Deep CSI Feedback for FDD MIMO Systems

Zibo He(✉), Long Zhao, Xiangchen Luo, and Binyao Cheng

The Key Lab of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT), Beijing, China
zibohe@bupt.edu.cn

Abstract. With the increasing number of antennas at the base station (BS), the feedback overhead of traditional codebook in frequency division duplexing (FDD) mode becomes overwhelming, since the number of codewords in codebook increases quickly. Alternatively, we can directly feedback the channel state information (CSI) to the BS for precoding. To reduce the overhead of CSI feedback, this paper proposes three CSI compression models based on autoencoder network. The first two of them, adopting deep learning (DL) structure, are named FCNet and CNet, respectively. FCNet employs full-connected network architecture, while CNet is designed based on convolutional neural network with lightweight convolution kernels and multi-channel architecture. By applying principal component analysis (PCA) on CSI feedback, the third one, i.e., PCANet, is also studied and analyzed in details. Experiments show that CNet has best accuracy performance at the cost of high computational complexity, while FCNet shows medium accuracy and complexity among the three models. Besides, the accuracy of PCANet is nearly the same as CNet in some specific channel conditions. Compared with the state-of-the-art of CsiNet, the proposed models have their own advantages and limitations in different scenarios.

Keywords: MIMO · CSI feedback · Deep learning · PCA

1 Introduction

In frequency division duplexing (FDD) multiple-input multiple-output (MIMO) systems, the user equipment (UE) selects suitable codeword in the codebook and transmits its index to the base station (BS) with few bits since the number of codewords is small [1,2]. However, with the increasing number of antennas at the BS, the number of codebooks increases and becomes overwhelming for MIMO systems [3,4]. Alternatively, we can directly feedback the downlink channel state information (CSI) to the BS through the feedback link for precoding [5].

In recent years, deep learning (DL) is widely used in image compression and has achieved great success, such as convolutional neural network (CNN) [6–8].

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 366–376, 2022.
https://doi.org/10.1007/978-3-030-99200-2_28

By taking into account that CSI compression is similar to image compression, DL has been applied for CSI feedback. Based on CNN, CsiNet has been proposed and shows remarkable performance [5]. With larger size of convolution kernels, CsiNetPlus is further introduced and shows better performance than CsiNet [9]. By leveraging multi-resolution architecture, another network called CRNet is also studied to improve the performance of CsiNet [10]. Combined with superimposed coding (SC), the CSI feedback performance of DL can also be improved [11].

Although the nonlinear representation ability of DL has been extensively studied, the feedback accuracy needs to be further improved for practical systems. On the other hand, it is true that DL for CSI feedback has outstanding performance, however the linear compression methods have advantages in reducing computational complexity, which is critical for UE with limited computing resources.

This paper adopts the antoencoder architecture for CSI feedback, where the encoder compresses the downlink CSI into bit information and feedbacks it to the BS; then the bit information is recovered to the CSI by the decoder for precoding. To achieve better feedback accuracy or low complexity, three different models, i.e., CNet, FCNet and PCANet, are proposed for the design of both encoder and decoder. By optimizing the number of neurons and network layers as well as designing a new activation function, FCNet based on full-connected network (FCN) can achieve low computational complexity at the cost of little accuracy degradation. By leveraging lightweight convolution kernels and multi-channel architecture, CNet based on CNN can achieve better accuracy than CsiNet. One linear compression scheme, i.e., principal component analysis (PCA), is also studied for designing the encoder and decoder, denoted as PCANet, which shows medium accuracy with extremely low complexity. The number of quantization bits are further analyzed for different schemes, by which we can determine the optimal number of quantization bits for CSI feedback.

The rest of this paper is organized as follows. Section 2 describes the system model and the overall structure of CSI feedback. The design of FCNet, CNet and PCANet as well as the quantization scheme are detailedly studied in Sect. 3. The numerical results and analysis are given in Sect. 4, while the conclusion is drawn in Sect. 5.

