Performance Enhancement of Massive MIMO Using Deep Learning-Based Channel Estimation

H M N Helmy.¹, S El Daysti.¹, H Shatila.¹, M Aboul-Dahab.¹

¹Electronics and Communication Engineering, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt.

*Corresponding author: hnabil110@gmail.com

Abstract. Massive Multiple-Input Multiple-Output (massive MIMO) system relies on channel state information (CSI) feedback to perform precoding and achieve performance gain in frequency division duplex (FDD) networks. However, transmission of massive MIMO system is subject to excessive feedback overhead. In this paper, we propose a Deep Learning (DL) approach-based channel estimation technique to enhance the performance of massive MIMO system. This technique is used to enhance recovery quality and improve trade-off between compression ratio (CR) and complexity of massive MIMO system. The proposed technique is based upon using the Channel State Information Network combined with gated recurrent unit (CsiNet-GRU). Moreover, the dropout method is used in the proposed technique to reduce overfitting during the learning process. The simulation results demonstrate that the proposed CsiNet-GRU technique results in a significant improvement in performance when compared with existing techniques used in conjunction with massive MIMO systems.

1. Introduction

The fifth generation (5G) wireless communications networks have a lot of novel requirements, such as the high system capacity with respect to the fourth generation (4G) networks, high frequency range (Covering through millimeter wave (mmWave) bands), increased data rate, ultra-low latency, reduced energy, and low cost [1]. Massive MIMO system was found very promising to be utilized in 5G systems to achieve its requirements. In addition, using massive MIMO systems brought new challenges on channel modeling [2]. It is quite notable that one of the performance bounds of 5G, like any other communication system, is determined by channel characteristics. Therefore, an accurate channel model plays an important role in designing, evaluating, and developing wireless communication systems. Massive MIMO channel state information (CSI) feedback techniques were devised in order to get more accurate and dynamic estimate of channel parameters, as compared with other channel models such as ITU-R IMT-2020, COST 2100 and the IEEE 802.11 ay models [3].

Artificial intelligence (AI) approaches were investigated in 5G systems to facilitate processing the signal received by massive MIMO antennas for the purpose of acquiring accurate channel estimation [4]. In this respect, a number of deep learning methods were proposed, with a variety of adopted algorithms. Compressive Sensing (CS) using the spatial and temporal correlation of CSI was used to obtain channel information with acceptable accuracy under substantially reduced feedback load [5]. Least Absolute Shrinkage and Selection Operator (LASSO) L1-solver [6] and Approximate Message-Passing (AMP) [7], were used CS to estimate channel parameters.
In [8], the authors utilized deep learning technology in massive MIMO to develop a CSI system, which was based upon using sensing and recovery network. The proposed network estimated the channel structure from training samples, and was called (CsiNet).

In addition, several algorithms using Auto-encoder CsiNet arrangement which utilized few neural network (NN) layers were proposed in order to reduce the channel state information (CSI) feedback and perform recovery with high degree of accuracy [9]. The authors in [10] adopted a simple and efficient approach to reduce the overhead of downlink channel estimation and feedback using linear regression (LR) and support vector regression (SVR) in machine learning. Channel estimation is very challenging when the receiver is equipped with a limited number of radio-frequency (RF) chains in beamspace millimeter-wave massive MIMO [11]. In this respect, an efficient online CSI prediction scheme, called OCEAN was proposed in [12], for predicting channel information from historical data in 5G wireless communication systems.

Convolutional Neural Network (CNN) is a deep learning technique which was proposed for channel estimation, combined with Long-Short Term Memory (LSTM) network using time varying method [13]. In [14] the authors proposed the CNN architecture invoking an (LSTM) module which admitted the neural network NN to benefit from exploiting temporal and frequency correlations of wireless channels. The concept of channel mapping in space and frequency, was adopted in [15] where the channels at one set of antennas and one frequency band are mapped to the channels at another set of antennas and frequency band using convolutional neural network CNN. Deep CNN was employed in [16] to perform wideband channel estimation for Millimeter Wave (mmWave) massive MIMO systems based on spatial correlation. The results showed that the performance of proposed technique was close to those of the ideal minimum mean-squared error (MMSE) estimator. In [17], the authors introduced an architecture of deep recurrent neural network (RNN) that was used for channel state information (CSI) by making use of compression and decompression capabilities. The architecture was based on splitting feature extraction into spatial and temporal domains, an approach which resulted in an improvement in channel state information (CSI) prediction.

