Millimeter Wave Hybrid Precoding Based on Deep Learning

Qing Liu(✉) and Ken Long

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

s180101008@stu.cqupt.edu.cn

Abstract. To overcome the high energy consumption, and insufficient use of spatial information in traditional hybrid precoding algorithms, a hybrid precoding algorithm based on deep learning was proposed for millimeter wave massive MIMO system. Firstly, the hybrid precoding design problem was transformed into an exhaustive search problem for the analog precoder/combiner using the equivalent channel matrix. Then, a deep learning model was constructed to learn how to optimize the cascaded hybrid precoder by an improved convolutional neural network model. Finally, the optimized cascaded hybrid precoder was used to predict the analog precoding/combiner matrix directly, and the digital precoding/combiner matrix was obtained by applying singular value decomposition (SVD) to the equivalent channel matrix. Simulation results show that the performance of the proposed cascaded hybrid precoder is close to that of the pure digital precoder and can maximize the achievable rate to enhance the spectral efficiency of the millimeter wave massive MIMO system.

Keywords: Millimeter wave · Deep learning · Hybrid precoding · MIMO

1 Introduction

The need to transmit large amounts of data in a short period of time has exploded in recent years, and it is expected that wireless data traffic may grow by more than 10,000 times by 2030 [1]. Millimeter wave communication is an alternative solution to improve spectral efficiency and frequency resource shortage [2]. In the past decade, hybrid beamforming technology has been widely used to support high data rates and energy efficiency. Channel estimation in millimeter-wave is challenging due to the training overhead and strict restrictions on the use of Radio Frequency (RF) chains by a large number of transmitting and receiving antennas [3].

Inspired by massive multiple input multiple output (MIMO), millimeter wave massive MIMO system is considered as a potential technology to improve system throughput. Hybrid precoding [1] emerged as a result of multi-multiplexing massive data streams and achieving more accurate beamforming in millimeter wave massive MIMO. In [4], a hybrid beamforming based on continuous interference elimination was proposed, which

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 13–27, 2022.
https://doi.org/10.1007/978-3-030-99200-2_2

can provide high performance at low complexity. Its main idea was to decompose the total realizable rate optimization problem with non-convex constraints into a series of simple sub-rate optimization problems. The author in [5] designed a hybrid precoding with low complexity for multi-user millimeter-wave system by configuring a hybrid precoder. However, the previously proposed hybrid precoding scheme is based on singular value decomposition (SVD), resulting in high communication complexity and requiring complex bit allocation strategy. In addition, the proposed scheme based on Geometric Mean Decomposition (GMD) [6] can avoid the bit allocation problem, but it still faces great challenges in solving the non-convex constraints of the analog precoder and utilizing the structural characteristics of millimeter wave system.

In the context of millimeter wave massive MIMO systems, although a lot of research has been done to enhance the performance of hybrid precoding, there are still many problems, the two main challenges are high cost and poor system performance. In the past few years, a number of hybrid precoding methods have been proposed to reduce costs or improve precoding performance [7–9]. At the same time, in order to achieve high spectral efficiency and reduce cost, the author in [10] proposed an alternating minimization scheme for effective design of hybrid precoder. In [11], a hybrid precoding method was proposed based on beam space SVD by using low-dimensional beam space channel state information (CSI) processed by a compressed sensing (CS) detector. These traditional methods cannot make full use of the structural characteristics of millimeter-wave systems, while the traditional low-cost schemes are realized at the cost of reducing the system's mixed precoding. Therefore, it is urgent to propose a new method to enhance the hybrid precoding performance of millimeter wave massive MIMO systems.

This paper intends to design a millimeter-wave hybrid precoding method based on deep learning, which jointly optimizes the channel measurement vector and designs a hybrid beamforming vector to achieve a near-optimal data rate with negligible training overhead.

2 System Model and Problem Description

2.1 System Model

The fully connected hybrid architecture depicted in Fig. 1, where a transmitter employs N_t antennas and N_t^{RF} RF (Radio Frequency) chains is communicating via N_s streams with a receiver which has N_r antennas and N_r^{RF} RF chains. The transmitter pre-codes the transmitted signal using a $N_t^{RF} \times N_s$ baseband precoder $\mathbf{F}_{\mathrm{BB}}$ and a $N_t \times N_t^{RF}$ RF beam-former $\mathbf{F}_{\mathrm{RF}}$, while the receiver combines the received signal using the $N_r \times N_r^{RF}$ RF combiner $\mathbf{W}_{\mathrm{RF}}$ and the $N_r^{RF} \times N_s$ baseband combiner $\mathbf{W}_{\mathrm{BB}}$. $\mathbf{F}_{\mathrm{RF}}$ is implemented in the analog domain using RF circuits, every entry of the RF precoders is assumed to have a constant-modulus, i.e., $\left|[\mathbf{F}_{\mathrm{RF}}]_{m,n}\right|^2 = N_t^{-1}$. Further, the total power constraint is enforced by normalizing the baseband precoder $\mathbf{F}_{\mathrm{BB}}$ to satisfy $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s$ [12].

