local perception and soft decision, and enhances the representation of temporal correlational feature information of the MIMO channels.

After the FCN and LSTM-Attention modules, the output of encoder is obtained by integrating feature information $\boldsymbol{s}_1$ from FCN and $\boldsymbol{s}_2$ from LSTM-Attention with addition, where the encode rate of encoder is r=$\frac{M}{N}$. This operation can be shown as $\boldsymbol{s} = \boldsymbol{s}_1 + \boldsymbol{s}_2$. Then, the M dimensional compressed vector is transmitted back to BS for CSI recovery through uplink channel. During the feedback transmission, the feedback channel is assumed to be perfect enough to transmit the compressed codeword without any loss.

At the BS, the codeword $\boldsymbol{s}$ received from UE is used to recover the truncated matrix $\boldsymbol{H}$ by adopting the decoder which is formed with feature decompression and channel recovery in series. Firstly, the feature decompression module is formed with LSTM-Attention and FCN in parallel, which is the same module as the feature compression in the encoder. The difference is that the input size of decompression is $M \times 1$ and the output size is $N \times 1$. Secondly, following the decompression operation, the output codeword $\boldsymbol{l}_d \in R^{N\times 1}$ is fed into the "RefineNet" units. These units can solve vanishing gradient problem during the process of refinement feature, which in turn will increase the accuracy of information reconstruction [17]. Finally, the reconstructed channel matrix $\widehat{\boldsymbol{H}}$ is imbedded into the original channel matrix $\widetilde{\boldsymbol{H}}$ recovery through nonzero connection and inverse DFT operations.

The reconstruction performance of the auto encoder-decoder depends on large-scale numbers training parameters. These parameters are mainly from compression and decompression modules. However, the training parameters stores vast redundant information, and not all parameters and substructures play an important role of network, which in turn increase the complexity of the feedback system and channel recovery time. The following new architecture is proposed based on the Reduced LSTM-Attention CsiNet with low complexity.

B. REDUCED LSTM-ATTENTION CSINET

The proposed LSTM-Attention CsiNet has the disadvantage of too many training parameters, which in turn increases the complexity of channel system. Here, we provide a more effective network framework. As shown in Figure. 4, the compression module consists of the FCN and the LSTM-Attention in series rather than in parallel. Moreover, in this way, we integrate the feature information from the FCN and the LSTM-Attention modules to alleviate the gradient disappearance. For feature compression module, a M dimensional sequence from FCN is transmitted to LSTM-Attention module directly, inspired by Lu. C in [23]. The feature vector $\boldsymbol{l}$ with N dimension from the feature exaction module is transformed into $\boldsymbol{s}_1$ with M dimension, keeping size unchanged from input to output in LSTM-Attention. Moreover, in the feature decompression module, the received $\boldsymbol{s}$ vector with M dimension from UE is translated into M dimension through LSTM-Attention, and then integrate the output of the LSTM-Attention with $\boldsymbol{s}$ vector. Consequently, the jump connection on the LSTM-Attention does not need any dimension transformation before the addition operation. Then the obtained M dimensional vector is fed into FCN to decompress the feature vector.

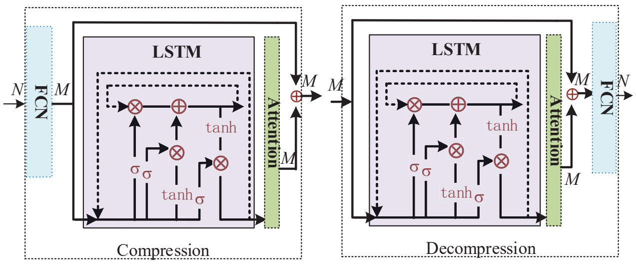


FIGURE 4. Reduced parameters LSTM-Attention.

