SE Block-Assisted ResNet for Channel Estimation in OFDM System

Yuanhai Liang and Zhengfa Zhu(✉)

Changsha University of Science and Technology, Changsha, China
1054575560@qq.com, zhuzhengfa@csust.edu.cn

Abstract. Channel estimation (CE) is an important part of wireless communication system, which has a significant impact on the quality of wireless communication. Considering a single-input single-output (SISO) downlink scenario, this paper proposes a "squeeze and excitation" (SE) block combined with residual neural network (SE-ResNet) method to improve the CE performance of the orthogonal frequency division multiplexing (OFDM) system. The SE-ResNet is inputted into a CSI matrix of the pilot position obtained by the least squares (LS) method, and a raw feature matrix is learned by the convolutional layer and a partial residual layer. And the global information of the original feature matrix in each channel is compressed into a descriptor through the squeeze operation in the "SE" block. Then attention map is obtained from information aggregated in the descriptors by an excitation operation which fully captures channel-wise dependencies. The attention map is multiplied by the original feature matrix to get a new feature matrix, which is resized by an interpolation layer to obtain CSI of entire frame. New feature matrix is helpful for interpolation layer to get more accurate complete CSI. To further improve CE performance and optimize network model, this paper adopts methods of network tailoring and Concrete Autoencoder (Concrete AE) for pilot design. Simulation results show that the proposed SE-ResNet is superior to the traditional LS and minimum mean squared error (MMSE) methods in various practical wireless environments. Network pruning will reduce the number of parameters of the network within an acceptable loss range, which will reduce computational costs. Using the pilot scheme designed by Concrete AE for CE will have better performance.

Keywords: Channel Estimation · Squeeze and Excitation Block · Residual Convolutional neural network · Concrete Autoencoder

1 Introduction

With the popularization of 5G technology, fresh challenges such as massive connectivity and ultra reliability has raised and need to be met [1]. Subsequently, CE, as one of the key technologies in wireless communication system, also needs to meet higher accuracy requirement. Among conventional CE methods, LS method is easily implemented and widely adopted in practical communication systems, which is an interpolation-based

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 386–396, 2023.
https://doi.org/10.1007/978-3-031-31733-0_32

approach and requires no information about the statistics of the channel. However, it cannot obtain satisfactory performance in some scenarios that require higher accuracy. In order to obtain more accurate CSI, MMSE method was proposed to refine CSI, but it is difficult to implement in practical scenarios because of complete channel statistics requirement and high computational complexity.

With the rapid development in computer vision, natural language processing, semantic recognition, and so on, deep learning (DL) has also been introduced into wireless communication system to estimate CSI. At present, DL is divided into two categories to realize CE, namely data-driven and model-driven. In data-driven method, the receiver in wireless communication system is considered as a black box, and wireless signals from transmitter can be accurately recovered at the receiver. For example, in 2, the receiver is instead of a simple designed five-layer deep neural network (DNN) for signal detection and CE. It is shown the DNN based approach, via training with a large amount of data, can achieve the performance comparable to the MMSE estimator. The data-driven scheme proposed above do not rely on channel statistical knowledge, so they may be a promising candidate when channel model is unknown or difficult to model analytically, for example, in high mobility vehicular communications, chemical communications, underwater communications, and so on [3]. Although data-driven method is relatively simple and effective, it requires a large amount of training data, which limiting its application. The model-driven CE approach, which improves neural network by characteristics of the model structure and then trains a large amount of data, is more powerful and has wider applications. For example, in 4 the authors apply DL to estimate the uplink channel of a hybrid analog-to-digital converter massive multiple-input multiple-output (MIMO) system. The received signals of all antennas are used by DNN to estimate the channel, and to eliminate the adverse effects of coarsely quantized signals, a selective input prediction DNN (SIP-DNN) is developed. Signals received by the high-resolution analog-to-digital converter antennas are utilized in SIP-DNN to predict the channels of other antennas as well as to estimate their own channels.