2 System Model

For simplification, we consider a narrow band FDD MIMO system, where the BS and UE are equipped with N_t and N_r antennas, respectively. Assuming that the transmitted pilot and data symbols are $\mathbf{x}_\mathrm{p} \in \mathbb{C}^{N_\mathrm{t}\times 1}$ and $x_\mathrm{d} \in \mathbb{C}$, respectively, then the received pilot signal $\mathbf{y}_\mathrm{p}$ and data signal $\mathbf{y}_\mathrm{d}$ can be respectively expressed as:

$$\begin{aligned} \mathbf{y}_\mathrm{p} &= \mathbf{H}\mathbf{x}_\mathrm{p} + \mathbf{z}_\mathrm{p} \\ \mathbf{y}_\mathrm{d} &= \mathbf{H}\mathbf{w}x_\mathrm{d} + \mathbf{z}_\mathrm{d} \end{aligned} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_\mathrm{r}\times N_\mathrm{t}}$, $\mathbf{w} \in \mathbb{C}^{N_\mathrm{t}\times 1}$ and $\mathbf{z}_\mathrm{p}, \mathbf{z}_\mathrm{d} \in \mathbb{C}^{N_\mathrm{r}\times 1}$ denote the channel matrix, precoding vector and additive white Gaussian noise (AWGN), respectively.

As shown in Fig. 1, when the UE received the pilot signal $\mathbf{y}_{\mathrm{p}}$, it can estimate channel $\mathbf{H}$, then compresses and feedbacks it to the BS via the uplink, so that the BS can design the precoder $\mathbf{w}$ for downlink transmission. Downlink channel estimation [12] is beyond the scope of this paper, so we assume that the perfect channel can be acquired by UE and only focus on the feedback scheme.

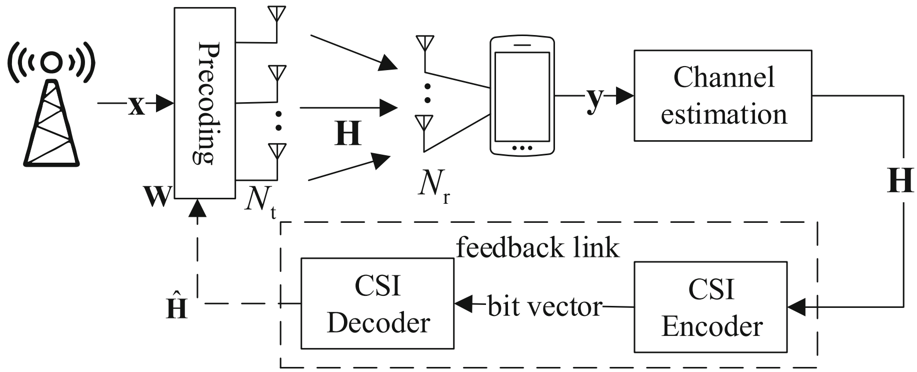

Fig. 1. Channel feedback model

In order to achieve better performance of CSI feedback, we borrow the autoencoder network, which generally consists of an encoder and a decoder. The objective of the encoder is to compress the channel matrix $\mathbf{H}$ to a (N/B)-dimensional compressed vector $\mathbf{s}$, i.e.,

$$\mathbf{s} = \left[s_1, s_2, \cdots, s_{N/B}\right]^{\mathrm{T}} = \text{Encoder}\left(\mathbf{H}\right) \tag{2}$$

where N and B denote the numbers of total feedback bits and quantization bits per element in $\mathbf{s}$, respectively. Then, $\mathbf{s}$ can be quantized into a bit vector $\mathbf{c}$ and transmitted to the BS:

$$\mathbf{c} = \left[c_1, c_2, \cdots, c_N\right]^{\mathrm{T}} = \text{Quantization}\left(\mathbf{s}\right) \tag{3}$$