The above two channel feedback networks are trained end-to-end by iteration, aiming at minimizing the loss function, and then obtaining the best kernel and biases of networks. Here, we use mean squared error (MSE) as the loss function to measure the reconstruction performance, and use adaptive moment estimation (ADAM) optimization algorithm to update the series of parameters via gradient in each iteration. The MSE of the estimated channel matrix is calculated as follows:

$$L(\boldsymbol{\theta}_e, \boldsymbol{\theta}_d) = \frac{1}{K}\sum_{i=1}^{K} \| f_d(f_e(\boldsymbol{H}_i, \boldsymbol{\theta}_e), \boldsymbol{\theta}_d) - \boldsymbol{H}_i\|_2^2 \tag{11}$$

where $\boldsymbol{\theta}_e$, $\boldsymbol{\theta}_d$ are parameters of the encoder and decoder, respectively. f_e , f_d are the networks of encoder and decoder, respectively. K denotes the batch size, and $\boldsymbol{H}_i$ is CSI at i th iteration.

The proposed two algorithm diagrams are shown in Table 1, including offline training and online recovery. The trained neural network can be directly used to CSI feedback.

TABLE 1. Algorithm diagram with training and recovery for CSI feedback

Algorithm: (Reduced) LSTM-Attention CsiNet channel feedback algorithm for massive MIMO System

Input: Generated channel matrix $\widetilde{\boldsymbol{H}}$
Output: The CSI feedback matrix $\widehat{\boldsymbol{H}}$
Offline training:
Step1: Generate original channel matrix $\widetilde{\boldsymbol{H}}$ based on COST 2100 MIMO module [24]. Preprocess the channel matrix through DFT and truncation operation. Separate the real and imaginary parts. And then concatenate them to be effect matrix $\boldsymbol{H}'$ with a new size.
Step2: At UE, input the matrix $\boldsymbol{H}'$ to the encoder. Extract the feature vector by convolution and reshaping operation. Compress the vector to codeword $\boldsymbol{s}$ by FCN and LSTM-Attention.
Step3: Transmit the codeword $\boldsymbol{s}$ received from encoder of UE to BS.
Step4: At the BS, decompress $\boldsymbol{s}$ through FCN and LSTM-Attention to a vector with $2N_tN_c$dimension, and then recover it back into CSI through the Refinenet units.
Step5: Compute the loss function L with MSE and update the training parameters of structure using ADAM optimization to reduce the error between the original effect matrix $\boldsymbol{H}$ and estimated matrix $\widehat{\boldsymbol{H}}$.
Repeat **Step2** to **Step4** through iteration until obtaining the optimal CSI feedback neural network.
Online recovery:
At UE, the estimated CSI can be imported to the auto encoder module of the optimal neural network and sent to BS. At BS, the received vectors can be transmitted to decoder and used for CSI feedback without lots of iterations, effectively reducing the complexity of channel feedback system.

IV. SIMULATION AND DISCUSSIONS

In this section, for indoor and outdoor two kinds of scenarios, experimental simulations are developed to verify the effectiveness of the proposed massive MIMO channel feedback algorithms respectively. We compare the performance of the proposed algorithms with some prevailing feedback algorithms through tensorflow on PC with Nvidia Geforce GTX 1080 Ti GPU. During offline training, channel matrices datasets, including training, validation and testing sets, are obtained by COST 2100 MIMO channel model [24]. Total parameters of the simulation of the MIMO channel model are listed in Table. 2.