Recently, many DL-based CE approaches have been proposed. In [5], the authors introduced an attention-aided DL framework for massive MIMO systems. By integrating attention mechanism into fully connected network, they improve CE performance significantly at the cost of small complexity overhead. In [6], the author takes the time-frequency response of the fast fading channel as a low-resolution image, and then proposes a ChannelNet network to use super-resolution method to obtain unknown channel response. Inspired by this, a residual learning based deep neural network, called ReEsNet, is designed and optimized for CE in [7], in which the up-sampling function is implemented as transposed convolution layer in order to scale up image height and width. However, the hyperparameters of the transposed convolution and subsequent convolution layers need to be modified according to different pilot patterns. Therefore, in [8], instead of the transposed convolutional layer, bilinear interpolation is proposed to interpolate CSI of pilot position to obtain the CSI of the entire frame, which improves the performance and reduces complexity by 82%. In [9] and [10], after the pilot design of Concrete AE, ChannelNet and a generative adversarial network are cascaded for CE respectively. The simulation results show that the performance of the networks cascaded

with Concrete AE are better than original networks. Although the performance of CE has been improved, it is not enough for high-precision communication requirements.

Motivated by the above papers, in this paper we propose a residual neural network combined with a SE block to improve CE accuracy. The full CSI is recovered by self-attention learning on the pilot CSI. Simulation results demonstrate that the presented SE-ResNet can achieve art-of-state performance and outperform other DL based estimation method. On this basis, we use Concrete AE to select the position with the most channel information in the time-frequency resource grid as pilot position, and cascade SE-ResNet for channel estimation. Finally, we will use network clipping to reduce parameters within the acceptable loss range. We conducted experiments and the results show that these methods are effective to improve the performance of CE. The rest of the paper is organized as follows: Sect. 2 introduces traditional CE methods, SE-ResNet is introduced in Sect. 3, followed by Sect. 4 with simulation results and a conclusion at the end.

2 Channel Estimation Based on Traditional Methods

OFDM technology is a multi-carrier modulation method that can make full use of spectrum resources, effectively resist frequency-selective interference and have a simple transmit and receive signal structure. In the frequency domain, the relationship between the transmit signal and the receive signal can be expressed as:

$$Y = H \circ X + W \tag{1}$$

where $X, Y \in \mathbb{C}^{N_c \times N_s}$ denotes the transmit and receive signals in the OFDM system, $H \in \mathbb{C}^{N_c \times N_s}$ denotes the channel coefficients, and $W \in \mathbb{C}^{N_c \times N_s}$ denotes the additive Gaussian white noise (AWGN). $\circ$ denotes the element-wise multiplication operation. And N_c, N_S represent the number of subcarriers and the number of OFDM symbols, respectively.

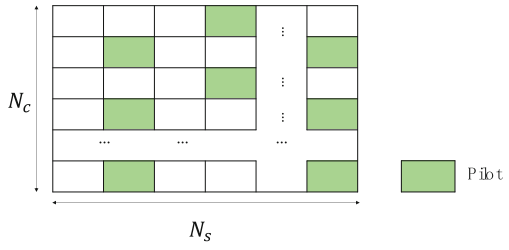

Fig. 1. Pilot and data symbols of one frame.

The frame structure is shown in Fig. 1. For each pilot OFDM symbol, some specific subcarriers are selected as the pilot carriers while the rest of the subcarriers are set to zero, while for the data OFDM symbols all subcarriers are used to transmit the modulated signal. The frequency domain channel estimation is performed based on the pilot OFDM symbols, and the channel coefficients of the whole frame are predicted based on this. The following two methods, LS and MMSE, are described.

2.1 LS Method

By neglecting the noise, the estimated channel gain matrix at the pilot position is derived and expressed as follows.

$$H_{LS} = \frac{Y_P}{X_p} \tag{2}$$

where Y_P, X_p denote the received and transmitted pilot signals, respectively, and the predicted channel gain of the whole frame is obtained by linear interpolation, because the noise calculation is ignored nor is the channel statistical information required, so the LS method is not ideal.

