2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW) | 978-1-7281-9507-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/WCNCW49093.2021.9420022

Knowledge-Driven Machine Learning-based Channel Estimation in Massive MIMO System

Daofeng Li*, YaMei Xu*, Ming Zhao*, Sihai Zhang*, Jinkang Zhu†
*Key Laboratory of Wireless-Optical Communications,
University of Science & Technology of China, Hefei, Anhui, P.R. China
† PCNSS, University of Science & Technology of China, Hefei, Anhui, P.R. China
Email: {df007,yamei}@mail.ustc.edu.cn, {zhaoming, shzhang, jkzhu}@ustc.edu.cn

***Abstract*—Accurate channel state information is critical for the massive multiple-input multiple-output(MIMO) system. However, due to the fundamental issue of pilot contamination, most existing works have to use additional channel information or complex machine learning network to design channel estimator. In this paper, we propose a knowledge-driven machine learning (KDML)-based channel estimator, which has a simple network structure and low training overhead. Firstly, we analyze the channel estimation problem under both uncorrelated and correlated channel models and convert it into the well-studied image denoising problem. Then, without any prior information about channel, we take advantage of the traditional channel estimation algorithms to construct the knowledge module in KDML, which maps the input data into object space without any training. After that, we use the denoising convolutional neural network (DnCNN) to build the learning module in KDML, where the residual learning accelerates the training process neural network. Finally, we obtain the estimation of the latent noise from the polluted channel matrix and then calculate the output of the proposed channel estimator. Simulation results demonstrate that the proposed channel estimator outperforms the traditional channel estimator in terms of the normalized mean-squared error (NMSE) without any statistical information about the channels. Besides, the proposed estimator also significantly better than the conventional machine learning-based estimator under the same network depth.**

***Index Terms*—Massive MIMO, Channel estimation, Knowledge-Driven, Machine learning, DnCNN**

I. Introduction

The next-generation wireless communication system is one of the cornerstones to realize a more convenient lifestyle and emerging technologies, like AR and VR. Massive MIMO is supposed to be a promising technology due to its high spectrum efficiency and power efficiency [1]. Accurate channel state information is an essential condition both for precoding transmitted signals and achieving the aforementioned advantages. However, due to the pilot contamination, it is still quite challenging to design a channel estimator with high accuracy and low complexity [2].

Nowadays, there are mainly two categories of methods to design channel estimator. One solves it by mathematics and statistics based on some statistical properties and assumptions of the wireless channel. Least square (LS) and minimum mean square error (MMSE) pilot-based channel estimator are representatives of this category of methods. The LS estimator does not rely on any prior information about the channel, thus it has low complexity but inferior performance than the MMSE estimator [3]. Although the MMSE estimator has better performance, it suffers from complex matrix operations, such as inversion and singular value decomposition. [4] proposes a modified MMSE estimator that uses partially decoded data to mitigate the pilot contamination between reused pilot sequences and achieves better performance. Based on it, [5] further points out that data-aided MMSE estimator can work without the perfect knowledge of inter-cell large scale fading, but will cause a little performance loss. Besides, to reduce the complexity, the channel covariance matrices is replaced by a sample covariance matrix obtained by a transformation of the LS estimator in [6]. Also, the second-order statistical information about user channels helps estimate channel as shown in [7]. Especially, it also points out that, in the large-number-of-antennas regime, the pilot contamination can vanish completely under certain conditions. But all the above estimator are doom to use some prior knowledge about the channel or additional user data, which will occupy a large amount of overhead and resource in practice.

The other category of method designs the channel estimator with machine learning. Luo et al. [8] propose an efficient online channel state information scheme by combing a convolutional neural network (CNN) with a long short term memory network (LSTM). And they further design an offline-online two-step training mechanism to make the prediction more stable. In [9], a spatial-frequency-temporal CNN (SFT-CNN) based channel estimator is proposed, which exploits the temporal correlation in time-varying channels to improve the accuracy of channel estimation. To perform channel estimation with unknown noise, [10] propose a channel estimator based the modified convolutional blind denoising network (CBDNet). The estimator can adjust the estimated noise level map to interactively reduce the noise in the channel matrix. In [11], an end-to-end deep neural network is proposed to design the pilot sequence and channel estimator at the same time, which outperforms the state-of-art compressive sensing approaches. However, most machine learning-based methods completely ignore domain knowledge and algorithms in wireless communication. As a result, neural networks often have complex structures and require a lot of data for training.