TABLE 2. Simulation parameters

Parameters	Setting
The cost 2100 Channel model [24]	Indoor picocellular: 5.3 GHz and Outdoor rural: 300MHz
Antennas	N_t= 32
Subcarriers	$\widetilde{N}_c$= 1024
Truncated length	N_c= 32
Duplex mode	FDD
Modulation	Flat Rayleigh fading

A. NMSE PERFORMANCE OF THE PROPOSED ALGORITHMS

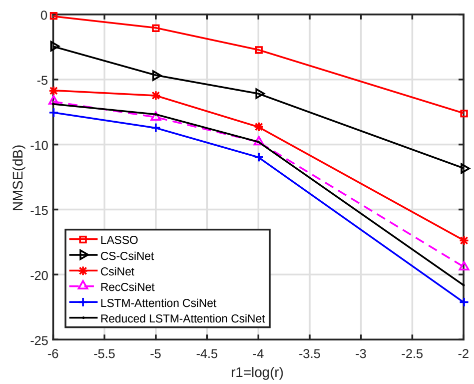
The training, validation and testing datasets have 100000, 30000 and 20000 samples, respectively. The batch size, epochs and learning rate are chosen as K= 200 , 1000 and 0.001, respectively. There are six LSTM units before attention mechanism at compression and decompression modules.

In this paper, we apply normalized mean square error (NMSE) to compare performance of the proposed novel LSTM-Attention CsiNet CSI feedback algorithms with LASSO [11], CS-CsiNet [17], CsiNet [17] and RecCsiNet [23] under compression rate r , $r = \frac{M}{N} \in \left\{\frac{1}{4}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}\right\}$, NMSE is shown as

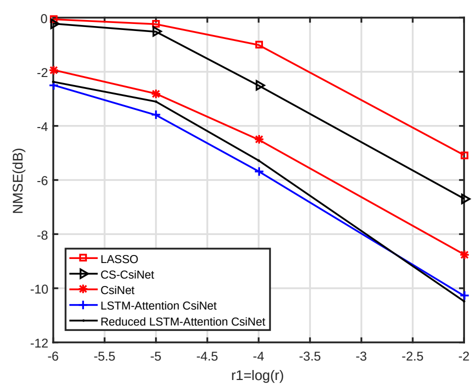
$$NMSE = E\left\{\frac{||\widehat{\boldsymbol{H}}-\boldsymbol{H}||_2^2}{||\boldsymbol{H}||_2^2}\right\} \tag{12}$$

It is noted that the smaller the NMSE is, the smaller the CSI recovery error rate and the better the performance will be.

As can be seen from Figure. 5, for communication system equipped with 32 transmit antennas at BS and single-user at receiving terminal. For the case of indoor picocellilar scenario, the NMSE value of the novel LSTM-Attention CsiNet CSI feedback framework is smaller than LASSO, CS-CsiNet, CsiNet and RecCsiNet algorithms, particularly, about 4 dB smaller than the CsiNet neural network for code rate $r = \frac{1}{4}$. Moreover, for outdoor rural scenario at 300MHz



(a) Indoor picocellular scenario



(b) Outdoor rural scenario

FIGURE 5. NMSE performance comparison with the prevailing algorithms

TABLE 3. The comparison of parameters and runtime with the prevailing algorithms

Code-rate	CsiNet		LSTM-Attention CsiNet		Reduced LSTM-Attention CsiNet	
	Parameters	Δt	Parameters	Δt	Parameters	Δt
$\frac{1}{4}$	2103904	0.0001	10247148	0.008	2924912	0.004
$\frac{1}{16}$	530656	0.0001	7484688	0.008	582704	0.001
$\frac{1}{32}$	268448	0.0001	7024272	0.008	281936	0.0007
$\frac{1}{64}$	137344	0.0001	6794064	0.008	141152	0.0005

band, the NMSE value is about 1.5-3 dB smaller than CsiNet. Figures. 5 shows that regardless of it is at indoor or outdoor, the proposed LSTM-Attention CsiNet has an outstanding NMSE performance. At the compression and decompression modules, the network is adopted LSTM units to make fully use of the time correlation of the massive MIMO channel information, with an employed attention mechanism to better learn the channel structure feature. Therefore, the LSTM-Attention CsiNet demonstrates a more effective performance in terms of reconstruction of the channel matrix.