Due to the limited scattering of the millimeter wave channel, the single-user model is taken as the research object, if only L paths reach the receiving antenna, the extended

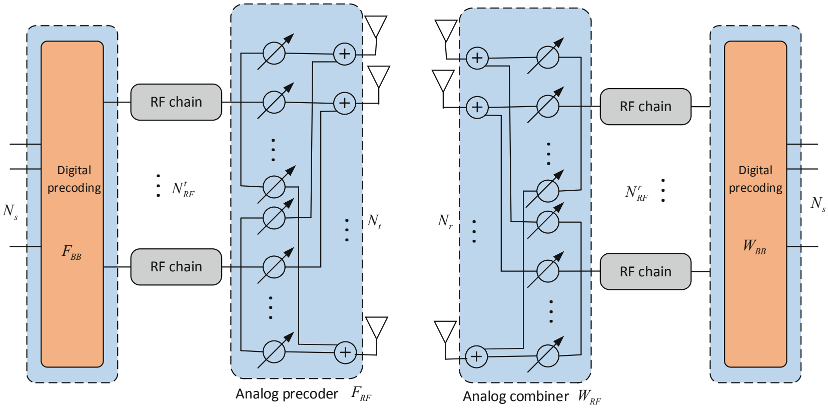

Fig. 1. Fully connected hybrid precoding model

Saleh-Valenzudel geometric channel model [12] is simplified as

$$\mathbf{H} = \sum_{\ell=1}^{L} \alpha_\ell a_r\left(\phi_{r,\ell}, \theta_{r,\ell}\right) a_t^H\left(\phi_{t,\ell}, \theta_{t,\ell}\right) \tag{1}$$

where α_ℓ denote the complex path gain of the ℓ th path. The angles $\phi_{r,\ell}, \theta_{r,\ell}$ represent the ℓ th path azimuth and elevation angles of arrival (AOAs) at the receive antennas while $\phi_{t,\ell}, \theta_{t,\ell}$ represent the ℓ th path angles of departure from the transmit array. $a_t(\cdot)$ and $a_r(\cdot)$ represent the sending and receiving array response vector. According to the arrangement of antenna elements, common antenna arrays can be divided into uniform linear array (ULA) and uniform planar array (UPA). The array response vector of ULA and UPA arrays are defined in [13]. The array antenna arrangement adopted in this paper is ULA.

2.2 Problem Description

The overall goal of this paper is to directly design hybrid precoders/combinators to maximize the rate of system implementation. If the channel is known and the RF beamforming/combinator vector is selected from the predefined quantization codebook, the hybrid precoding/combinator design problem can be expressed as

$$\begin{aligned}
&\left\{\mathbf{F}_{\mathrm{BB}}^*, \mathbf{F}_{\mathrm{RF}}^*, \mathbf{W}_{\mathrm{BB}}^*, \mathbf{W}_{\mathrm{RF}}^*\right\} = \arg\max \log_2\left|\mathbf{I} + \mathbf{R}_n^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{H}\mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W}\right| \\
&s.t.\ \mathbf{F} = \mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{RF}} \\
&\mathbf{W}{=}\mathbf{W}_{\mathrm{BB}}\mathbf{W}_{\mathrm{RF}} \\
&[\mathbf{F}_{\mathrm{RF}}]_{:,n_t} \in \mathcal{F}, \forall n_t \\
&[\mathbf{W}_{\mathrm{RF}}]_{:,n_r} \in \mathcal{W}, \forall n_r \\
&\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = \mathrm{N_s}
\end{aligned} \tag{2}$$

where $\mathbf{R}_n$ represent the noise covariance matrix with $\mathbf{R}_n = \left(N_s\sigma_n^2/P_T\right)\mathbf{W}^{\mathrm{H}}\mathbf{W}$, where P_T denote the total transmit power and σ_n^2 denote noise power. Further,

$\mathcal{F}=\{c^1, c^2, \ldots, c^{|\mathcal{F}|}\}$ represent the predefined quantization codebook at the base station and $\mathcal{W} = \{g^1, g^2, \ldots g^{|\mathcal{W}|}\}$ represent the predefined quantization codebook on the user side.