All the above works motivate us to propose the knowledge-driven machine learning model (KDML), which combines the knowledge processing with the machine learning to simplify network structures and reduce training overhead for a machine learning-based estimator. We then design a channel estimator with the denoising convolutional neural network (DnCNN) [12] in the Massive MIMO system based on KDML. Specifically, to simplify the neural network structures, we construct a knowledge module, which can reduce input feature dimensions of the neural network, based on the LS algorithms. Besides, it also converts the channel estimation into an image denoising problem. Then, we design the learning module of KDML from the view of image denoising rather than the general nonlinear mapping. And the residual learning [13] further accelerates the training process of the proposed estimator.

The major contributions of this paper are as follows:

- We propose the knowledge-driven machine learning model as a general approach to design a channel estimator without any statistical characteristics about the wireless channel. The domain knowledge is utilized to reconstruct the learning tasks to simplify the network structures and reduce training overhead.
- We design a KDML-based channel estimator with simple neural network structure and low training cost in Massive MIMO system by the DnCNN network. Simulation experiments prove that the proposed estimator outperforms traditional LS method and conventional machine learning-based estimator.

II. System Model

In this section, we introduce settings of the massive MIMO system firstly. Then, we present the wireless channel model used in the following sections and the channel estimation with pilot contamination in this scenario.

A. Massive MIMO System

A massive MIMO system with L time-synchronized cells is considered, which is shown in Fig. 1. Time synchronization makes sure that pilots from any cell is correlated, therefore it can be the biggest challenge for the channel estimation. The cellular network works in time-division duplexing (TDD) mode with full spectrum reuse. Hence, the downlink channel matrix can be obtained at the base station (BS) by transposing the uplink channel matrix, which is regarded as the channel reciprocity. Each BS is equipped with M antennas and each cell contains K single-antenna users synchronously communicated with the BS. All BS will estimate the channel state information of its K users by pilots during a coherence time interval. For one cell, it is easy to avoid intra-cell interference by making pilot sequences orthogonal. But for all L cell, the above approach will lend the channel estimation can not be accomplished in a coherence time interval. Thus, it is necessary to reuse pilot sequences in different cells and this eventually leads to pilot contamination.

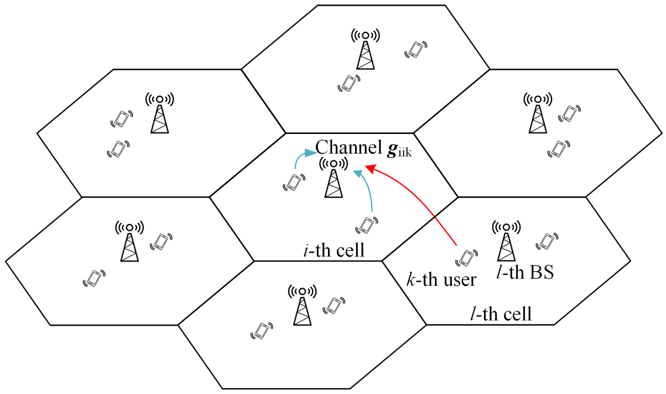

Fig. 1: Uplink transmission in massive MIMO systems.

B. Wireless Channel Model

Since the channel estimation will be executed in a coherence time interval, therefore we use the block fading model to describe the wireless channel in this paper. Under the model, the channel state information keeps constant during any coherence time interval and the channel information of any two intervals is independently and identically distributed (i.i.d). As for the channel in a interval, we consider both correlated and uncorrelated Rayleigh fading model.