Our contribution in this paper, we introduced a technique based on Deep Learning approach that can enhance the accuracy of channel estimation to improve the performance of massive MIMO utilized in 5G systems. The proposed technique is a modification of the DL-based channel state information (CSI) feedback network (CsiNet) as adopted in frequency division duplex (FDD) for massive (MIMO) system which was introduced in [9]. Gated Recurrent Unit (GRU) network is introduced to the system in order to increase the accuracy of CSI. In addition, the Dropout method is used to reduce the overfitting during the training processes. The performance of the proposed technique is investigated and compared to other similar techniques available in literature. The rest of the paper is organized as follows: - section 2. describes the utilized system model, section 3. describes the proposed channel state information gated recurrent unit (CsiNet-GRU) with the use of the Dropout method. Section 4. gives the simulation results and analysis and finally section 5. gives the conclusion of the paper.

2. Utilized System Model

The utilized system model is based on CSI network (CsiNet) configuration which was introduced in [9] with an image as input. A single-cell frequency division duplex (FDD) massive MIMO combining orthogonal frequency division multiplexing (MIMO-OFDM) structure is adopted. The structure contains \( N_t \gg 1 \) transmit antennas at the base station (BS), and a single receiver antenna at the user equipment (UE). The transmitter is assumed to employ precoding process, where independent and appropriate weightings are applied to each transmitted antenna signal such that the link throughput is maximized at the receiver output. The OFDM is assumed to have a number of \( N_c \) subcarriers. The received signal at the \( n^{th} \) subcarrier is expressed as:

\[
y_n = \bar{h}_n^H v_n x_n + z_n
\]
where $\tilde{H} = \left[ \tilde{h}_1, ..., \tilde{h}_N \right]$ $\in \mathbb{C}^{N \times N_t}$ be matrix of the estimated CSI then the UE should return $\tilde{H}$ to the BS through the feedback links. In these links, the total number of parameters are $N_t \cdot N_c$. This process is carried out continuously so that the time varying characteristics of the channel is estimated. Figure 1. illustrates the Pseudo-color plot of the strength of $H$ in case of $N_t \times N_c = 32 \times 32$ as an example of the CsiNet technique given in [9].

### Figure 1. Pseudo-color plot of the strength of $H \in \mathbb{C}^{32 \times 32}$ [9]

#### 3. The Proposed CsiNet-GRU Technique Using Dropout Method

In the FDD system, we shall take care of the feedback links included in both of the UE and BS. For this reason, we focus on the feedback scheme which allows autoencoder processing. We shall use GRU as a gating mechanism which is an advanced modeling of RNN, where processing is carried out either to keep or forget the information. The GRU operation (processing) is based upon utilizing two gates namely, the reset gate and update gate. A block diagram of the GRU is shown in figure 2. where the update gate acts similar to the forget and input gate of a LSTM [14]. It decides what information to throw away and what new information to add. The reset gate is used to decide how much past information to forget.

![Figure 2. the block diagram of (GRU) [14]](image)

We will extend CsiNet given in [9] by adding some GRU layers. The proposed technique is called the Channel State Information Network combined Gated Recurrent Unit (CsiNet-GRU) and is illustrated in figure 3. The GRUs are used to extend the CsiNet decoders for time correlation extraction and final reconstruction of the CSI. Each GRU has an inherent memory cells and can keep the previously extracted information for a long period for later prediction. To facilitate comparison with the results of the CsiNet structure given in [9], we shall consider a $3 \times 3$ convolution layers and an M-unit dense layers for sensing and a decoder with $2N_t \cdot N_c$ unit dense layers and two RefineNet for reconstruction. Each RefineNet contains four $3 \times 3$ convolution layers with different channel size as illustrated in figure 3. The outputs of the CsiNet decoders form sequences, each of length $T$ before being fed into
three-layer GRUs. Each GRU has $2\mathbf{N}_c \times \mathbf{N}_t$ hidden units, which have the same as the output dimension.