2.2 MMSE Method

By further processing the H_{LS} obtained by the LS method using priori information of the channel, the channel gain matrix of the pilot position can be expressed as follows.

$$H_{MMSE} = R_{HH_P}(R_{H_PH_P} + I\frac{\sigma_N^2}{\sigma_X^2})^{-1}H_{LS} \tag{3}$$

where H denotes the channel gain matrix of the pilot symbol, and H_P is the actual measured channel gain matrix of the pilot subcarrier. $\frac{\sigma_N^2}{\sigma_X^2}$ is the numerical inverse of the signal-to-noise ratio, while σ_X^2, σ_N^2 denote the average power of AWGN noise and transmit signal, respectively. R_{HH_P} and $R_{H_PH_P}$ are the inter-correlation matrix of H and H_P and the autocorrelation matrix of H_P.

Although the MMSE method increases the accuracy of CE by using the a priori information of the channel, it also increases the computational complexity and the difficulty to implementation as well.

3 Deep Learning Based Channel Estimation

Compared with the traditional CE methods mentioned above, although DL methods have advantages, neural network schemes may have gradient disappearance problem, and ResNet are designed to mitigate gradient disappearance. Therefore, researchers proposed many CE methods based on ResNet, such as 3.1 Interpolation-ResNet, and achieved good results. This section details the ResNet-based method proposed in this paper.

3.1 Interpolation-ResNet

The Interpolation-ResNet is an improved network model based on ResNet. The channel gain matrix of the pilot part obtained by LS method is used as the input of Interpolation-ResNet. Since the existing DL framework does not support complex number operations. The input of Interpolation-ResNet is divided into two parts: the real part, and the imaginary part, which are inputted into the network as two channels. Interpolation-ResNet

consists of 4 neural blocks, 3 convolutional layers, and an interpolation layer. Neural blocks and convolutional layers learn the features of pilot channel gain. Then, the channel gain matrix of pilot is interpolated by the interpolation layer to the channel gain matrix of the whole frame.

3.2 SE-ResNet

We keep the network structure of Interpolation-ResNet but introduce the SE attention module after the last neural block of the network. The SE-ResNet network structure is shown in Fig. 2, and the overall architecture of this network is divided into three parts.

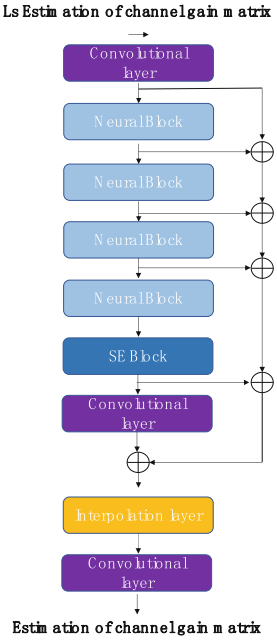

Fig. 2. SE-ResNet.

The first part is a convolutional layer, which has N convolution kernels, and the size of the convolution kernel is $3 \times 3 \times 2$. The second part is connected to the residual block composed of 4 neural blocks, an SE block, and a convolutional layer. Each neural block has two convolutional layers and a Relu layer in the middle. Each convolutional layer consists of N convolution kernels, and each convolution kernel is $3 \times 3 \times N$ in size. The last part consists of an interpolation layer and a convolutional layer, which has two convolution kernels with size is $36 \times 7 \times N$.

Input is divided into two parts: the real part, and the imaginary part, which are inputted to the first convolutional layer. In each neural block, the input of each neural block goes through two convolutional layers and a Relu layer, then the output of the Relu layer is added by the input, which is called residual operation, and summation is used

as input of next layer. Output of all the layers before interpolation layer are summed together and the result is forwarded to the interpolation layer, and then resized by bilinear interpolation. The resized data is input to the last convolutional layer to obtain the final output.

The original CNN uses all features with equal importance for all data, while in fact, some features will be more important to some data. Therefore, the introduced SE module can learn these features will play a more important role. The SE module is shown in Fig. 3.

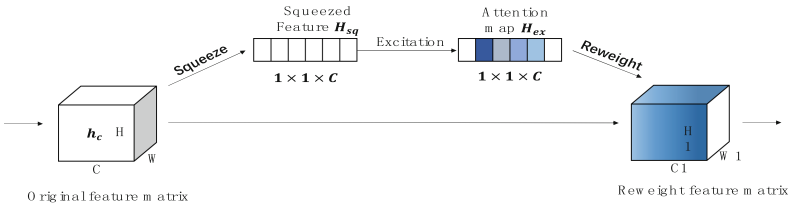

Fig. 3. SE Block.