Based on above assumptions, the $M \times 1$ channel vector generated from the k-th user in the l-th cell to the i-th BS is denoted as Eq.(1):

$$\boldsymbol{g}_{ilk} = [g_{ilk1}, g_{ilk2}, \cdots, g_{ilkM}]^T \sim \mathcal{CN}(\mathbf{0}_M, \boldsymbol{R}_{ilk}) \tag{1}$$

where $\boldsymbol{R}_{ilk} \in \mathcal{C}^{M \times M}$ is the positive semi-definite channel covariance matrix. That is to say, the small-scale fading is modeled as the Gaussian distribution and the average large-scale fading between k-th user and the i-th BS is determined by the normalized trace of $\boldsymbol{R}_{ilk}$ as $\beta_{ilk} = \frac{1}{M} tr(\boldsymbol{R}_{ilk})$. And the eigenvalues of $\boldsymbol{R}_{ilk}$ describe the spatial channel correlation.

For the correlated Rayleigh fading model, we adopt the Exponential correlation model [14] to describe spatial correlation, which is given as Eq.(2).

$$r_{ij} = \begin{cases} r^{j-i}, i \leq j \\ r^*_{ji}, i \geq j \end{cases}, |r| \leq 1 \tag{2}$$

And the correlation factor r is set as 0.5 in the rest of this paper. For the uncorrelated Rayleigh fading model, we let the channel covariance matrix $\boldsymbol{R}_{ilk} = \beta_{ilk} \boldsymbol{I}_M$. So that the elements of $\boldsymbol{g}_{ilk}$ are i.i.d circularly-symmetric complex normal variables for all i, l and k.

C. Channel Estimation and Pilot Contamination

We assume that all BSs and users are perfectly synchronized and before data transmission phase an uplink pilot phase for channel estimation is executed. To make channel estimation as difficult as possible, we also let all L cell reuse the same pilot sequences. The sequence assigned to the k-th user in any cell is denoted as Eq.3.

$$\phi_k = [\phi_{k_1}, \phi_{k_2}, \cdots, \phi_{k_\tau}]^T \tag{3}$$

where τ is the length of the pilot sequence. Without loss of generality we normalize the average power of pilot symbols:

$$\phi_{k_1}^H \phi_{k_2} = \begin{cases} 0, k_1 \neq k_2 \\ 1, k_1 = k_2 \end{cases} \tag{4}$$

Extending it to the $K \times \tau$ matrix containing all pilots, $\Phi = [\phi_1, \phi_2, \cdots, \phi_K]^T$ and $\Phi^H \Phi = \boldsymbol{I}_M$. Then, the received uplink pilot sequence at the $i-$th BS can be represented as a $M \times N$ matrix, which is given as Eq.(5).

$$\boldsymbol{Y}_i = \sqrt{p} \sum_{l=1}^{L} \boldsymbol{G}_{il} \boldsymbol{\Phi} + \boldsymbol{N}_i \tag{5}$$

where $\boldsymbol{G}_{il} = [\boldsymbol{g}_{il1}, \boldsymbol{g}_{il2}, \cdots, \boldsymbol{g}_{ilK}]$. p is the transmitted power or average signal to noise ratio (SNR) and $\boldsymbol{N}_i \in \mathcal{C}^{M \times N}$ stands for the additive white Gaussian noise (AWGN) with i.i.d elements following $\mathcal{CN}(0,1)$.

We briefly introduce the LS channel estimator in our scenario. Without loss of generality, we set the i-th cell as the target cell, so that $\boldsymbol{G}_{ii}$ is the desired channel while $\boldsymbol{G}_{il}, \forall l \neq i$ are the interference channels. Then, from the view of channel vector, $\boldsymbol{g}_{iik}$, at the i-th BS, LS channel estimation is a sufficient statistic, which is given as Eq.(6).

$$\hat{\boldsymbol{g}}_{iik}^{LS} = \frac{1}{\sqrt{p}} \boldsymbol{Y}_i \phi_k = \boldsymbol{g}_{iik} + \sum_{l=1, l \neq i}^{L} \boldsymbol{g}_{ilk} + \boldsymbol{z}_{ik} \tag{6}$$

where $\boldsymbol{z}_{ik} = (1/\sqrt{p}) \boldsymbol{N}_i \phi_k$ and $\hat{\boldsymbol{g}}_{iik}^{LS} \sim \mathcal{CN}(\boldsymbol{0}_M, \boldsymbol{W}_{ik})$ while $\boldsymbol{W}_{ik} = \sum_{l=1}^{L} \boldsymbol{R}_{ilk} + (1/p) \boldsymbol{I}_M$. Given by (6), the estimation error vector can be calculated as $\overline{\boldsymbol{g}}_{iik} = \hat{\boldsymbol{g}}_{iik}^{LS} - \boldsymbol{g}_{iik} \sim \mathcal{CN}(\boldsymbol{0}_M, \boldsymbol{W}_{ik} - \boldsymbol{R}_{iik})$. Therefore, the mean square error(MSE) per antenna of the LS estimator is given as (7).