If the RF beamforming/combined codebook consists of orthogonal vectors (such as the DFT codebook), then for any selected RF precoder $\mathbf{F}_{\mathrm{RF}}$ and RF combiner $\mathbf{W}_{\mathrm{RF}}$, the best baseband precoder/combiner is expressed as

$$\mathbf{F}_{\mathrm{BB}}^{*} = \left(\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}[\overline{\mathbf{V}}]_{:,1:\mathrm{N}_s} \tag{3}$$

$$\mathbf{W}_{\mathrm{BB}}^{*} = [\overline{\mathbf{U}}]_{:,1:\mathrm{N}_S} \tag{4}$$

where $\overline{\mathbf{V}}$ and $\overline{\mathbf{U}}$ represent the left and right singular value orthogonal matrix of the effective channel matrix with $\overline{\mathbf{H}} = \mathbf{W}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}$. Then the optimal hybrid precoder can be found through exhaustive search of candidate RF beamforming/combination vectors, which means the hybrid precoding design problem can be transformed into an exhaustive search problem of RF precoding/combination matrix

$$\begin{aligned}&\left\{\mathbf{F}_{RF}^{*}, \mathbf{W}_{RF}^{*}\right\} = \arg\max\ \log_2\left|\mathbf{I} + \mathrm{SNR}\mathbf{W}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}} \times \left(\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}\right)^{-1}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W}_{\mathrm{RF}}\right|\\&s.t.\ [F_{RF}]\ :, \mathrm{n}_t \in \mathcal{F}, \forall n_t\\&[W_{RF}]\ :, \mathrm{n}_r \in \mathcal{W}, \forall n_r\end{aligned} \tag{5}$$

3 Problem Solving

3.1 Millimeter Wave Compression Channel Sensing

This paper proposes a novel neural network architecture that can directly find the RF precoding/combination vector in the hybrid precoding architecture, and the vector can maximize the best achievable rate. According to the system model in Sect. 2.1, the achievable rate of the encoder/combiner is expressed as

$$R = \log_2\left|\mathbf{I} + \mathbf{R}_n^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{H}\mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W}\right| \tag{6}$$

where $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$, $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$.

Let $\mathbf{P}$ and $\mathbf{Q}$ denote the $N_t \times M_t$ and $N_r \times M_r$ channel measurement matrix adopted by both the transmitter and receiver to sense the channel $\mathbf{H}$, with M_t and M_r representing the number of transmit/receiver measurements. If the pilot symbol is equal to 1, then the received measurement matrix, $\mathbf{Y}$, can be written as

$$\mathbf{Y} = \sqrt{P_{\mathrm{T}}}\mathbf{Q}^H\mathbf{H}\mathbf{P} + \mathbf{Q}^H\mathbf{V} \tag{7}$$

where $[\mathbf{V}]_{\mathrm{m,n}} \sim \mathcal{N}\left(0, \sigma_n^2\right)$ denote receive measurement noise, and when vectorizing this matrix, we get

$$\mathrm{y} = \sqrt{\mathrm{P_T}}\left(\mathbf{P}^{\mathrm{T}} \otimes \mathbf{Q}^{\mathrm{H}}\right)\mathrm{h}+\mathrm{v_q} \tag{8}$$

where $y = \mathrm{vec}(\mathbf{Y})$, $\mathrm{v} = \mathrm{vec}\left(\mathbf{Q}^{\mathrm{H}}\mathbf{V}\right)$, $\mathrm{h} = \mathrm{vec}(\mathbf{H})$.

3.2 Deep Learning Model

In Sect. 2.2, the hybrid precoding design problem has been transformed into the solution problem of analog precoding/combination matrix or the binary classification problem of analog precoding/combination matrix. Since convolutional neural network has significant effect on classification problems, this paper constructs a deep learning model based on the improved convolutional neural network.

The performance of deep learning models can be improved by increasing the depth and width of the network (the depth refers to the number of layers of the network, and the width refers to the number of channels at each layer), but it will lead to a very large number of network parameters. However, the massive parameters not only easily produce overfitting, but also greatly increase the amount of calculation. The Inception module can reduce the dimension of large-sized matrices and aggregate visual information in different sizes to facilitate feature extraction from different scales and achieve better performance. The cascade hybrid precoder structure proposed in this paper is mainly composed of two parts, as shown in Fig. 2.