The final outputs are then reshaped into two $(\mathbf{N}_c' \times \mathbf{N}_t')$ matrices, as the final recovered $\hat{\mathbf{H}}_t$, which allow CR- CsiNet encoder to perform on the remaining $(T - 1)$ channel matrices to generate a series of $M_1 \times 1$ codewords ($M_1 > M_2$), given that less information is required due to channel correlation. The $(T - 1)$ codewords are all concatenated with the first ($M_1 \times 1$) codeword before being fed into the low-CR CsiNet decoder to fully utilize feedback information. Each CsiNet outputs two matrices with size $(\mathbf{N}_c' \times \mathbf{N}_t')$ as extracted features from the angular delay domain, as the final recovered $\hat{\mathbf{H}}_c$. The spatial frequency domain CSI can then be obtained via inverse 2D-DFT. At each time step, the GRUs implicitly learns time correlation from the previous inputs and then merges them with the current inputs to increase low CR recovery quality. the channel group in angular-interval domain is given by:

$$\{H'_t\}_{t=1}^{T} = \{H'_1, \ldots, H'_t, \ldots, H'_T\}$$

Figure 3. The structure of Proposed CsiNet-GRU using Dropout Technique

In order to better enhance the performance of the CsiNet-GRU structure, we adopt the Dropout regularization method that is utilized in neural networks to reduce overfitting. In this method, randomly selected neurons are ignored (dropped-out) during training. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass [18]. In the proposed structure, the input and recurrent connections to GRU units are probabilistically excluded from activation and weight updates while training a network. In fact, there are two dropout parameters in RNN layers, namely: dropout, applied to the first operation on the inputs, and recurrent dropout applied to the other operation on the recurrent inputs.

It is worth mentioning that we are interested in designing the encoder which can transform the channel matrix into an M-dimensional vector (codeword), where $M < N$. Thus, we define the data compression ratio $\gamma$ as $(\gamma = M/2N_cN_t)$. The encoder first extracts CSI features via a convolutional layer with two $3 \times 3$ filters, followed by an activation layer. Then, a fully connected (FC) layer with M
neurons is adopted to compress the CSI features to a lower dimension. The compression ratio (CR) of this encoder can be expressed as $\text{CR} = 1/\gamma$

The final reconstruction of the CSI is performed by three $2 \times N_c$ unit GRUs with Dropout techniques. Moreover, we adopt depth wise separable convolutions in feature recovery to reduce the size of the model and interact information between channels. We shall introduce the delay $\theta$ as a parameter used in the encoder and decoder i.e. $\theta = \{ \theta_{en}, \theta_{de} \}$. It is worth mentioning that $H_i$ are standardized with all components scaled into the $[0; 1]$. This standardization is required for CsiNet.

In order also to have fair comparison with the CsiNet given in [9], we shall adopt the same Adaptive Moment Estimation (ADAM) as the optimization algorithm and use mean squared error (MSE) as the loss function, which is defined as,

$$L(\theta) = \frac{1}{M} \sum_{m=1}^{M} \sum_{t=1}^{T} \| f(H_0; \theta) - H_t \|_2^2$$

where $M$ is the total number of samples in the training set and $\| \cdot \|_2$ is the Euclidean norm. The mechanism of the proposed CsiNet-GRU is described as follows. Multiple CR CsiNet encoders are deployed at UE, whereas the CsiNet decoders and GRUs are deployed at the BS. Each side has a counter. At the beginning, $H_1$ is compressed with high CR at the UE and recovered by a high-CR CsiNet decoder and initialized by the GRUs at the BS. In the subsequent time step $t$ ($2 \leq t \leq T$), $H_t$ is transformed into a lower-dimensional codeword $S_t$ at the UE, which is expected to contain the learned correlation information. The lower-dimensional codeword, $S_t$ is then concatenated with the first one $S_1$ and inversely transformed by the GRUs at the BS.

### 4. Simulation and Results Analysis

The COST 2100 model [3] is used to simulate MIMO channels and generate training samples. We set the MIMO-OFDM system to work on a 20 MHz bandwidth using Uniform Linear Array (ULA). The parameters utilized in both indoor and outdoor channel scenarios are given in table 1. Data sets are generated by randomly setting different start places with for indoor and outdoor scenarios. We perform the simulations at CR values with the first channel $H_1$ compressed under 1/4. The Training, validation, and testing sets given in table 1. Some parameters are preloaded from the CsiNet for initialization (Epochs from 500 to 1000, Learning rate is 0.001 and batch size is 200) as in illustrated in table 1.
Simulation is carried out through m-files implemented on a Python 3.7 platform. We compare the performance of the proposed architecture with the previous similar modeling approaches of CSI with different Deep Learning approaches, namely Conv-LSTM CsiNet [17], LASSO [6], TVAL3 [19] and CsiNet [9]. We utilize the default setup in the open source codes of the previously mentioned techniques for reproduction. Note that in the CsiNet, we have taken into consideration the slight distinction between datasets. LASSO uses simple sparsity priors but achieves good performance. TVAL3 is a minimum total variation method that provides remarkable recovery quality but with high computing efficiency. Convolutional-LSTM CsiNet adopting RNN and Depth wise separable convolution in feature extraction and recovery modules, respectively. Distinction between datasets in addition to refining the CsiNet parameters on our training set for several epochs for fairness.