In addition, with the same experiment condition, the test results of the Reduced LSTM-Attention CsiNet channel feedback system are shown in Figure. 5. For the indoor scenario, the Reduced LSTM-Attention CsiNet achieves better NMSE performance compared to the other prevailing algorithms, in which the curves are smaller than CsiNet about 1-3 dB. For the outdoor scenario, the Reduced LSTM-Attention CsiNet also shows preferable performance, which better 0.5-2dB than that of the CsiNet. FCN in compression and decompression modules of the Reduced LSTM-Attention CsiNet, as a jump connection network, maintains the input and output dimensions of LSTM-Attention module unchanged, which in turn decreases the number of parameters and reduces complexity of the feedback network.

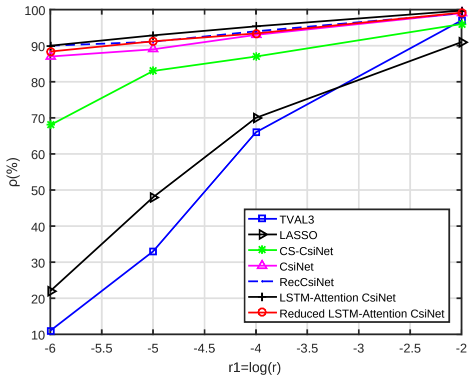
B. COSINE SIMILARITY ANALYSIS OF THE PROPOSED ALGORITHMS

In order to further verify the effectiveness of the proposed algorithms, we consider cosine similarity by calculating

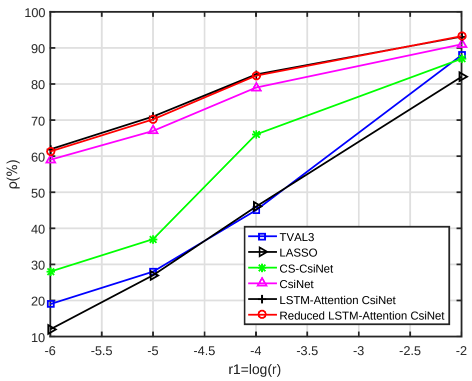
$$\rho = E\left\{ \frac{1}{\widetilde{N}_c} \sum_{n=1}^{\widetilde{N}_c} \frac{|\widehat{\boldsymbol{h}}_n^H \widetilde{\boldsymbol{h}}_n|}{||\widehat{\boldsymbol{h}}_n^H||_2 ||\widetilde{\boldsymbol{h}}_n||_2} \right\} \tag{13}$$

where $\widehat{\boldsymbol{h}}_n$ denotes the reconstructed channel vector of the nth subcarrier, and $\widehat{\boldsymbol{h}}_n^H / ||\widehat{\boldsymbol{h}}_n||_2$ is used as a beamforming vector. The greater the ρ is, the better the performance will be.

The cosine similarity performance, as beamforming gain, is used to measure channel reconstruction accuracy of the massive MIMO system. For single-user at UE, 32 transmit antennas at BS, the Figures. 6 shows the ρ performance of some algorithms, including TVAL3 [13], LASSO [11], CS-CsiNet [17], CsiNet [17], RecCsiNet [23] and the proposed two feedback algorithms. For indoor environment with 5.3 GHz, the performance of the proposed two algorithms,



(a) Indoor picocellular scenario



(b) Outdoor rural scenario

FIGURE 6. ρ performance comparison with the prevailing algorithms

the Reduced LSTM-Attention CsiNet and LSTM-Attention CsiNet, are better than other conventional algorithms and better 1% and 1.5% than CsiNet in average respectively. For the case of outdoor environment, the ρ performance also improves as the figures show. Since the features extracted from the channel matrix is compressed with the FCN to a lower dimension, it could bring some information loss compared to LSTM-Attention CsiNet. Besides, the greater the M is, the greater the code rate is and the better the CSI feedback accuracy will be.