A three-dimensional original feature matrix compressed according to dimension C, each two-dimensional feature matrix of size $W \times H$ is compressed into a descriptor. Each compressed descriptor represents global information of the corresponding two-dimensional matrix, and the final result is $1 \times 1 \times C$. The above compression process is the squeeze operation, which is implemented in global average pooling. The specific formula is as follows:

$$H_{sq} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} h_c(i, j) \tag{4}$$

The next two fully connected layers are taken as excitation operations. The first fully connected layer followed by a Relu layer, and the second one followed by a sigmoid layer, as in Eq. 5:

$$H_{ex} = \sigma(W_2 \delta(W_1 H_{sq})) \tag{5}$$

Multiplying the result obtained in (4) by W_1 whose dimension is $C/r \times C$, the reduction ratio r is a hyperparameter, which play an import role on SE block capacity and computational cost and in this paper it equals 1. The dimension of $W_1 H_{sq}$ is $1 \times 1 \times C$, and after a Relu layer. Similarly, after a fully connected layer with a weight of W_2 and a Sigmoid layer, the final dimension of H_{ex} is still $1 \times 1 \times C$. These two fully connected layers are used to generate weights for each feature attention map. Finally, the attention-based original features are recalibrated by multiplying the attention feature map and the original feature matrix.

3.3 Pilot Design Based on Concrete AE Network

Concrete AE will look for a feature subset that effectively identifies the largest amount of channel information. There is a selection layer in this network, which has I neurons

to select input features. Specifically, a noisy time-frequency grid h_n is selected as the noise channel input, and the grid size is $N_C \times N_S$. Flatten h_n to get $h_n = h_1, h_2 \ldots h_j$, where $J = N_C \times N_S$ is the length of the vector. The selection layer obtains the result by selecting features, that is, $h_{p,I} = h_n m_I$. Where $h_{p,I} = [h_{p,1}, h_{p,2} \ldots h_{p,I}]$ is the most informative feature subset and m_I is a J-dimensional random variable sampled from a concrete distribution, the elements are defined as:

$$m_I = \frac{\exp((\log \alpha_l + g_l)/T)}{\sum_{j=1}^{J} \exp((\log \alpha_j + g_j)/T)} \tag{6}$$

In the above formula, $\alpha_l \in \mathbb{R}^{J}_{>0}$ is the concrete parameter, $T \in (0, \infty)$ is the temperature parameter, g_l is sampled from a Gumbel distribution.

When T is close to 0, the concrete random variable will approach a discrete distribution, and the output vector ml will be approximated as a one-hot vector in probability $\alpha_l / \sum_p \alpha_p$ (only m_I the remaining elements are zero).

4 Simulation Results

Consider the downlink scenario in the SISO system. The deployed channel models are extended pedestrian a model (EPA), extended vehicle a model (EVA) and extended typical urban model (ETU) in 3GPP. It is worth mentioning that all simulations were performed under the same conditions. The settings of the OFDM system are shown in Table 1.

Table 1. Baseband Parameters

Parameter	Value
Pilot Subcarriers	24
Total number of subcarriers	72
Pilot Symbols	2
Number of OFDM symbols per slot	14
CP length	16
Bandwidth	1.08 MHz
Carrier Frequency	2.1 GHz
Subcarrier Spacing	15 kHz
Frames per slot	1

4.1 Model Training

Training data is collected from the EPA channel model. A total of 100,000 samples were collected, and 20,000 data were collected every 5 dB from a signal-to-noise ratio (SNR)

ranging from 0 dB to 20 dB (maximum Doppler shift from 0 Hz to 97 Hz, corresponding to 0 km/h ~ 50 km/h). The training set occupy 95% of the total data set, and the remaining 5% of the data set is validation data set. The training parameters are shown in Table 2.

Table 2. Train Parameters

Parameter	Value
Optimizer	Adam
Maximum epoch	100
Initial learning rate	0.001
Loss function	Mean Squared Error (MSE)
Drop period for learning rate	20
Drop factor for learning rate	0.5
Minibatch size	128
L2 regularization	0.001

We compare the MSE of SE-ResNet network by setting different number of convolution kernels (2, 4, 6, 8, 10) of convolutional layers. The network performance of different convolution kernels is shown in Fig. 4. Considering the space limitation of this paper and the increased computational cost of convolution kernels, we select 8 convolution kernels for subsequent experiments.