We run the modeling CsiNet-GRU on Collaboratory (python) according to Zero configuration required, Free access to GPUs and Easy sharing are used for training and testing of the CsiNet, Conv-LSTM CsiNet and CsiNet-GRU on python colab editor. Comparisons are investigated based on the Normalized Mean Square Error (NMSE), the cosine similarity, accuracy and the best runtime in indoor and outdoor channel, as well as the complexity incorporated. The Normalized Mean Square Error (NMSE) utilized for comparisons quantifies the difference between the input \( \{H_t\}_{t=1}^T \) and the output \( \{\hat{H}_t\}_{t=1}^T \) and is given in [9] by:

\[
NMSE = \mathbb{E}
\left[
\frac{1}{T}
\sum_{t=1}^{T}
\left[
\frac{\|H_t - \hat{H}_t\|_2^2}{\|H_t\|_2^2}
\right]
\right]
\]  
(4)

To measure the degree of similarity between the actual channel \( h_{n,t} \) and the estimated channel value \( \hat{h}_{n,t} \) of the \( n^{th} \) subcarrier at time \( t \), we used the cosine similarity coefficient \( \rho \), which is given in [9] as:

\[
\rho = \mathbb{E}
\left[
\frac{1}{TN_c}
\sum_{t=1}^{T}
\sum_{n=1}^{N_c}
\left[
\frac{|\hat{h}_{n,t}^H h_{n,t}|}{\|\hat{h}_{n,t}\|_2 \|h_{n,t}\|_2}
\right]
\right]
\]  
(5)

| Table 1. COST 2100 Model Data – Sets and System, Parameters |
|-------------------------------------------------------------|
| **MIMO OFDM** | 20 MHz bandwidth |
| **H** | \( 32 \times 32 \) |
| **N_t** | 32 Antennas |
| **N_c** | 1024 Subcarriers |
| Training Samples | 100,000 |
| Validation Samples | 30,000 |
| Testing Samples | 20,000 |
| Epochs | 500 - 1000 |
| Learning Rate | 0.001 |
| Batch size | 200 |
| \( \Delta t \) | 0.04 s |
| \( T \) | 10 s |
| **CR** | 4, 16, 32, 64 |
| **Channel Model** | **Indoor Scenario** | **Outdoor Scenario** |
| Speed | Pico, 5.3 GHz | Rural, 300 MHz |
| \( v \) | 0.0036 km/h | 4.24 km/h |
| \( \Delta t \) | 30 s | 0.56 s |
We introduce a new parameter for comparison which we shall call the accuracy. We define it as the ratio of number of correct predictions to the total number of input samples, that means accuracy is the ratio of recovered channel vector to the original channel vector \( H_t^T / H_1 \) so the accuracy is defined as:

\[
\text{Accuracy} = \frac{1}{T N_r} \sum_{t=1}^{T} \sum_{n=1}^{N_r} \frac{\| \hat{H}_t^N \|_2}{\| H_r^N \|_2}
\]

Figures (4,5) illustrate the relation between CR and NMSE in cases of indoor and outdoor scenarios respectively for all structures. It is clear from figure 4 that the proposed CsiNet-GRU has the lowest NMSE, whereas in figure 5, it has the lowest NMSE among others except that of Conv-LSTM CsiNet at values of CR>20. Figures (6,7) illustrate the relation between the CR and accuracy in cases of indoor and outdoor scenarios respectively for all structures. It is quite clear that the CsiNet-GRU outperforms the other structures, with higher values of accuracy observed at lower values of CR. Figures (8,9) illustrate the relation between the cosine similarity \( \rho \) and CR in indoor and outdoor scenarios respectively for all structures. Again, the proposed CsiNet-GRU outperforms the other structures, and in addition, it exhibits a near linear performance with a lowest slope.