$$\lambda_{ik}^{LS} = \frac{1}{M} \mathbb{E}\{ \| \hat{\boldsymbol{g}}_{iik}^{LS} - \boldsymbol{g}_{iik} \|^2 \} = \frac{1}{M} tr \left[\boldsymbol{W}_{ik} - \boldsymbol{R}_{iik} \right] \tag{7}$$

Similarly, for the uncorrelated channel model, the estimation error is given as $\lambda_{ik}^{LS} = \varepsilon_{ik} - \beta_{iik}$, where $\varepsilon_{ik} = (\sum_{l=1}^{L}) \beta_{ilk} + (1/p) \boldsymbol{I}_M$. It is obvious that the dominating factor of estimation error is the interference from other cells, which is called as the pilot contamination. Due to the pilot contamination,

$$\lambda_{ik}^{LS} \to (1/M) tr \left[\sum\nolimits_{l=1, l \neq i}^{L} \boldsymbol{R}_{ilk} \right], as\ p \to \infty \tag{8}$$

That is to say, even there is lack of AWGN, the performance of LS channel estimator is still subjected to the pilot contamination. Besides, due to time-synchronized, any BS can not separate users who use the same pilot sequence. Thus, it is necessary to rise new algorithm to solve the channel estimation problem in this scenario.

III. KDML-based Channel Estimation

The framework of the proposed KDML-based channel estimator is presented in Fig. 2. From the view of KDML, once we get the received pilot sequences, there are two key modules including knowledge module and learning module to construct the channel estimator.

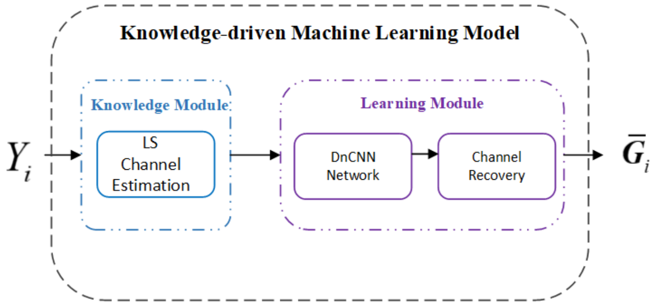

Fig. 2: The framework of the proposed method

A. Knowledge-Driven Machine Learning Model

Due to the high non-linear fitting ability and successful applications in the field of image and speech, deep learning tends to be emerging research hot spots in the field of channel estimation. However, most of these estimators ignore well-established traditional algorithms, which have already been summarized as domain knowledge in wireless communication. Besides, these estimators often have complex network structures and high training overhead. As a result, these estimators only have limited interpretability and reliability.

Meanwhile, the basic idea of KDML is to use domain knowledge to rebuild learning tasks. As mentioned above, compared with deep learning, KDML has a unique knowledge module which accounts for mining the commonalities over dataset based on domain knowledge. More specifically, for a certain optimization problem, it maps all instances of data from input space directly into output space without any training. Thus, the neural network structure of learning module will be simplified and the training costs will be reduced correspondingly. After the knowledge module, the function of learning module is similar with ordinary deep learning network. Thus, the objective function of KDML is given as:

$$\arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathbb{E}\{ \| \boldsymbol{y} - f(\boldsymbol{k}(\boldsymbol{x}), \boldsymbol{\theta}) \|^2 \} \tag{9}$$

where $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{k}$ stands for the labels, input data, neural network parameters, and knowledge module function, respectively.