Once the BS received $\mathbf{c}$, it is dequantized into the estimated $\mathbf{s}$, i.e.,

$$\hat{\mathbf{s}} = \left[\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_{N/B}\right]^{\mathrm{T}} = \text{Dequantization}\left(\mathbf{c}\right) \tag{4}$$

which is further recovered to the reconstructed channel matrix $\hat{\mathbf{H}}$ by the decoder:

$$\hat{\mathbf{H}} = \text{Decoder}\left(\hat{\mathbf{s}}\right) \tag{5}$$

The objective of the entire network is to minimize the normalized distance between the original $\mathbf{H}$ and the reconstructed $\hat{\mathbf{H}}$, i.e., normalized mean square error (NMSE), given by:

$$\text{NMSE} = \mathrm{E}\left\{\frac{\left\|\mathbf{H} - \hat{\mathbf{H}}\right\|_2^2}{\left\|\mathbf{H}\right\|_2^2}\right\} \tag{6}$$

3 Design of Deep CSI Feedback

By leveraging different methods or DL network structures, three schemes for CSI feedback are proposed in this section and detailedly given as follows.

3.1 FCNet

FCN is first considered for CSI feedback due to its low computational complexity. Based on the FCN, Fig. 2 illustrates the proposed FCNet architecture, the encoder and decoder of which consist of a series of full-connected layers.

The encoder consists of one input layer, three hidden layers and one quantization layer, which have L, P and N/B neurons, respectively. Firstly, we reshape the original CSI matrix $\mathbf{H}$ into a vector $\mathbf{h} \in \mathbb{C}^{M\times 1}$, where $M = 2N_\mathrm{t}N_\mathrm{r}$ and 2 represents the real and imaginary parts of the CSI matrix. The CSI vector $\mathbf{h}$ serves as the input of encoder and the output is a bit vector $\mathbf{c}$. Once the BS received $\mathbf{c}$ from UE through feedback link, it utilizes the decoder to reconstruct the CSI matrix $\hat{\mathbf{H}}$, where the decoder contains three hidden layers with P neurons and one output layer with M neurons.

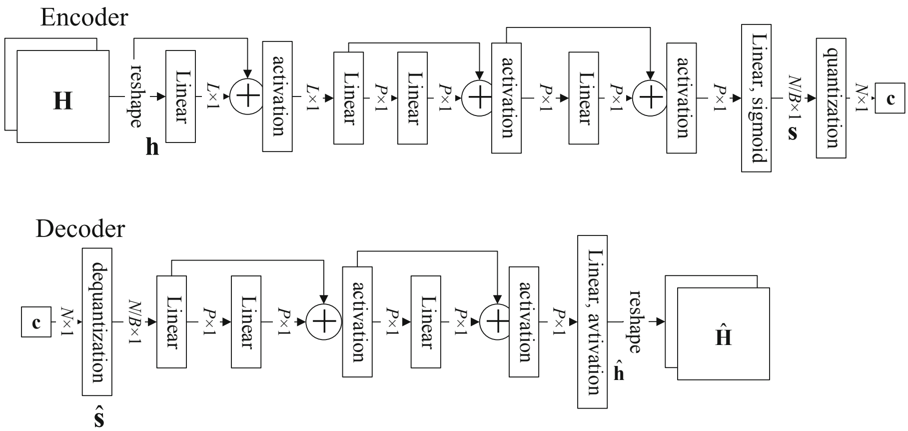

Fig. 2. Architecture of FCNet for CSI feedback

In order to further extract the channel feature and avoid gradients vanishing or exploding caused by deep FCNet, we adopts the shortcut connection [13] in FCNet. What's more, we create a new avtivation function in hidden layers to improve the nonlinear representation ability of the model, and it can be written as:

$$\mathrm{activation}\,(x) = x \cdot \tanh\,(\mathrm{softplus}\,(x)) \tag{7}$$

Moreover, uniform quantization is developed as the quantizer of FCNet. Uniform quantization is a rounding operation, where each value is rounded to the nearest value in a finite set of predefined quantization levels [14], which can be written as:

$$\hat{s}_k = \frac{\text{round}\left(s_k \cdot 2^B - 0.5\right) + 0.5}{2^B}, k \in [1, 2, \cdots, N/B] \tag{8}$$

where B represents the number of quantization bits and the function round() can convert its input to its nearest integer value[1].

