# Channel Estimation for Cell-Free mmWave Massive MIMO Through Deep Learning

Yu Jin ⓘ, Jiayi Zhang ⓘ, *Member, IEEE*,
Shi Jin ⓘ, *Senior Member, IEEE*, and Bo Ai ⓘ, *Senior Member, IEEE*

*Abstract*—The combination of cell-free massive multiple-input multiple-output (MIMO) systems along with millimeter-wave (mmWave) bands is indeed one of most promising technological enablers of the envisioned wireless Gbit/s experience. However, both massive antennas at access points and large bandwidth at mmWave induce high computational complexity to exploit an accurate estimation of channel state information. Considering the sparse mmWave channel matrix as a natural image, we propose a practical and accurate channel estimation framework based on the fast and flexible denoising convolutional neural network (FFDNet). In contrast to previous deep learning based channel estimation methods, FFDNet is suitable a wide range of signal-to-noise ratio levels with a flexible noise level map as the input. More specifically, we provide a comprehensive investigation to optimize the FFDNet based channel estimator. Extensive simulation results validate that the training speed of FFDNet is faster than state-of-the-art channel estimators without sacrificing normalized mean square error performance, which makes FFDNet as an practical channel estimator for cell-free mmWave massive MIMO systems.

*Index Terms*—Channel estimation, cell-free massive MIMO, FFDNet, millimeter wave.

## I. INTRODUCTION

Fifth-generation (5G) and beyond 5G systems are expected to achieve Gbit/s data rates [1], [2]. To meet the high data rate demand, the combined use of millimeter-wave (mmWave) and massive multiple-input multiple-output (MIMO) is regarded as a pivotal and significant technology [3]. Moreover, a promising network densification concept has been recently termed as the so-called cell-free (CF) massive MIMO,

Y. Jin and J. Zhang are with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, P. R. China. They are also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, P. R. China (e-mail: 15221269@bjtu.edu.cn; jiayizhang@bjtu.edu.cn).

S. Jin is with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, P. R. China (e-mail: jinshi@seu.edu.cn).

B. Ai is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, P. R. China (e-mail: boai@bjtu.edu.cn).

Digital Object Identifier 10.1109/TVT.2019.2937543

where a large number of multiple-antenna access points (APs) are distributed over a wide area to coherently serve a small number of user equipments (UEs) in the same time-frequency resources [4]–[7]. Due to the exploitation of favorable propagation and macro-diversity gain, CF massive MIMO can use simple linear signal processing schemes to provide uniformly good services to all UEs, especially for the cell-edge UEs [8].

It is well known that the unused bandwidth at mmWave frequencies is much higher than the one at conventional sub-6 GHz frequencies. In order to further improve the system performance, there has been a growing interest in utilizing the less-congested mmWave bands in CF massive MIMO systems mainly because: First, the dimension and hardware complexity of antenna arrays at APs is substantially reduced by employing mmWave; Second, distributed path-diversity is generated to alleviate the signal blockage which is the intrinsic challenging at mmWave frequencies [9], [10]. More specifically, the authors in [9] proposed a energy-efficient power control algorithm for CF mmWave massive MIMO systems. In [10], a fronthaul quantization optimization approach has been developed to achieve the balance of fronthaul requirement and data rate.

Typically, it is an extremely challenging task to obtain accurate channel state information (CSI) for the CF mmWave massive MIMO systems, partially due to its large bandwidth. The joint channel from the APs to a UE is strongly spatially correlated due to the closed location of the APs. The simple linear channel estimator (LS) eludes the computational burden, while it cannot achieve acceptable normalized mean squared error (NMSE) for the considered system over strong spatially correlated channels. Furthermore, the classical MMSE estimator with high complexity can only obtain suboptimal estimations when using non-orthogonal pilot sequences [10]. Compared with collocated massive MIMO, the computing ability of APs in CF is limited by power consumption and cost. So it is interesting to improve the NMSE performance and to reduce the complexity at the same time.

The channel matrix in mmWave massive MIMO systems can be characterized as a sparse matrix [11]. Therefore, the compressive sensing (CS) technology can be used to iteratively recover sparse channel matrix with a simple sparsity prior [12], which has promoted establishing the framework of distributed compressive channel estimation. However, existing CS algorithms cannot estimate the accurate channel matrix. First, the channel matrix is approximately instead of perfectly sparse. Moreover, the variations in the sparse mmWave channel matrix between most adjacent elements are small. Therefore, it is very complicated to model the priors of mmWave channel matrix. Finally, CS-based algorithms perform time-consuming with iterative approaches, and may not accurately and fully exploit channel structures.