For the problem of channel estimation in massive MIMO system, there is no doubt that the input dataset is consisted of received pilot sequences from different cells with different SNR. Accordingly, the labels are made of channel state information, which can be presented a $M \times K$ matrix. Different from other works, we choose the estimation of channel state information rather than the real channel, which is more in line with the practical situation. As for the knowledge module, we adopt the mentioned LS estimator given as (6). By the way, we will study other traditional estimators like MMSE in the future. Thereafter, the output of knowledge module will be $\hat{\boldsymbol{G}}_{ii}^{LS} \in \mathcal{C}^{M \times K}$. It worth noting that $K \leq \tau$, so the size of input data for the learning module is always smaller than the size of pilot sequences.

B. DnCNN Network Structure

Considering that $\hat{G}_{ii}^{LS}$ is a two-dimensional matrix, we regard it as an image of the wireless channel environment. Thus, we design the learning module from the view of deep learning-based image denoising algorithms. Apart from the advantages of KDML itself, we think that the neural network should be easy to achieve well-trained. Based on the above considerations, we decide to use DnCNN network, whose structure is shown in Fig. 3, to construct the learning module.

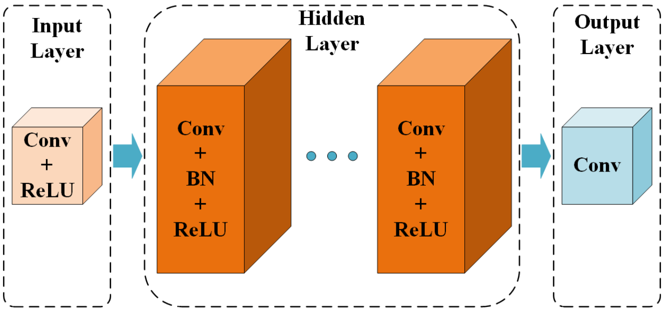

Fig. 3: Network Structure of the DnCNN

Conventional discriminative denoising networks, like MLP and CSF [15], try to map a noisy image to the latent clean image. But DnCNN takes advantage of residual learning formulation to accelerate the training process. Formally, the objective function of it describes the MSE between the desired residual(noise) images and predicted ones, which is given as

$$\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{E}\{\|f(\boldsymbol{k}(\boldsymbol{x}),\boldsymbol{\theta}) - (\boldsymbol{y}-\boldsymbol{x})\|^2\} \tag{10}$$

And based on (10), all parameters will be trained and updated by the optimization algorithm, Adam [16].

It is not difficult to find that there are types of layers in the DnCNN network with D layers. The first type is the input layer, which is consisted of a convolution layer and rectified linear units($ReLU$). The convolution layer contains 64 filters whose size is $3\times3\times c$. Here, c stands for the number of input image channels, such as for color image $c = 3$. Although the $\hat{G}_{ii}^{LS}$ is supposed to be a gray wireless channel image, nowadays neural network only accepts real numbers. Therefore, we split it into real and imaginary two parts and set $c = 2$ in the rest of this paper. And $ReLU(x) = max(0, x)$ is a kind of non-linear activation function. Layer 2 through $D-1$ share the second type. Compared with the first type, it adds a new structure called batch normalization(BN), which can further accelerate network training. Lastly, the output layer is made up of a convolution layer with c filters of size $3\times3\times64$. Formally, let f_d^t denote the function of the d layer with type t and $*$ denote the convolution operator. Then the mathematical representation of the DnCNN is given as (11).

$$\boldsymbol{y} = f_D^3(f_{D-1}^2(f_{D-2}^2(\cdots f_1^1(\boldsymbol{x})))) \tag{11}$$

where

$$\begin{cases} f_d^1 = ReLU(\boldsymbol{\theta}_d * \boldsymbol{x}) \\ f_d^2 = ReLU(BN(\boldsymbol{\theta}_d * \boldsymbol{x}_{d-1})) \\ f_d^3 = (\boldsymbol{\theta}_D * \boldsymbol{x}_{D-1}) \end{cases} \tag{12}$$

C. Channel Recovery

Based on the DnCNN network, we can only get an image that describes latent noise information rather than the channel matrix itself. Let $\boldsymbol{\omega}$ denote the output of DnCNN. Then we have

$$\hat{\boldsymbol{G}}_{ii}^{LS} = \bar{\boldsymbol{G}}_{ii} + \boldsymbol{\omega}_{ii} = \bar{\boldsymbol{G}}_{ii} + \mathcal{R}(\hat{\boldsymbol{G}}_{ii}^{LS}) \tag{13}$$

where $\bar{\boldsymbol{G}}_{ii}$ is the desired output of the proposed channel estimator, and $\mathcal{R}$ stands for the residual learning function. Thereby, the depth of DnCNN will not cause the performance degradation problem, i.e. the accuracy will decrease along with the increasing of network depth. Besides, given (13), we can regard the channel recovery module as a skip connection go through the whole network, which significantly speeds up training as well as improves the performance.