3.2 CNet

Although FCNet can achieve low computational complexity, it shows low feedback accuracy. In order to improve the accuracy of CSI feedback, CNet is designed based on CNN, the detailed architecture of which is shown in Fig. 3.

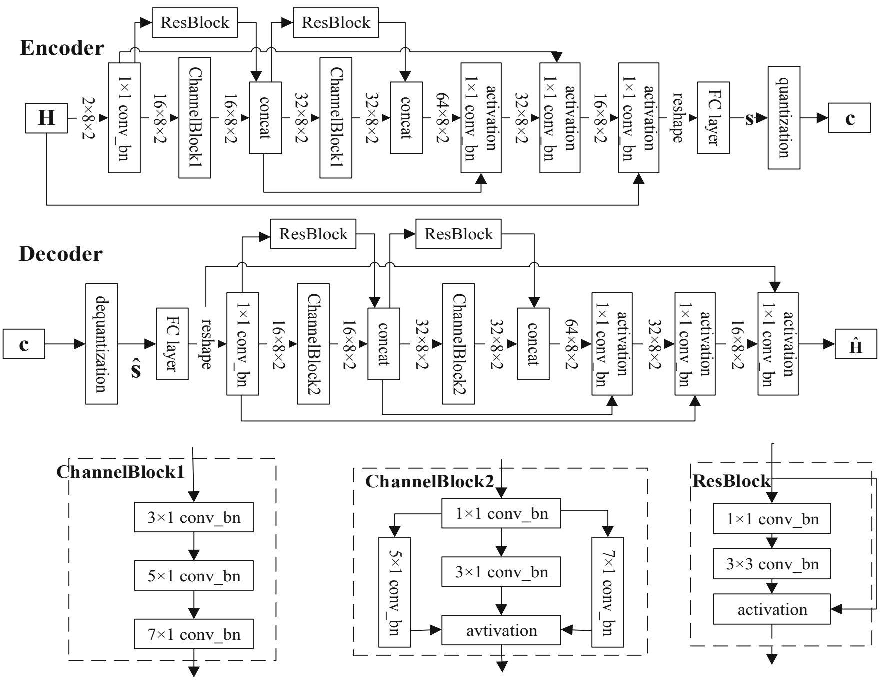

Fig. 3. Architecture of CNet for CSI feedback

Compared with CsiNet relying on CNN, we introduce lightweight convolution kernels into our CNet. It can efficiently extract the channel features for the dense CSI matrix with low computational complexity. On the other hand, we

---

[1] Since each value in **s** belongs to (0, 1) (the sigmoid function), we minus 0.5 before round function, quantize the rounded value to bits and then transmit them to the decoder. Once the decoder received these bits, it dequantizes them, plus 0.5 and then divide them by 2^B to get the quantization value.

apply multi-channel architecture on CNet, i.e., ChannelBlock1, ChannelBlock2 and ResBlock in Fig. 3. The multi-channel architecture convolutes the channel matrix along different paths to extract the channel features, which are further concatenated into a three-dimension channel feature matrix.

Similar to FCNet, the shortcut connection and the proposed activation function (7) are also adopted in CNet. The quantization process in CNet is also the same to that in FCNet.

3.3 PCANet

PCA is a linear method for dimension reduction, and it can map the original data onto a new feature space, the orthogonal basis of which is called principal component.