It is worth mentioning that the promising performance of LSTM-Attention CsiNet channel feedback depends on a series of training parameters including weights and biases, which increases complexity of the recovery network. The parameters mainly from compression and decompression modules. For LSTM-Attention, the $\boldsymbol{W}_x$, $\boldsymbol{W}_y$ and $\boldsymbol{b}$ have $N_{In} \times N_{lstm}$, $N_{In} \times N_{In}$ and N_{In} parameters at LSTM module respectively, attention mechanism has $M^2 + M$ parameters, where N_{In} and N_{lstm} are input dimension of LSTM and the number of LSTM units respectively. Furthermore, FCN has $NM + M$ parameters. For the Reduced LSTM-Attention CsiNet compression module, the $\boldsymbol{W}_x$,$\boldsymbol{W}_y$ and $\boldsymbol{b}$ also has $N_{In} \times N_{lstm}$, $N_{In} \times N_{In}$ and N_{In} parameters at LSTM module respectively, attention mechanism has $(\frac{M}{N_{In}})^2 + \frac{M}{N_{In}}$ parameters, and FCN has $NM+M$ parameters. In general, the number of parameters is different which is mainly from attention mechanism. In this paper, Table. 3 contains statistically the total parameters of the CsiNet, LSTM-Attention CsiNet and Reduced LSTM-Attention CsiNet algorithms. We can see that the less training parameters are, the simpler the neural network and the faster the channel information feedback will be. The LSTM-Attention CsiNet and Reduced LSTM-Attention CsiNet apply LSTM units to selectively remember information and learn temporal correlation, and then adopt attention mechanism to visualize the decision process of LSTM units and compute a soft alignment directly, improving reconstruction accuracy of channel matrices. Because the FCN compresses vector from N to M size, which leads to the loss of some characteristic information, the performance of the Reduced LSTM-Attention CsiNet is inferior to that of the LSTM-Attention CsiNet.

V. CONCLUSIONS

In this paper, the CSI auto encoder-decoder feedback algorithm based on temporal correlation for FDD massive MIMO system is studied. Aiming at the problem of the inefficient and high complexity of traditional based on codebook and compressive sensing CSI feedback algorithms, LSTM-Attention CsiNet is proposed to FDD massive MIMO system, which is considered as auto encoder-decoder mechanism based on deep learning. Based on massive datasets, the algorithm explores temporal correlation by selective memorizing with LSTM, allocates more attention to the important feature information by automatic weighting with attention mechanism, and then, learns CSI feature information, improving the accuracy of channel matrices recovery. Moreover, in order to reduce the complexity of CSI feedback network, this paper propose the Reduced LSTM-Attention CsiNet algorithm, which transforms feature vectors from N into M dimension with FCN, reducing the number of training parameters and complexity of the network, and therefore speeding up the CSI feedback.

REFERENCES

[1] C. Li, P. Liu, C. Zou, F. Sun, J. M. Cioffi and L. Yang, "Spectral-Efficient Cellular Communications with Coexistent One- and Two-Hop Transmissions," in IEEE Transactions on Vehicular Technology, vol. 65, no. 8, pp. 6765-6772, Aug. 2016.

[2] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari and F. Adachi, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," IEEE Wireless Communications Magazine, to be published, doi: 10.1109/MWC.2019.1900027.

[3] Chunguo L I, Yanshan L I, Song K, et al, "Energy efficient design for multiuser downlink energy and uplink information transfer in 5G," Science China, Vol.59, NO.2, pp.1-8, February, 2016.

[4] H. Huang, Y. Peng, J. Yang, W. Xia, G. Gui, "Fast beamforming design via deep learning," IEEE Transactions on Vehicular Technology, to be published, doi: 10.1109/TVT.2019.2949122.

[5] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," IEEE Transactions on Vehicular Technology, vol. 67, no. 9, pp. 8440-8450, Sept. 2018.

[6] 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 Transactions on Vehicular Technology, vol. 67, no. 9, pp. 8549-8560, Sept. 2018.