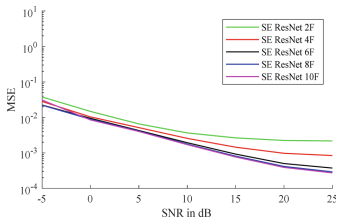

Fig. 4. SE-ResNet performance.

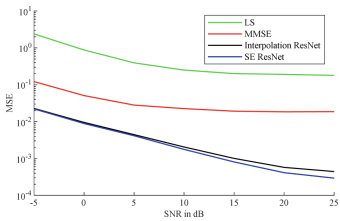

Fig. 5. MSE of estimation for the EPA channel.

4.2 MSE Performance

This paper uses MSE to evaluate the error between the estimated channel and the real channel. To verify that the proposed model is effective under extended range, the SNR is extended from –5 dB to 25 dB in test dataset, and the maximum Doppler shift is still from 0 Hz to 97 Hz. 5000 channel data are collected every 5 dB to form a test data set with a total capacity of 35000.

Figure 5 shows the MSE results of the traditional method, Interpolation-ResNet and SE-ResNet tested under the EPA channel model. It can be observed that Interpolation-ResNet and SE-ResNet have significant advantages over traditional methods. SE-ResNet

does not have much improvement over Interpolation-ResNet at lower SNR but has better performance at high SNR, which shows that the proposed model with the SE module can learn the characteristics of channel coefficients with high SNR more easily.

4.3 Generalization Performance

The EPA data described above is used to train the network model as well. The test data set of the network is collected using EPA, EVA and ETU models respectively. Each channel model collects SNR from –20 dB to 25 dB, maximum doppler shift from 0 Hz to 97 Hz, and a total of 50000 (5000 × 10) channel data for testing. The results are shown in Fig. 6.

Figure 6 shows the result that the EPA-trained network being tested on different channels. It can be seen that when the test channel and trained channel are the same, both SE-ResNet and Interpolation-ResNet can have higher performance, while SE-ResNet is superior. But when testing EVA and ETU channels, SE-ResNet will perform better in low SNR range of –20 dB ~5 dB. This means that SE-ResNet has better performance at low SNR when testing non-training channels.

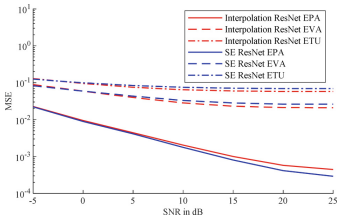

Fig. 6. MSE testing under different channel models.

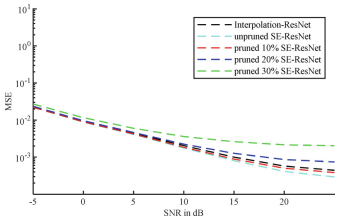

Fig. 7. Performance comparison of different crop rates.

4.4 Cropped Neural Network

Reducing network weight parameters within acceptable loss ranges can be used to reduce SE-ResNet redundancy calculations. Specifically, we try to sort weights by numerical value, and reset some of the weights to 0 according to a certain proportion, there are three cropping results shown in Fig. 7. SE-ResNet performance degrades significantly, even worse than Interpolation-ResNet, when clipping rate is above 10%. On the contrary, when the cropping rate is below 10%, the performance of SE-ResNet is still better than Interpolation-ResNet although there is a certain loss. This proves that SE-ResNet is a low-complexity network.

4.5 Performance of Concrete AE

In order to clearly reflect the performance improvement of SE-ResNet by pilot design of Concrete AE, we adopt two schemes for simulation. In the first scheme, we use 48 pilots and uniform distribution pilot design method, while the second scheme uses 8

pilots and Concrete AE pilot design method, the rest of the conditions are completely identical. Figure 8 shows the performance of the above two schemes, when SNR is less than 19, the Concrete AE cascaded SE-ResNet still shows extremely good performance although it has only 8 pilots. While SE-ResNet has better performance when SNR is greater than 19, it has higher complexity and computational cost.