In the proposed PCANet, the encoder first calculates the covariance matrix $\mathbf{D}$ of the original CSI vector $\mathbf{h}$, which is reshaped from $\mathbf{H}$. Then, the eigenvalues and eigenvectors of $\mathbf{D}$ are calculated and sorted. The first N/B largest eigenvectors are selected to constitute the new orthogonal basis $\mathbf{P}$. Finally the CSI vector $\mathbf{h}$ is mapped on the new orthogonal basis $\mathbf{P}$ to generate the weight vector $\mathbf{s}$. After quantization of $\mathbf{s}$, the obtained bit vector $\mathbf{c}$ will be transmitted to the decoder.

Once the decoder received $\mathbf{c}$, it can be dequantized and then mapped into the weight vector $\hat{\mathbf{s}}$. Further, the decoder projects $\hat{\mathbf{s}}$ on the transposed orthogonal basis, i.e., $\mathbf{P}^{\mathrm{T}}$, to acquire the reconstructed CSI matrix $\hat{\mathbf{H}}$.

The detailed PCANet for CSI feedback is given in Algothrim 1.

Algorithm 1 : PCANet algorithm

Encoder

1: Get the original CSI matrix $\mathbf{H}$ and convert it to a vector $\mathbf{h}$;
2: Calculate the covariance matrix $\mathbf{D} = \mathbf{h}\mathbf{h}^{\mathrm{T}}$;
3: Compute the eigenvalues $\boldsymbol{\lambda}$ and eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_M\}$ of $\mathbf{D}$;
4: Sort $\boldsymbol{\lambda}$ in descending order and choose corresponding eigenvectors to the first N/B largest eigenvalues as the new orthogonal basis $\mathbf{P}$;
5: Project $\mathbf{h}$ on the new orthogonal basis $\mathbf{P}$ and obtain the mapping vector $\mathbf{s} = \mathbf{h}\mathbf{P} = \left[s_1, s_2, \cdots, s_{N/B}\right]$;
6: Each value in $\mathbf{s}$ is quantized into B bits and form $\mathbf{C} = \left[\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{N/B}\right]^{\mathrm{T}}$, where $\mathbf{c}_i$ denotes the quantization bit vector of s_i $(i = 1, 2, \cdots, N/B)$;
7: Reshape $\mathbf{C}$ into the bit vector $\mathbf{c}$ and is transmitted to the decoder;

Decoder

8: Dequantize $\mathbf{c}$ and get the estimated weight vector $\hat{\mathbf{s}} = [\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_{N/B}]$;
9: Calculate the reconstructed CSI vector $\hat{\mathbf{h}} = \hat{\mathbf{s}}\mathbf{P}^{\mathrm{T}}$ and is then converted to the CSI matrix $\hat{\mathbf{H}}$.

Since each value in the weight vector $\mathbf{s}$ does not belong to (0, 1), quantization formula in (8) cannot be directly applied on PCANet. In order to map $\mathbf{s}$ into $(-1, 1)$, we divide each element of $\mathbf{s}$ by $\max\{s_i, i = 1, \cdots, N/B\}$. After quantization and dequantization, these values will be multiplied by $\max\{s_i, i = 1, \cdots, N/B\}$ in order to obtain the original values.

4 Simulation Results and Analysis

4.1 Experiment Setup

We consider two types of typical channels: CDL-C and CDL-D. The carrier frequency is 2.6 GHz and the delay spreads of CDL-C and CDL-D channels are 300 ns and 100 ns, respectively. The number of antennas at the UE and BS are $N_{\mathrm{r}} = 2$ and $N_{\mathrm{t}} = 8$, respectively. We generate 70000 independent channel samples, which is divided into 50000 training samples, 10000 validation samples and 10000test samples.

The whole simulation is implemented in Pytorch. Both FCNet and CNet adopt random initialization and Adam optimizer. To improve the feedback accuracy of FCNet and CNet, the NMSE is directly adopted as the loss function instead of mean square error (MSE). The initial learning rate is 0.001 and the batch size is 32. Moreover, FCNet and CNet are trained for 50 and 100 epochs, respectively.