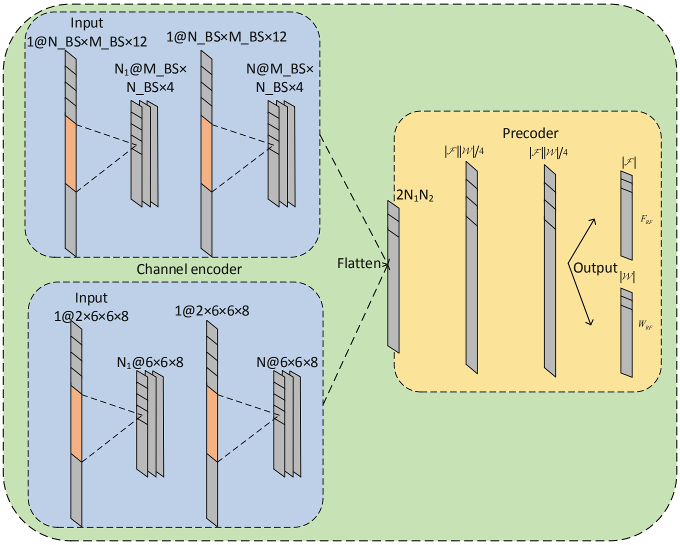

Fig. 2. Cascade hybrid precoding model

The first part is the channel encoder, which consists of two parallel convolution modules. This module is inspired by inception module in GoogLeNet architecture, and the convolution in the inception module is modified to a two-dimensional complex convolution. The first module is shown in Fig. 3, which contains three complex convolution blocks with kernel sizes of M_BS, 12 and 10, respectively. Where N_BS represents the number of antennas at the base station, and M_BS represents the channel measurement vector. Complex convolution blocks create a complex convolution kernel, which is convolved with a complex input layer to produce a complex output tensor. The complex convolution block has complex weights during initialization, which means that the input of the neural network is essentially complex channel coefficients. Therefore, the signal contains real and imaginary feature maps after the convolution operation. Two blocks with kernel sizes of 12 and 10 are connected behind the complex con-volution block to reduce the dimensionality. The output of the block is in series at a depth dimension that preserves the inclination characteristic. The second module is shown in Fig. 4, which contains three complex convolution blocks with kernel sizes of 6, 5 and 4, respectively. In the two modules, different kernel sizes are selected from low to high, with low kernel extracting local features and high kernel extracting global features.

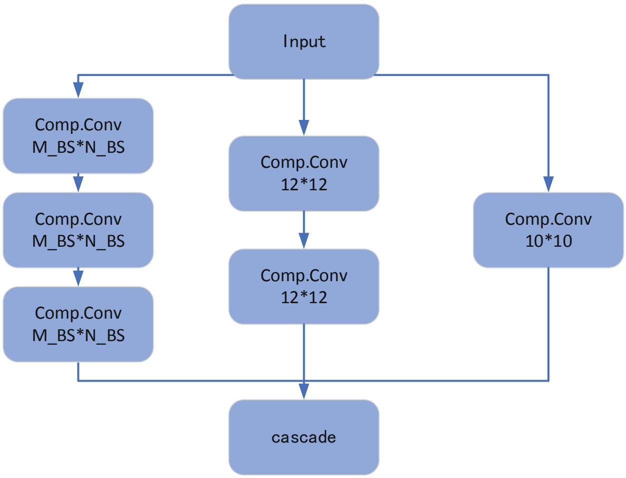

Fig. 3. Model 1

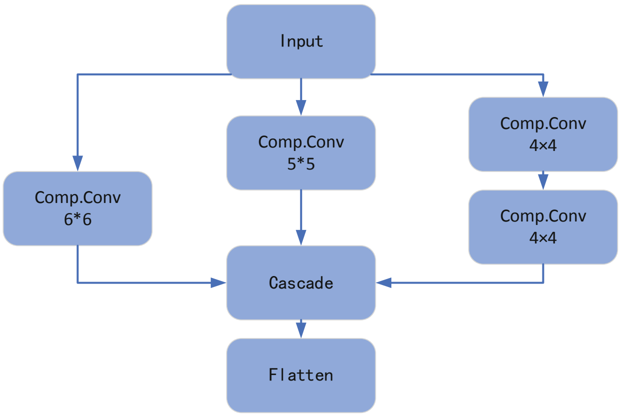

Fig. 4. Model 2

The second part is the precoder that consists of two fully connected layers and two output layers, which takes the output of the channel encoder as input. Through two fully connected layers and fully connected layer maps the learned distributed feature representation to the sample label space for classification. The number of neurons in the two fully connected layers are equal to $|\mathcal{F}||\mathcal{W}|/4$, sigmoid activation and Batch Norm standardization are performed after each fully connected layer. The size of the two output layers is equal to the number of transmit and receive codebooks $|\mathcal{F}|$ and $|\mathcal{W}|$, which means the number of candidate beamforming/combination vectors, and the activation function is Sigmoid.

3.3 Neural Network Training

In the training stage, the cascade hybrid precoder performs end-to-end training in a supervised manner. The data set of millimeter-wave channels and the corresponding RF beamforming/combination matrix are first constructed, and then the cascade hybrid precoder is trained to be able to predict the index of RF precoding/combination vectors given the input channel vector.