The performance comparison between the proposed CsiNet-GRU and other available techniques is given in table 2, where corresponding values of NMSE, \( \rho \), and accuracy are calculated for different values of \( \gamma \) for indoor and outdoor scenarios. It is clear that all techniques have better performances in the indoor scenario as compared to the outdoor one. From the table 2, it is notable that the CsiNet techniques considerably outperform the other CS based methods (LASSO and TVAL3). However, the proposed CsiNet-GRU outperforms the channel state information network (CsiNet). As far as the NMSE is concerned, the CsiNet-GRU achieves the lowest values at all compressed ratios (CRs), especially when CR is low. Notably, CsiNet-GRU is characterized by very short run time periods as compared to LASSO and TVAL3 techniques. However, compared with the other CsiNet technique, the proposed CsiNet-GRU slightly loses time efficiency. This is due to the complexity that result from adding the GRU units. It is quite notable that although there is a considerable added complexity due to the introduction of the GRU layers, the run time is still comparable to that of the CsiNet. This is due to utilizing the dropout method, which has compensated for the expected excessive run time.

Figure 10 shows the reconstruction results of the proposed technique as compared to the other modeling techniques, namely LASSO [6], TVAL3 [19], CsiNet [9], Conv-LSTM CsiNet [17] in indoor Picocellular scenario as an example. The figure represents the average performance at different CRs which is reflected on the reconstructed images in case of utilizing the different techniques. Clearly, CsiNet [9], Conv-LSTM CsiNet [17] and CsiNet-GRU continue to offer acceptable correlation coefficients \( \rho \) at low CRs, where CS based methods fail to work. It is quite remarkable that the proposed CsiNet-GRU technique outperforms the others.
Figure 4. NMSE (dB) performance comparison between CS methods INTDOOR scenario

Figure 5. NMSE (dB) performance comparison between CS methods OUTDOOR scenario

Figure 6. Accuracy performance comparison between CS methods INTDOOR scenario

Figure 7. Accuracy performance comparison between CS methods OUTDOOR scenario

Figure 8. ρ performance comparison between CS methods INDOOR scenario

Figure 9. ρ performance comparison between CS methods OUTDOOR scenario
Table 2. Performance Comparison between the Proposed Structure and Other Available Ones

|       | γ    | LASSO [6] | TVAL3 [19] | CsiNet [9] | Conv-LSTM CsiNet [17] | CsiNet-GRU |
|-------|------|-----------|------------|------------|------------------------|------------|
|       | 1/4  | -7.59     | -8.87      | -17.36     | -27.5                  | -51.73     |
|       | 1/16 | -2.96     | -3.2       | -8.65      | -23                    | -27.3      |
|       | 1/32 | -1.18     | -0.46      | -6.24      | -22.3                  | -23.81     |
|       | 1/64 | -0.18     | -0.6       | -5.84      | -21.2                  | -22.11     |
|       | 1/4  | 0.91      | 0.87       | 0.98       | 0.958                  | 0.99       |
|       | 1/16 | 0.72      | 0.73       | 0.9        | 0.957                  | 0.98       |
|       | 1/32 | 0.53      | 0.45       | 0.93       | 0.955                  | 0.97       |
|       | 1/64 | 0.3       | 0.24       | 0.93       | 0.942                  | 0.93       |
| Indoor| NMSE | 1/4       | 0.68       | 0.34       | 0.81                   | 0.82       |
|       |      | 1/16      | 0.55       | 0.22       | 0.63                   | 0.64       |
|       |      | 1/32      | 0.34       | 0.15       | 0.51                   | 0.55       |
|       |      | 1/64      | 0.55       | 0.23       | 0.48                   | 0.56       |
|       | ρ    | 1/4       | 0.345      | 0.314      | 0.0001                 | 0.0001     |
|       |      | 1/16      | 0.555      | 0.314      | 0.0001                 | 0.0001     |
|       |      | 1/32      | 0.604      | 0.286      | 0.0001                 | 0.0001     |
|       |      | 1/64      | 0.708      | 0.285      | 0.0001                 | 0.0001     |
| Outdoor| NMSE| 1/4       | -5.08      | -0.9       | -8.75                  | -10.9      |
|       |      | 1/16      | -1.09      | -0.53      | -4.51                  | -9.86      |
|       |      | 1/32      | -0.27      | 0.42       | -2.81                  | -9.18      |
|       |      | 1/64      | -0.06      | 0.74       | -1.93                  | -8.83      |
|       | ρ    | 1/4       | 0.82       | 0.58       | 0.91                   | 0.93       |
|       |      | 1/16      | 0.49       | 0.46       | 0.79                   | 0.94       |
|       |      | 1/32      | 0.32       | 0.28       | 0.67                   | 0.92       |
|       |      | 1/64      | 0.19       | 0.19       | 0.59                   | 0.88       |
|       | Accuracy | 1/4    | 0.66       | 0.54       | 0.68                   | 0.72       |
|       |      | 1/16      | 0.45       | 0.22       | 0.5                    | 0.53       |
|       |      | 1/32      | 0.2        | 0.2        | 0.36                   | 0.39       |
|       |      | 1/64      | 0.18       | 0.15       | 0.25                   | 0.28       |
|       | Run | 1/4       | 0.2122     | 0.15       | 0.0001                 | 0.0001     |
| time |      | 1/16      | 0.2409     | 0.3145     | 0.0001                 | 0.0001     |
|       |      | 1/32      | 0.598      | 0.2985     | 0.0001                 | 0.0001     |
|       |      | 1/64      | 0.6758     | 0.285      | 0.0001                 | 0.0001     |
5. Conclusions