4.2 Simulation Results and Analysis

The NMSEs of both FCNet and CNet versus the number of epochs are shown in Fig. 4a and Fig. 4b. From the two figures, we can know that the NMSEs on both the training set and test set gradually decrease and finally become convergent with the increasing of training epochs. Figure 4c shows that the largest eigenvalue of the covariance matrix gradually decreases and finally becomes convergent with the increasing of training samples, which shows the feasibility of PCANet.

Figure 5 shows the NMSEs of different models versus the number of total feedback bits under $B = 2$. It can be seen that the NMSE of each model decreases with the increasing number of total feedback bits, which means the accuracy of CSI feedback could be improved at the cost of high feedback overhead.

The number of quantization bits B determines the accuracy of uniform quantization and the number of elements in $\mathbf{s}$, i.e., N/B, and further affects the feedback accuracy of the proposed models, especially for PCANet. In order to find the optimal number of quantization bits for different models, Fig. 6 illustrates the NMSE versus the number of quantization bits under $N = 12$. It can be seen from Fig. 6 that the NMSE of each model first decreases and then increases with the increasing number of quantization bits, therefore there should exist an optimal number of quantization bits B for each model. Moreover, different models have different optimal numbers B; even for the same model, the optimal numbers B may be different under different channels. Taking CDL-D channel as an example, the optimal numbers of quantization bits for FCNet, CNet and PCANet are 4, 6 and 2, respectively. As for FCNet, the optimal numbers under CDL-C and CDL-D channels are 6 and 4, respectively.