In this paper, the neural network is trained based on the channel vector of hybrid architecture and the corresponding data set of RF beamforming/combination vector, and the target RF precoding/combination matrix is calculated based on the approximately optimal Gram-Schmidt hybrid precoding algorithm in [11]. The training and learning process of cascaded hybrid pre-coding neural network is shown in Table 1.

Table 1. The training and learning process of cascaded hybrid pre-coding

Algorithm 1 Neural Network Training
1: Set data parameters, start the data generator, and import training data and test data
2: Design a cascaded precoding neural network structure, set the convolution kernel, learning rate, loss function, number of training rounds epoch and batch size
3: Loop
4: for i=1:1: epoch
5: Update neural network weights according to Adam optimization algorithm
6: end for
7: Output the trained cascade precoding neural network

3.4 Neural Network Prediction

The hybrid precoding neural network architecture is divided into two parts in the prediction stage. The channel encoder is implemented in the RF circuit, and the weights of the two parallel convolution modules of the channel encoder network are used as the weights of the analog/RF measurement matrix at the transmitter and receiver. The precoder uses the output of the channel encoder as input to directly predict the index of the rf beamforming/combination vector of the hybrid architecture.

After the neural network is trained, the trained neural network is used to design the hybrid precoding matrix. The hybrid precoding/combination matrix is predicted by the neural network first, and then the baseband precoding/combination matrix is calculated by the formula (3) and (4). The prediction process is shown in Table 2.

Table 2. The prediction process of cascaded hybrid pre-coding

Algorithm 2 Neural Network Prediction
1: Start the trained cascade precoding neural network
2: Predict the mixed precoding matrix $\mathbf{F}_{\mathrm{RF}}$ and the combined matrix $\mathbf{W}_{\mathrm{RF}}$
3: Estimate the channel matrix $\tilde{\mathbf{H}} = \mathbf{QYP}^H$
4: Perform singular value decomposition on the channel matrix $\tilde{\mathbf{H}} = \mathbf{U\Sigma V}^H$
5: Calculate baseband precoding $\mathbf{F}_{\mathrm{BB}}$ and combination matrix $\mathbf{W}_{\mathrm{BB}}$ according to equations (3) and (4)
6: Output the matrix $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$

The loss function can evaluate the difference between the predicted value and the real value of the model, and the better the loss function, the better the performance of the model. In this paper, the binary cross entropy function is adopted as the loss function of two multi-label classification tasks. The total loss function is the arithmetic mean of binary cross entropy of pre-coding and combination tasks, which is expressed as

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\left[y_{true}\log y_{pred} + (1 - y_{true})\log\left(1 - y_{pred}\right)\right] \tag{9}$$

4 Simulation Analysis

4.1 Simulation Settings

In this paper, ray-tracing mode is adopted for simulation, and the data set is the universal Deep MIMO [14] data set that can be obtained publicly to generate simulation parameters, as shown in Table 3. The size of the training dataset [14] is 31200, and the size of the prediction dataset is 5000. During the experiment, Keras library and TensorFlow were used, and Adam optimizer was used to optimize the model.

Table 3. The adopted Deep MIMO dataset parameters

Parameter	Value
Number of base stations	4
Number of base station/user terminal antennas	64
System bandwidth/GHz	0.5
Receive noise figure/dB	5
Antenna spacing/wavelength	0.5
Total transmit power/dBm	5,10,15,20,25,30
Number of subcarriers	1024
OFDM sampling coefficient	1
OFDM limitations	1
Number of channel paths	3
Transmitter/receiver RF link number	3
Base station/user side codebook size	64

4.2 Achievable Rate

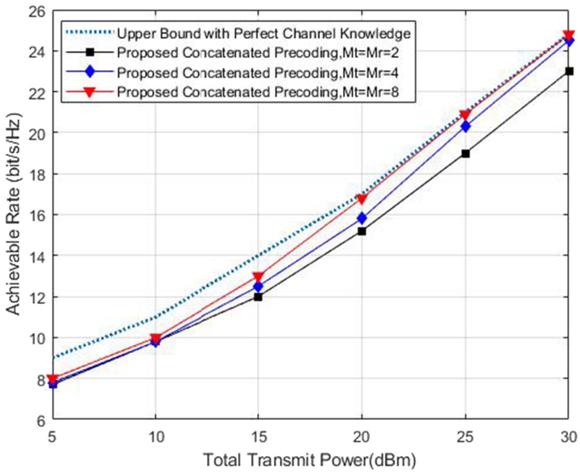

Fig. 5. Achievable rate of cascade hybrid precoder

Figure 5 shows the relationship between the achievable rate of the proposed concatenated precoding and the total transmit power when the channel measurement value is equal to 2, 4, and 8. It can be seen from the figure that the performance of the proposed cascaded precoder is close to the upper limit under a reasonable transmit power value. Additionally, the figure illustrates that the training overhead of the cascaded precoder compared with the traditional hybrid precoding method is significantly reduced. For example, with only 4 channel measurements at both transmitter and receiver, i.e., 16 pilots, the proposed cascaded precoder based on deep learning almost reaches the upper limit of spectral efficiency, which replaces the exhaustive search method required 4096 pilots.