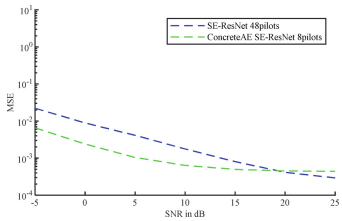

Fig. 8. Performance of concrete AE.

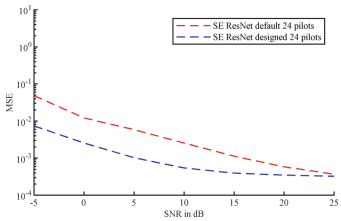

Fig. 9. Comparison of network performance under different SNR.

To comparatively highlight the power of concrete AE, we simultaneously take 24 pilots to train SE-ResNet and SE-ResNet cascaded with concrete AE. Taking the same channel conditions, the training results are shown in Fig. 9. Although the number of pilots is different, Fig. 9 shows that SE-ResNet cascade concrete AE has better performance than single SE-ResNet. This subsection proves that concrete AE can find the most informative points of channel information in the resource grid, and neural networks assisted by it can achieve better results in CE.

5 Conclusion

In this paper, ResNet with a SE block learns the distribution features of channels with weights, thereby improving the accuracy of CE. The traditional method, Interpolation-ResNet and SE-ResNet are compared, and the results show that SE-ResNet outperforms the other approaches in MSE. Regarding generalization performance, SE-ResNet has an advantage at low SNR. To further enhance the quality of CE, the position with the most channel information is selected as the pilot position by Concrete AE and used to replace the traditional pilot schemes in this paper. The simulation experiments prove that SE-ResNet cascaded with concrete AE has better performance than single SE-ResNet. In the future, we will consider using more channel model data to train SE-ResNet, so that the generalization ability of the model becomes better in the high SNR range.

References

1. Xiaohu, Y., Zhiwen, P., Xiqi, G., ShuMin, C.A.O., HeQuan, W.U.: The 5G mobile communication: the development trends and its emerging key techniques. Science China Inf. Sci. **44**(5), 551–553 (2014)
2. Hao, Y., Geoffrey Ye, L., Biing-Hwang, J.: Power of deep learning for channel estimation and signal detection in OFDM systems. IEEE Wireless Commun. Lett. **7**(1), 114–117 (2017)

3. Yang, Y., Gao, F., Ma, Xiaoli, M., Shun, Z.: Deep Learning-Based Channel Estimation for Doubly Selective Fading Channels. IEEE Access 7, 36579–36589 (2019)
4. Shen, G., Peihao, D., Zhiwen, P., Geoffrey Ye, L.: Deep Learning based Channel Estimation for Massive MIMO with Mixed-Resolution ADCs. IEEE Commun. Lett. **23**(11), 1989–1993 (2019)
5. Gao, J., Hu, M., Zhong, C., Li, G.Y., Zhang, Z.: An attention-aided deep learning framework for massive mimo channel estimation. IEEE Trans. Wireless Commun. **21**(3), 1823–1835 (2022)
6. Soltani, M., Pourahmadi, V., Mirzaei, A., Sheikhzadeh, H.: Deep learning-based channel estimation. IEEE Commun. Lett. **23**(4), 652–655 (2019)
7. Li, L., Chen, H., Chang, H.-H., Liu, L.: Deep residual learning meets ofdm channel estimation. IEEE Wireless Commun. Lett. **9**(5), 615–618 (2020)
8. Luan, D., Thompson, J.: Low complexity channel estimation with neural network solutions. In: WSA: 25th International ITG Workshop on Smart Antennas, pp. 1–6. VDE, French Riviera, France (2022)
9. Soltani, M., Pourahmadi, V., Sheikhzadeh, H.: Pilot pattern design for deep learning-based channel estimation in OFDM systems. IEEE Wireless Commun. Lett. **9**(12), 2173–2176 (2020)
10. Kang, X.-F., Liu, Z.-H., Yao, M.: Deep learning for joint pilot design and channel estimation in MIMO-OFDM systems. Sensors **22**(11), 4188 (2022)