Then, there are two essential steps to recover the channel. As mentioned above, we adopt the estimations rather than the real channel as labels and split the input wireless channel image into two parts. Thus, we pair the output of DnCNN and restore it as a new complex value matrix at first. Finally, we subtract that matrix from the corresponding channel estimation and get the final output of KDML-based channel estimator.

IV. Performance Evaluation

A. Experiments Setup

In the following experiments, the proposed channel estimator will be compared with the traditional LS and MMSE channel estimator with both correlated and uncorrelated channel model. Besides, we train an MLP network as the baseline of the machine learning-based channel estimator. And we also try to replace DnCNN with a novel network called FFDNet [17] to analyze the effect of different network structures. The performance metric is the normalized channel estimation error

$$NMSE = 10\log_{10}(\mathbb{E}\{\frac{\sum_{k=1}^{K}\|\bar{\boldsymbol{g}}_{iik} - \boldsymbol{g}_{iik}\|^2}{\sum_{k=1}^{K}\|\boldsymbol{g}_{iik}\|^2}\}) \tag{14}$$

The major parameters of massive MIMO system and DnCNN are listed in TABLE I and TABLE II, respectively. For the large scale fading β_{ilk}, we fix its value and set $\beta_{iik} = 1$, $\beta_{ilk} = \beta(0 < \beta < 1), \forall l \neq i$. The received pilot sequences dataset has a training set of 27,000 examples and a testing set of 3,000 examples, a total of 30,000 examples in 5 types which contains data with a certain SNR. It is worth noting that both epoch and the number of layers are far smaller than other works. When testing different values of these two parameters, performance degradation happens if the depth of DnCNN is larger than 5. This phenomenon proves that KDML indeed simplifies the network structure and reduce the training cost.

TABLE I: Communication system parameters

Number of Cells	7				
Number of Users per Cell	10				
Number of Antennas	30	60	90	120	150
Length of Pilot	16				
Large scale fading(β)	0.1	0.2	0.3	0.4	0.5
Traditional Algorithm	LS, MMSE				
SNR	0	5	10	15	20

TABLE II: Learning module parameters

Number of Layers	5
Number of Feature Maps	64
Optimizer	Adam
Loss function	MSE
Learning rate	0.01 & 0.001
Batch size	500
Padding	1
Kernel size	3×3
Training dataset size	25000
Test dateset size	5000
Epoch	50

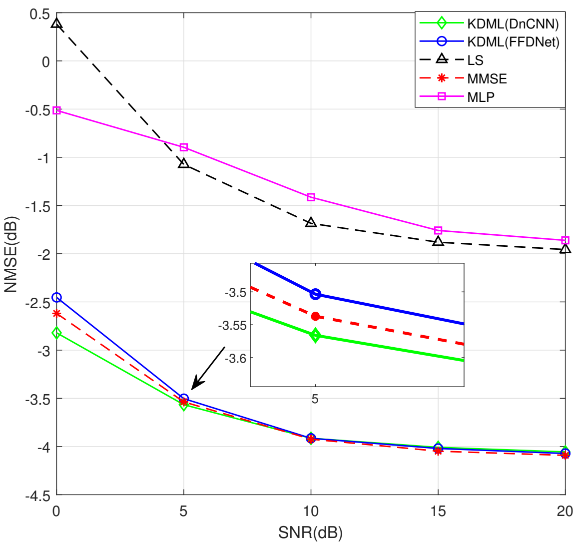

Fig. 4: NMSE performance comparison of the proposed method and other methods versus transmitted SNR.