In this paper, we proposed a channel state information (CSI) feedback network by extending the DL-based CsiNet technique to incorporate GRUs and making use of the Dropout method. The GRU layers were used to extend the CsiNet decoders for time correlation extraction and final reconstruction of CSI, whereas the Dropout method was used to reduce the overfitting of the channel modeling. The proposed CsiNet-GRU technique achieved the lowest NMSE, the best cosine similarity coefficient and the best accuracy as compared to other CS-based and CSI-based techniques. However, the introduction of the GRUs layers had increased the complexity, with the subsequent expected increase in the run time. Thus, the proposed technique allows for the tradeoff between accuracy and run time parameters in designing massive MIMO utilized in conjunction with FDD networks.
References

[1] Liu, G. and Jiang, D. "5G: Vision and requirements for mobile communication system towards year 2020," Chinese Journal of Engineering, 2016, p. 8, 2016.

[2] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for FDD multiuser massive MIMO systems," IEEE Trans. Signal Process. , vol. 62, no. 12, p. 3261–3271, Jun. 2014.

[3] L. Liu, C. Oestges, J. Poutanen, and K. Haneda, "The COST 2100 MIMO channel model," IEEE Wireless Commun. , vol. 19, no. 6, p. 92–99, Dec. 2012.

[4] R. Li et al., "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," IEEE Wireless Communications, vol. 24, no. 5, pp. 175-183, October 2017.

[5] P. H. Kuo, H. Kung, and P. A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," Proc. IEEE WCNC, p. 492–497, Apr. 2012.

[6] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Comm. Pure and Applied Math, vol. 75, p. 1412–1457, 2004.

[7] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," Proc. Natl. Acad. Sci., vol. 106, no. 45, pp. 18914 - 18919, 2009.

[8] P. Dong, H. Zhang, and G. Y. Li, "Machine learning prediction-based CSI acquisition for FDD massive MIMO downlink," in Conf. " in (Globecom’18), Proc. IEEE Global Commun., Abu Dhabi, UAE, Dec. 2018.

[9] C. Wen, W. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," IEEE Wireless Communications Letters, vol. 7, p. 748–751, Oct 2018.

[10] 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, 2018.

[11] 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.

[12] Luo, Changqing, Jinlong Ji, Qianlong Wang, Xuhui Chen, and Pan Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," IEEE Transactions on Network Science and Engineering, vol. 7, no. 1, pp. 227 - 236, Jan 2020.

[13] Wang, Tianqi, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," IEEE Wireless Communications Letters, vol. 8, no. 2, pp. 416-419, 2018.

[14] Lu, Chao, Wei Xu, Hong Shen, Jun Zhu, and Kezhi Wang., "MIMO channel information feedback using deep recurrent network," IEEE Communications Letters , vol. 23, no. 1, pp. 188-191, 2018.

[15] Alrabeiah, M. and Alkhateeb, A., "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in 2019 53rd Asilomar Conference on Signals, Systems, and Computers, November, 2019.

[16] P. Dong, H. Zhang, G. Y. Li, N. Naderi Alizadeh and I. S. Gaspar, "Deep CNN for Wideband MmWave Massive Mimo Channel Estimation Using Frequency Correlation," ICASSP 2019 -, in 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019.

[17] Li, Xiangyi, and Huaming Wu., "Spatio-temporal representation with deep neural recurrent network in MIMO CSI feedback," IEEE Wireless Communications Letters, vol. 9, no. 5, pp. 653-657, 2020.
[18] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[19] Li, Chengbo, Wotao Yin, and Yin Zhang, "User’s guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 4, pp. 46-47, 2009.