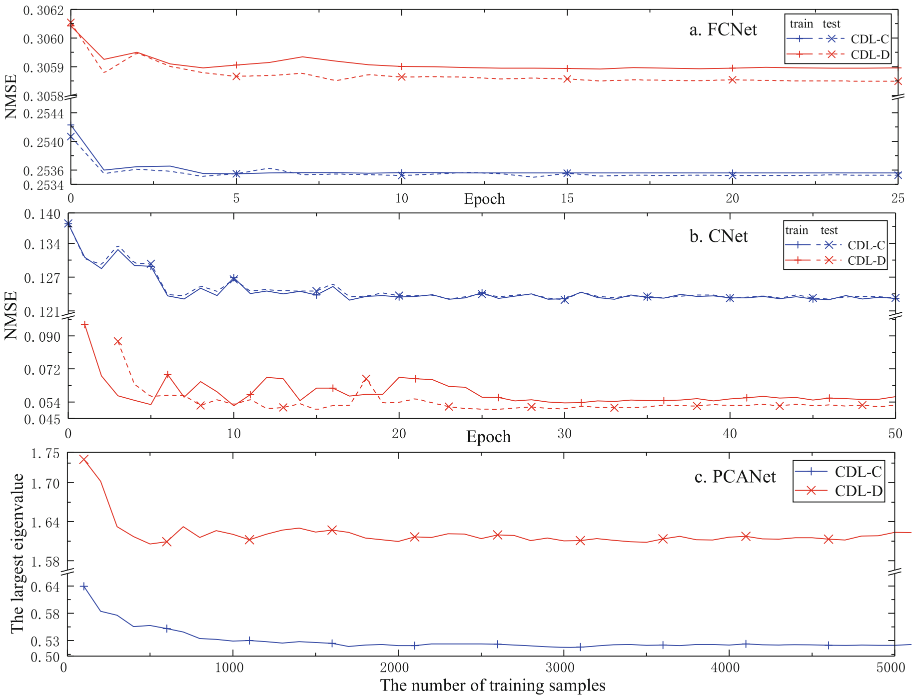

Fig. 4. Convergence of the proposed models

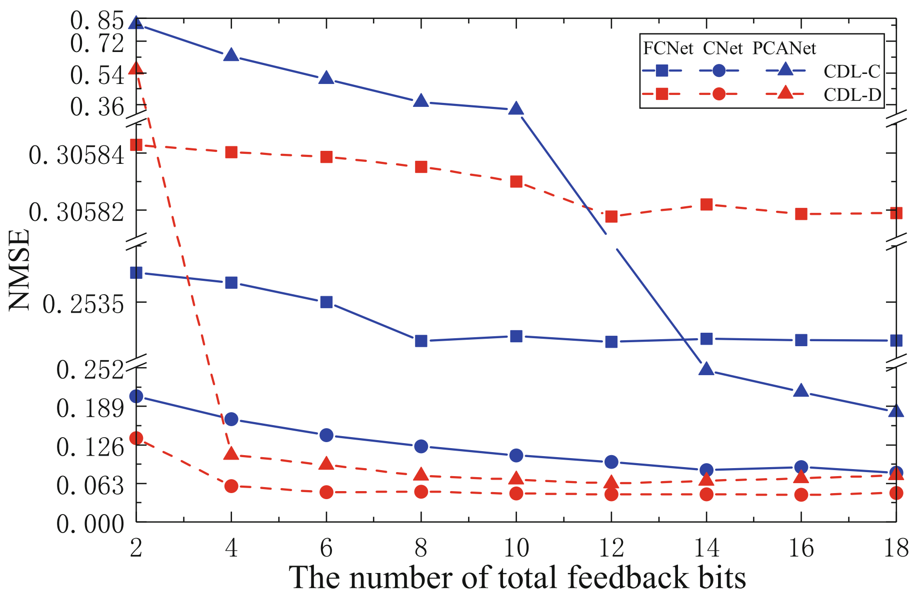

Fig. 5. NMSE versus the number of feedback bits ($B = 2$)

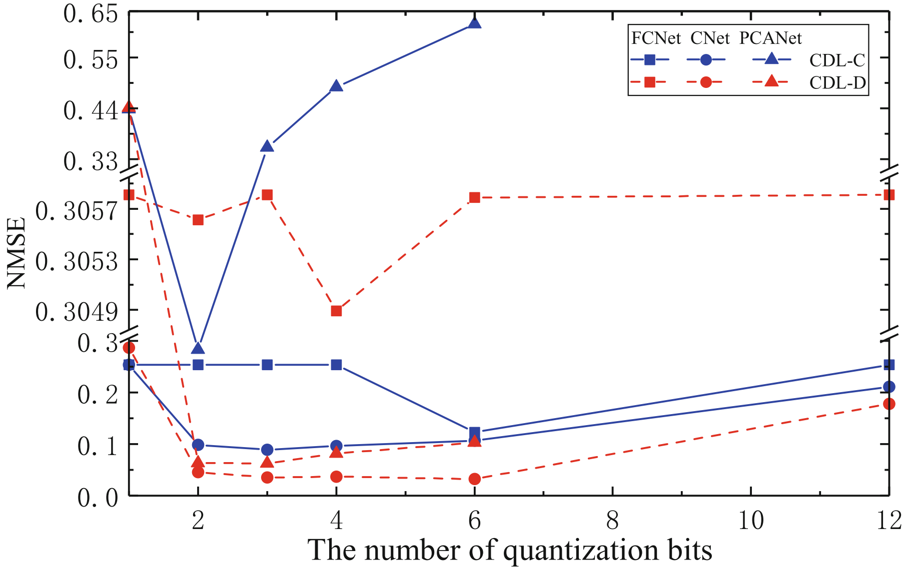

Fig. 6. NMSE versus the number of quantization bits ($N = 12$)

In order to evaluate the computational complexities of different models, the total number of multiplication and addition operations, i.e., floating point operation (FPO), is calculated as the evaluated metric. Table 1 shows both the NMSE and FPO of the proposed models and CsiNet under $N = 12$ and $B = 2$. It can be seen that CNet performs much better in feedback accuracy compared with the other three models under CDL-C channel. As for CDL-D channel, the accuracies of CNet and PCANet are nearly the same and are much better than FCNet and CsiNet. In terms of FPO metric, CNet has the highest complexity but can obtain the best accuracy; in contrast, PCANet achieves much lower complexity than the other three models.