[7] C. K. Sung, I. Ahmad, G. Lechner and H. Suzuiki, "Performance Analysis of Distributed Transmit Beamforming with Quantized Channel Feedback," 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia, 2019, pp. 1-5.

[8] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, G. Gui, "UL-CSI data driven deep learning forpredicting DL-CSI in cellular FDD systems," IEEE Access, vol. 7, no. 1, pp. 1-10, 2019.

[9] W. Shen, L. Dai, B. Shim, Z. Wang and R. W. Heath, "Channel Feedback Based on AoD-Adaptive Subspace Codebook in FDD Massive MIMO Systems," in IEEE Transactions on Communications, vol. 66, no. 11, pp. 5235-5248, Nov. 2018.

[10] X. Rao and V. K. N. Lau, "Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems," in IEEE Transactions on Signal Processing, vol. 62, no. 12, pp. 3261-3271, June15, 2014.

[11] Daubechies I, Defrise M, De Mol C, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Communications on Pure & Applied Mathematics, Vol. 57, NO. 11, pp. 1413-1457, August, 2003.

[12] B. C. Jung and Woohyuk Chang, "A message passing algorithm for compressed sensing in wireless random access networks," 2013 19th Asia-Pacific Conference on Communications (APCC), Denpasar, 2013, pp. 463-464.

[13] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill and T. Q. Nguyen, "An Augmented Lagrangian Method for Total Variation Video Restoration," in IEEE Transactions on Image Processing, vol. 20, no. 11, pp. 3097-3111, Nov. 2011.

[14] C. A. Metzler, A. Maleki and R. G. Baraniuk, "From Denoising to Compressed Sensing," in IEEE Transactions on Information Theory, vol. 62, no. 9, pp. 5117-5144, Sept. 2016.

[15] N. Samuel, T. Diskin and A. Wiesel, "Deep MIMO detection," 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, 2017, pp. 1-5.

[16] H. Huang, Y. Song, J. Yang, G. Gui, "Deep-Learning-Based Millimeter-Wave Massive MIMO for Hybrid Precoding," in IEEE Transactions on Vehicular Technology, vol. 68, no. 3, pp. 3027-3032, March 2019.

[17] C. Wen, W. Shih and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," in IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748-751, Oct. 2018.

[18] T. Wang, C. Wen, S. Jin and G. Y. Li, "Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels," in IEEE Wireless Communications Letters, vol. 8, no. 2, pp. 416-419, April 2019.

[19] Y. Wang, M. Liu, J. Yang, G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 4074-4077, Apr. 2019.

[20] F. Karim, S. Majumdar, H. Darabi and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification," in IEEE Access, vol. 6, pp. 1662-1669, 2018.

[21] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[22] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv preprint arXiv: 1409.0473, 2014.

[23] C. Lu, W. Xu, H. Shen, J. Zhu and K. Wang, "MIMO Channel Information Feedback Using Deep Recurrent Network," in IEEE Communications Letters, vol. 23, no. 1, pp. 188-191, Jan. 2019.

[24] L. Liu et al., "The COST 2100 MIMO channel model," in IEEE Wireless Communications, vol. 19, no. 6, pp. 92-99, December 2012.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2963896, IEEE Access

VI. ABBREVIATIONS

TABLE 4. Abbreviations Table

Abbreviations	Full Name
CSI	Channel State Information
MIMO	Multi-input Multi-output
LSTM	Long Short-term Memory
CNN	Convolutional Neural Network
BS	Base Station
UE	User Equipment
FDD	Frequency Division Duplex
CS	Compressive Sensing
LASSO	Least Absolute Shrinkage and Selection Operator
AMP	Approximate Message Passing
TVAL	Total Variation Augmented Lagrangian
BM3D-AMP	Block-Matching and 3D-AMP
DL	Deep Learning
FCN	Fully-connected Network
ULA	Uniform Linear Array
OFDM	Orthogonal Frequency Division Multiplexing
DFT	Discrete Fourier Transform
MSE	Mean Squared Error
ADAM	Adaptive Moment Estimation
NMSE	Normalized Mean Square Error

QI LI received the B.S. degree from Zhongyuan University of Technology, Zhengzhou, China, in 2017. She is Master's degree currently studying at Zhongyuan University of Technology. Her research interests include wireless communication, massive MIMO, application of deep learning in wireless comunication.