B. NMSE Performance Analysis Versus SNR

Fig. 4 depicts the NMSE performance of all five methods versus SNR with the correlated channel model and $\beta = 0.1, M = 30$. Firstly, we directly use the perfect channel correlation matrix, which is unavailable in practice, for calculating the NMSE performance of the MMSE channel estimator. Therefore, it is natural that MMSE outperforms the LS estimator. Meanwhile, as SNR> 5dB, the NMSE performance of the proposed estimator achieves the same as that of MMSE. It is a remarkable fact that the proposed estimator is based on the LS estimator but does not have any statistical channel information other than the received pilot sequence. Further, when SNR $<$ 5dB, the proposed estimator even outperforms MMSE. This phenomenon implies that the learning module of KDML can effectively extract and utilize channel correlations.

Secondly, we try to prove the advantage of the proposed methods against other machine learning-based estimators. For the sake of fairness, the MLP network we used also has 5 layer. Considering it is more suitable for handling vectors rather than matrix, we vectorize the channel matrix and increase the training dataset size of it by five times. Although we have tried a variety of parameter settings, its performance is still the worst. Therefore, we have reason to believe that when the network depth of pure machine learning-based methods is greatly restricted, it cannot effectively solve the channel estimation in our scenario.

Besides, we also use FFDNet, which can be regarded as an advanced version of DnCNN neural network, to construct a new KDML-based estimator. The initialization parameters of the FFDNet network are almost the same as that of DnCNN. As shown in Fig.4, there is almost no performance difference between these two KDML-based channel estimators. That is to say, without performance loss, we can use KDML model to simplify structures of different deep learning networks. Actually, we have also confirmed the performances with uncorrelated channel model. However, it does not significantly differentiate from that of correlated channel model. Thus, we no longer analyze the simulation results in this scenario in detail.

C. NMSE Performance Analysis Versus β

Fig. 5 presents the NMSE performance of all four methods versus β with SNR = 5dB, $M = 30$. Here, β not stands for the large scale fading of interfering cells, but also represents the degree of pilot contamination to some extent. As we can see, the NMSE performances of all four estimators get worse with the increase of β. And the performance loss of LS is more evident than other estimators due to its simple algorithm design. Interestingly, the performance variation trend of the proposed estimator is basically the same as MMSE, which further verifies the generalization ability of the proposed estimator.

D. NMSE Performance Analysis Versus Number of Antennas

Fig. 6 exhibits the NMSE performance comparison of all four estimators versus M with $L = 7, \beta = 0.05$, SNR=5dB. First of all, as M increases, the NMSE performances of all estimators become better. It is reasonable due to the characteristics of the Massive MIMO system. As expected, the LS estimator results in the highest NMSE value, However, for the lowest value, it is surprising to find that the proposed DnCNN-based estimator has a little improvement than MMSE. It is necessary to highlight that the MMSE estimator obtains the perfect channel correlation matrix in advance. Meanwhile, the

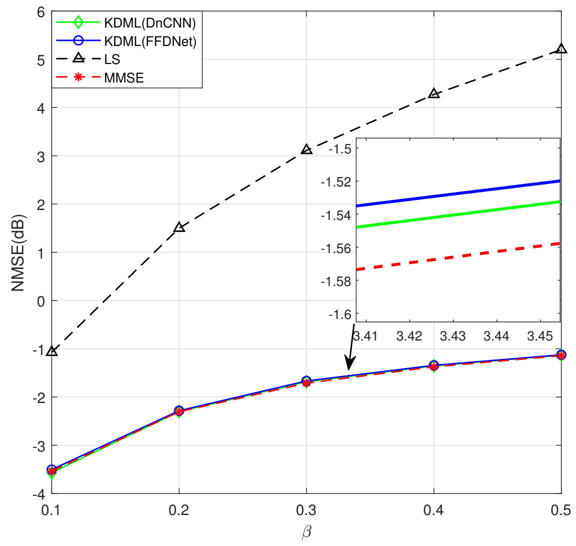

Fig. 5: NMSE performance comparison versus β.

proposed estimator knows nothing about the channel. This phenomenon indicates that the proposed estimator does not use the neural network as a new implementation of the MMSE estimator, but a new machine learning-based method with higher performance and more potential.