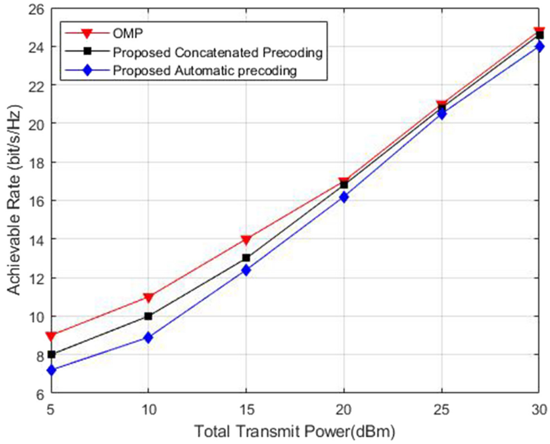

Fig. 6. The achievable rates of several hybrid precoding algorithms

Figure 6 shows the relationship between the achievable rate and the total transmit power of the cascaded hybrid precoder, the spatial sparse precoding method, and the automatic precoder. It can be seen from the figure that in all the schemes, the achievable rate increases as the total transmit power increases. In addition, the performance of the cascaded hybrid precoding scheme is better than that of the automatic precoder. When total transmit power increases, the performance gap between the cascaded precoding scheme and the automatic precoding becomes larger and larger. This is because of the excellent mapping and learning capabilities of the deep learning model to achieve better hybrid precoding performance.

Figure 7 shows the relationship between the achievable rate and the total transmit power for different learning rates. It can be seen from the figure that the proposed cascaded precoding is optimized by a lower learning rate, because a higher learning rate will lead to a higher verification error. However, the convergence speed will be slower due to the lower learning rate. Therefore, the optimal learning rate should be selected reasonably to achieve better performance. This paper sets the learning rate to 0.01.

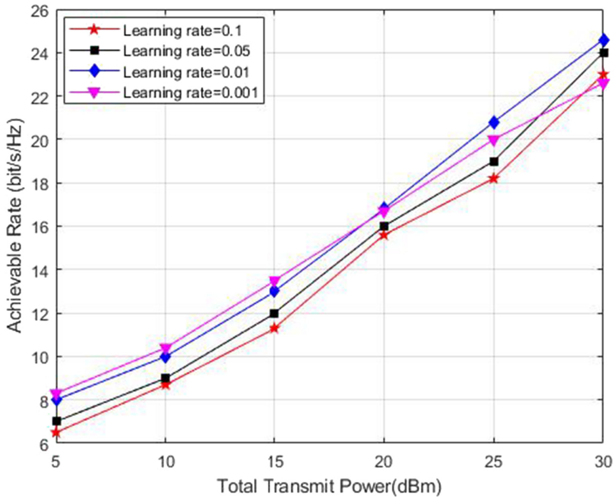

Fig. 7. The achievable rates of the cascaded hybrid precoding method based on deep learning for different total transmit power values and different learning rates

4.3 Predict Accuracy

Since the performance of the deep learning model in various tasks requires quantitative indicators to be evaluated, the sample accuracy of the prediction indicators is calculated in this paper to evaluate the prediction performance of the proposed algorithm. Let $y_{true}^{(i)}$ denote the set of label index, $y_{pred}^{(i)}$ denote the set of prediction index, and the accuracy of sample i is defined as

$$a_i = \frac{\left| y_{true}^{(i)} \cap y_{pred}^{(i)} \right|}{\left| y_{true}^{(i)} \right|} \tag{10}$$

where $\cap$ denote the intersection of two sets and $|\cdot|$ denote cardinality of the set. The sample accuracy can be expressed as

$$a = \frac{\sum_{i=1}^{N} a_i}{N_{samples}} \tag{11}$$

where $N_{samples}$ represent the number of samples.

Under different total transmitting powers, the classification accuracy of the simulated precoding vector at the base station end and the simulated combination vector at the client end is shown in Table 4.