AIHUA ZHANG received the B.S. and Ph.D. degrees from Zhengzhou University of china in 1998 and 2014, respectively. She received the master degree from China University of Mining & Technology-Beijing in 2003. From 2003 to 2014, she was a Lecturer. Since 2014, she has been an Associate Professor with Zhongyuan University of Technology. Now she is a visiting researcher with the Department of Electrical Engineering and Computer Science, University of Louisville, Kentucky, USA. Her current interests are in the areas of signal processing in communication, channel estimation, sparse signal processing, massive MIMO, NOMA.

PENGCHENG LIU received the B.Eng. degree in measurement and control and the M.Sc. degree in control theory and control engineering from Zhongyuan University of Technology, China, in 2007 and 2012, respectively, and the Ph.D. degree in robotics and control from Bournemouth University, UK, in 2017. He is currently a Lecturer at the Department of Computer Science, University of York, UK. Before joining York, he has held several academic positions at Cardiff, Lincoln, Bournemouth and China. He also held academic positions as a Visiting Fellow at Institute of Automation, Chinese Academy of Sciences, China and Shanghai Jiao Tong University, China. He is a member of IEEE, IEEE Robotics and Automation Society (RAS), IEEE Control Systems Society (CSS) and International Federation of Automatic Control (IFAC). He is also a member of the IEEE Technical Committee on Bio Robotics, Soft Robotics, Robot Learning, and Safety, Security and Rescue Robotics. He is an Associate Editor of IEEE Access and he received the Global Peer Review Awards from Web of Science in 2019, and the Outstanding Contribution Awards from Elsevier in 2017. He has published over 60 papers on flagship journals and conferences. He was nominated as a regular Funding/Grants reviewer for EPSRC, NIHR and NSFC and he has been leading and involving in several research projects and grants, including EPSRC, Newton Fund, Innovate UK, Horizon 2020, Erasmus Mundus, FP7-PEOPLE, NSFC, etc. He serves as reviewers for over 30 flagship journals and conferences in robotics, AI and control, e.g NEUNET, NODY, JINT, ICRA, IROS, RA-L, NCAA, etc. His research interests include robotics, machine learning, dynamical systems control and optimization.

JIANJUN LI received the B.S. and the M.S. degree from Xidian University, Xi'an, China, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China in 1996, 1999 and 2002, respectively. From 2002 2007, he was a senior researcher on 4G wireless communication in Beijing Samsung Telecom. R&D Center in Beijing China. From 2007 2015, he has worked in Posdata, Pantech and Innovative technology Lab. Co. as a senior researcher in Korea. During these period, his research mainly focus on Wimax and 3GPP LTE standard including MIMO, CoMP, small cell and LAA. Since 2016, he has been an Associate Professor at Zhongyuan University of Technology. His current research interest is on 5G wireless communication including massive MIMO, NOMA, new waveform and application of digital signal processing in communication systems.

CHUNLEI LI received his B.S. degree in computer science from Zhengzhou University, China, in 2001, M.S. degree from Hohai University, China, in 2004 and Ph.D. degree in computer science from Beihang University, China, in 2012. He is an associate professor at School of Electronics and Information in Zhongyuan University of Technology. In recent years, he published more than 50 technical papers and has authored two books. He has also obtained three patents. His research interests include machine vision for fabric defect detection, pattern recognition and low-rank representation.

• • •