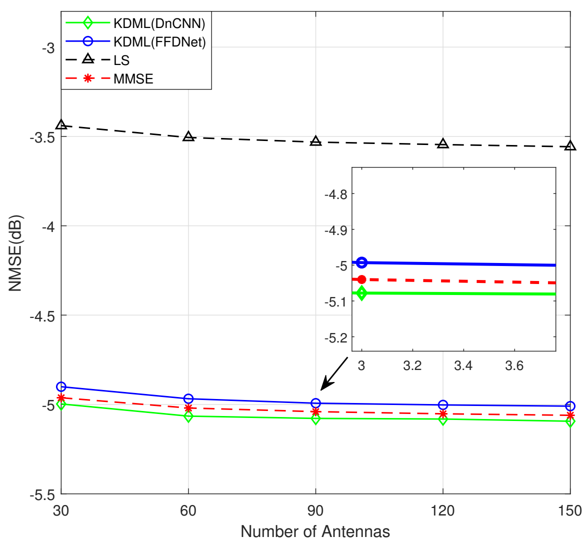

Fig. 6: NMSE performance comparison versus number of antennas.

V. Conclusion

In this paper, we proposed a channel estimator based on the knowledge-driven machine learning model for the massive MIMO system. We establish a typical massive MIMO scenario with an uncorrelated and correlated block fading channel model. And then, we regard the channel estimation matrix as an image of the wireless environment and took advantage of the well-known denoising deep learning network, DnCNN, to construct the learning module of KDML. Simulation results exhibit that the proposed estimator can achieve better NMSE performance compared with conventional machine learning-based estimators and also outperforms than the traditional channel estimators.

Acknowledgment

This work was partially supported by Key Program of Natural Science Foundation of China under Grant(61631018), Huawei Technology Innovative Research(YBN2018095087).

References

[1] E. Bjornson, J. Hoydis, and L. Sanguinetti, “Massive mimo has unlimited capacity,” *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2017.

[2] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive mimo: Benefits and challenges,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[3] F. A. P. D. Figueiredo, F. A. C. M. Cardoso, I. Moerman, and G. Fraidenraich, “Channel estimation for massive mimo tdd systems assuming pilot contamination and flat fading,” *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2018.

[4] M. Ju, L. Xu, L. Jin, and D. Defeng Huang, “Data aided channel estimation for massive mimo with pilot contamination,” in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[5] M. Ju, L. Xu, L. Jin, and D. D. Huang, “Data aided channel estimation for massive mimo with pilot contamination,” in *ICC 2017 - 2017 IEEE International Conference on Communications*, 2017.

[6] F. P. D. Figueiredo, D. A. M. Lemes, C. F. Dias, and G. Fraidenraich, “Massive mimo channel estimation considering pilot contamination and spatially correlated channels,” *Electronics Letters*, 2020.

[7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, “A coordinated approach to channel estimation in large-scale multiple-antenna systems,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, 2012.

[8] Changqing, Luo, Jinlong, Ji, Qianlong, Wang, Xuhui, Chen, Pan, and Li, “Channel state information prediction for 5g wireless communications: A deep learning approach,” *IEEE Transactions on Network ence and Engineering*, vol. 7, no. 1, pp. 227–236, 2018.

[9] P. Dong, H. Zhang, G. Y. Li, N. Naderializadeh, and I. S. Gaspar, “Deep cnn based channel estimation for mmwave massive mimo systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 989–1000, 2019.

[10] Y. Jin, J. Zhang, B. Ai, and X. Zhang, “Channel estimation for mmwave massive mimo with convolutional blind denoising network,” *IEEE Communications Letters*, vol. PP, no. 99, pp. 1–1, 2019.

[11] X. Ma and Z. Gao, “Data-driven deep learning to design pilot and channel estimator for massive mimo,” *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2020.

[12] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[14] S. L. Loyka, “Channel capacity of mimo architecture using the exponential correlation matrix,” *Communications Letters IEEE*, vol. 5, no. 9, pp. 369–371, 2001.

[15] U. Schmidt and S. Roth, “Shrinkage fields for effective image restoration,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[16] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *Computer Science*, 2014.

[17] K. Zhang, W. Zuo, and L. Zhang, “Ffdnet: Toward a fast and flexible solution for cnn based image denoising,” *IEEE Transactions on Image Processing*, pp. 1–1, 2017.