It can be seen from Table 2 that the accuracy for the transmit and receive predicted beams are almost the same since we treat them as equally important tasks when training the auto-precoder neural network model. Besides, the classification accuracy of the cascaded precoder increases continuously with the increase of the total transmit power when the number of channel measurement vectors remains. The classification accuracy

Table 4. Classification accuracy of cascaded precoder for different values of total transmit power and number of channel measurement vectors

P_T(dBm)	5	10	15	20	25	30
Tx acc (Mt = Mr = 2)	0.7449	0.8217	0.8584	0.8879	0.9217	0.9373
Rx acc (Mt = Mr = 2)	0.7553	0.8344	0.8537	0.8934	0.9288	0.9412
Tx acc (Mt = Mr = 4)	0.7565	0.8387	0.8623	0.8984	0.9316	0.9354
Rx acc (Mt = Mr = 4)	0.7642	0.8392	0.8679	0.9030	0.9369	0.9437
Tx acc (Mt = Mr = 8)	0.7683	0.8416	0.8711	0.9120	0.9374	0.9448
Rx acc (Mt = Mr = 8)	0.7713	0.8467	0.8693	0.9137	0.9355	0.9451

of cascaded precoder increases continuously with the increase of the number of channel measurement vectors when the total transmit power remains because a larger number of pilots can obtain more channel state information, which makes the neural network use more features for classification during training. However, when the number of pilots continues to increase, the classification accuracy tends to stabilize, whose value stabilizes at 0.94. This is because no more channel state information can be extracted when the number of pilots exceeds a certain threshold, which means the channel state information has reached the maximum. Therefore, the classification accuracy will no longer continue to increase.

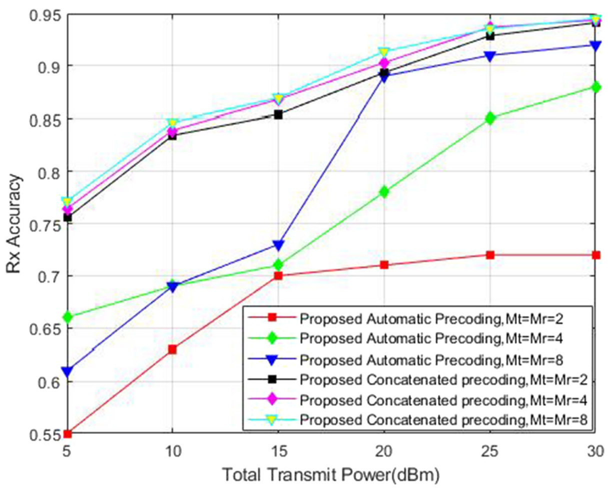

Fig. 8. Comparison of classification accuracy between cascaded hybrid precoding and automatic precoding

Figure 8 shows the relationship between the classification accuracy of the proposed cascaded precoder and auto-precoder [15] methods with different total transmit power. It can be seen from the figure that the classification accuracy of these two schemes increases with the increase of the total transmission power, and the classification accuracy of the cascaded precoder is better than the classification accuracy of the auto-precoder. In addition, the stability of cascaded hybrid precoder is better than that of automatic precoder. Classification accuracy is significantly improved especially with less pilots. For example, for a transmit power of 20 dBm, the accuracy of the auto-precoder is 0.77, and the accuracy of the cascaded precoding is increased to 0.9030, which improves by around 17.27%. This shows that the performance of the cascaded precoder is superior than the auto-precoder.

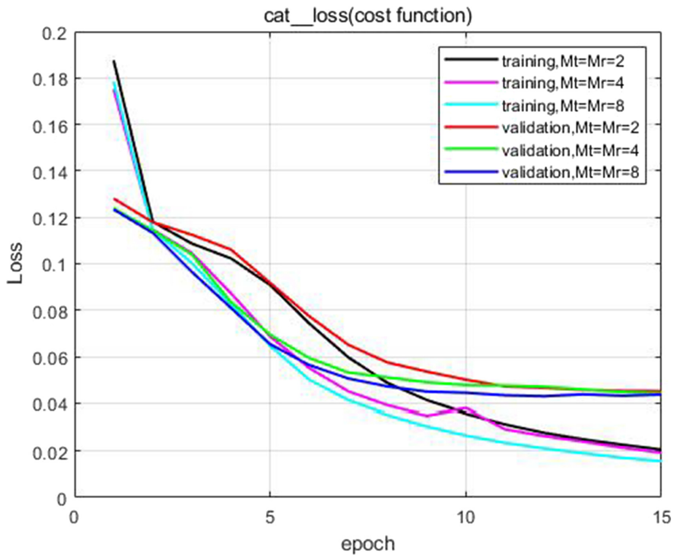

Fig. 9. Training vs Validation Graph for 15 epochs

Figure 9 shows the relationship between the size of the cascaded precoding loss function and the number of iterations. It can be seen from the figure that the loss function curve changes smoothly on the training set and the test set, without excessive fluctuations or jaggedness, which indicate that the learning rate set is reasonable, and it is close to a constant value under a limited number of iterations. There is no obvious gap between the loss function of the training set and the loss function of the test set, indicating that the network does not appear to be under-fitting and overfitting.