Table 1. NMSE and FPO between the proposed models and CsiNet ($N = 12$, $B = 2$)

Model	NMSE		FPO
	CDL-C	CDL-D	
FCNet	0.2535	0.3058	352010
CNet	0.0983	0.0454	2250970
PCANet	0.2830	0.0630	984
CsiNet	0.2530	0.2633	231450

5 Conclusion

In order to efficiently feedback the CSI from the UE to BS, this paper proposes three models, named FCNet, CNet and PCANet, respectively, based on the autoencoder network in FDD MIMO systems. By optimizing the numbers of neurons and network layers as well as designing a new activation function, we first propose FCNet based on FCN architecture. By leveraging lightweight convolution kernels and multi-channel architecture for extracting channel feature, CNet is designed to further improve the accuracy of CSI feedback. Moreover, based on the linear method of PCA for dimension reduction, PCANet is studied and analyzed for CSI feedback with low complexity. Simulation results indicate that CNet performs best in the feedback accuracy at the cost of high complexity, while FCNet shows medium accuracy and complexity among the three models. Besides, the accuracy of PCANet is nearly the same as CNet in some specific channel conditions. Compared with the state-of-the-art of CsiNet, the feedback accuracy of CNet (or FCNet) is much better (or slightly worse). With much lower complexity, PCANet performs better in different scenarios.

Acknowledgment. This work was supported in part by the China Natural Science Funding under Grant 61731004.

References

1. 3GPP Technical Specification (TS) 38.211: NR; Physical channels and modulation (Release 15) (2017)
2. 3GPP Technical Specification (TS) 38.214: NR; Physical layer procedure for data (Release 15) (2017)
3. Zhao, L., Zhao, H., Zheng, K., Xiang, W.: Massive MIMO in 5G Networks: Selected Applications, 1st edn. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-319-68409-3
4. Zheng, K., Zhao, L., Mei, J., Shao, B., Xiang, W., Hanzo, L.: Survey of large-scale MIMO systems. IEEE Commun. Surv. Tutor. **17**(3), 1738–1760 (2015)
5. Wen, C., Shih, W., Jin, S.: Deep learning for massive MIMO CSI feedback. IEEE Wirel. Commun. Lett. **7**(5), 748–751 (2018)
6. Akbari, M., Liang, J., Han, J.: DSSLIC: deep semantic segmentation-based layered image compression. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, pp. 2042–2046 (2019)
7. Huang, C., Nguyen, T., Lai, C.: Multi-channel multi-loss deep learning based compression model for color images. In: IEEE International Conference on Image Processing (ICIP), Taipei, pp. 4524–4528 (2019)
8. Cai, C., Chen, L., Zhang, X., Gao, Z.: Efficient variable rate image compression with multi-scale decomposition network. IEEE Trans. Circuits Syst. Video Technol. **29**(12), 3687–3700 (2019)
9. Guo, J., Wen, C., Jin, S., Li, G.Y.: Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: design, simulation, and analysis. IEEE Trans. Wirel. Commun. **19**(4), 2827–2840 (2020)

10. Lu, Z., Wang, J., Song, J.: Multi-resolution CSI feedback with deep learning in massive MIMO system. In: IEEE International Conference on Communications (ICC), Dublin, pp. 1–6 (2020)
11. Qing, C., Cai, B., Yang, Q., Wang, J., Huang, C.: Deep learning for CSI feedback based on superimposed coding. IEEE Access **7**, 93723–93733 (2019)
12. Liu, Y., You, S.D., Chung, C.: Channel estimation method for LTE downlink MIMO transmission. In: IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, pp. 719–722 (2019)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp. 770–778 (2016)
14. Chen, T., Guo, J., Jin, S., Wen, C., Li, G.Y.: A novel quantization method for deep learning-based massive MIMO CSI feedback. In: IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, pp. 1–5 (2019)