5 Conclusions

This paper proposes a cascade hybrid precoding algorithm based on deep learning for millimeter wave hybrid precoding. Cascaded hybrid precoder learns how to optimize channel sensing vectors to concentrate sensing power in the direction most desired, and how to predict RF beamforming and combination vectors of hybrid architectures directly from the received sensing vectors. Compared with the traditional method, the proposed deep learning method can directly predict the simulated precoding/combination vector under the condition of unknown channel state information, which overcomes the huge training cost required to select the optimal beamforming vector in the large antenna array millimeter wave system. At the same time, the data rate is close to the best data bit rate. Compared with some existing deep learning schemes, the performance of cascade hybrid precoding is closer to that of all-digital precoding algorithm, and it can accurately predict the analog precoding matrix/combination matrix with fewer channel measurement vectors.

References

1. Ghosh, A., Thomas, T.A., Cudak, M.C., et al.: Millimeter-wave enhanced local area systems: a high-data-rate approach for future wireless net-works. IEEE J. Sel. Areas Commun. **32**(6), 1152–1163 (2014)
2. Alkhateeb, A., Mo, J., Gonzalez-Prelcic, N., et al.: MIMO precoding and combining solutions for millimeter-wave systems. IEEE Commun. Mag. **52**(12), 122–131 (2014)
3. Heath, R.W., Gonzalez-Prelcic, N., Rangan, S., et al.: An overview of signal processing techniques for millimeter wave MIMO systems. IEEE J. Sel. Top. Signal Process. **10**(3), 436–453 (2016)
4. Gao, X., Dai, L., Han, S., et al.: Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. IEEE J. Sel. Areas Commun. **34**(4), 998–1009 (2016)
5. Alkhateeb, A., Leus, G., Heath, R.W.: Limited feedback hybrid precoding for multi-user millimeter wave systems. IEEE Trans. Wirel. Commun. **14**(11), 6481–6494 (2015)
6. Chen, C.E., Tsai, Y.C., Yang, C.H.: An iterative geometric mean decomposition algorithm for MIMO communications systems. IEEE Trans. Wirel. Commun. **14**(1), 343–352 (2014)
7. Jin, J., Zheng, Y.R., Chen, W., et al.: Hybrid precoding for millimeter wave MIMO systems: a matrix factorization approach. IEEE Trans. Wirel. Commun. **17**(5), 3327–3339 (2018)
8. Zhang, E., Huang, C.: On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems. In: 2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall), pp. 1–5. IEEE (2014)
9. Wang, G., Ascheid, G.: Joint pre/post-processing design for large milli-meter wave hybrid spatial processing systems. In: European Wireless 2014; 20th European Wireless Conference, VDE, pp. 1–6 (2014)
10. Yu, X., Shen, J.C., Zhang, J., et al.: Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems. IEEE J. Sel. Top. Signal Process. **10**(3), 485–500 (2016)
11. Chen, C.H., Tsai, C.R., Liu, Y.H., et al.: Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems. IEEE Trans. Signal Process. **65**(6), 1412–1424 (2016)
12. Alkhateeb, A., El Ayach, O., Leus, G., et al.: Channel estimation and hybrid precoding for millimeter wave cellular systems. IEEE J. Sel. Top. Signal Process. **8**(5), 831–846 (2014)
13. El Ayach, O., Rajagopal, S., Abu-Surra, S., et al.: Spatially sparse pre-coding in millimeter wave MIMO systems. IEEE Trans. Wirel. Commun. **13**(3), 1499–1513 (2014)
14. Alkhateeb, A.: DeepMIMO: a generic deep learning dataset for millimeter wave and massive MIMO applications. arXiv preprint arXiv:1902.06435 (2019)
15. Li, X., Alkhateeb, A.: Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems. In: 2019 53rd Asilomar Conference on Signals, Systems, and Computers, pp. 800–805. IEEE (2019)

16. Huang, H., Song, Y., Yang, J., et al.: Deep-learning-based millimeter-wave massive MIMO for hybrid precoding. IEEE Trans. Veh. Technol. **68**(3), 3027–3032 (2019)
17. Bao, X., Feng, W., Zheng, J., et al.: Deep CNN and equivalent channel based hybrid precoding for mmWave massive MIMO systems. IEEE Access **8**, 19327–19335 (2020)