Note that the channel matrix of beamspace mmWave massive MIMO systems can be considered as a 2-dimensional (2D) image [13]. In order to further improve the channel estimation performance, deep learning (DL) has been recently used in wireless communication networks [14]–[16]. For example, a learned denoising-based approximate message passing (LDAMP) neural network, which combines denoising convolutional neural network (DnCNN) and the iterative sparse signal recovery algorithm, has been applied for the channel estimation of beamspace mmWave massive MIMO systems [15]. It is interesting to learn that the LDAMP network outperforms the most advanced compressed sensing-based algorithms in channel estimation. However, DnCNN is tailored to specific noise levels, and it works well only when

the noise level is in the trained range. This problem causes the DnCNN denoiser limited by flexibility and efficiency in the practical channel estimation.

To fill this gap, we use a fast and flexible denoising convolutional neural network (FFDNet) [17] for the channel estimation of CF mmWave massive MIMO systems. It can deal with different noise levels using the same neural network by inducing a noise level map as input. Moreover, the FFDNet model can divide an image into several sub-images to reduce the training and testing latency. To this end, we investigate the performance of employing FFDNet for the channel estimation of CF mmWave massive MIMO systems. The simulation results show that the proposed FFDNet-based channel estimator outperforms the existing DnCNN-based method by a large range of noise levels, while with a marginal loss in NMSE.

## II. SYSTEM MODEL

Let us consider the CF mmWave massive MIMO system with $M$ APs and $K$ UEs, which randomly distributes in a wide area. All UEs are simultaneously served by $M$ APs by using the same time and frequency resources. The number of antennas at each AP and UE is $N_r$ and $N_t$, respectively. Furthermore, we assume all APs are connected to a central processing unit (CPU) via infinite-capacity fronthaul links [4].

Let $\mathbf{G}_{m,k}$ denotes the channel matrix between the $m$th AP and the $k$th UE, it is modeled as

$$\mathbf{G}_{m,k} = \sqrt{\mathbf{L}(d_{m,k})}\mathbf{H}_{m,k}, \qquad (1)$$

where $\mathbf{L}$ is the distance-dependent pathloss and shadowing matrix, $d_{m,k}$ denotes the distance between the $m$th AP and the $k$th UE, $\mathbf{H}_{m,k} \in \mathbb{C}^{N_r \times N_t}$ is the small-scale fading matrix and is given as

$$\mathbf{H}_{m,k} = \sum_{l=1}^{L} \boldsymbol{\alpha}_R \left(\phi_{R,l}^{\text{azi}}, \phi_{R,l}^{\text{ele}}\right) \boldsymbol{\alpha}_T^{\text{H}} \left(\phi_{T,l}^{\text{azi}}, \phi_{T,l}^{\text{ele}}\right), \qquad (2)$$

where $L \ll \min(N_r, N_t)$ denotes the number of multipath. Moreover $\phi_{T,l}^{\text{ele}}$, $(\phi_{T,l}^{\text{azi}})$ and $\phi_{R,l}^{\text{ele}}$, $(\phi_{R,l}^{\text{azi}})$ denote the elevation (azimuth) angle of departure and arrival for the $l$th path, respectively. Additional, $\boldsymbol{\alpha}_R \left(\phi_{T,l}^{\text{azi}}, \phi_{T,l}^{\text{ele}}\right)$ and $\boldsymbol{\alpha}_R \left(\phi_{R,l}^{\text{azi}}, \phi_{R,l}^{\text{ele}}\right)$ denote the steering vectors at the transmitter and the receiver, respectively.

These steering vectors depend on the array geometry. For the typical $N_1 \times N_2$ uniform planar arrays, $\boldsymbol{\alpha}\left(\phi_{T,l}^{\text{azi}}, \phi_{T,l}^{\text{ele}}\right)$ is given by [18]

$$\boldsymbol{\alpha}(\phi^{\text{azi}}, \phi^{\text{ele}})$$
$$= [1, e^{j2\pi \text{d} \sin \phi^{\text{azi}} \sin \phi^{\text{ele}}/\lambda}, \ldots, e^{j2\pi(N_1-1)\text{d} \sin \phi^{\text{azi}} \sin \phi^{\text{ele}}/\lambda}]^T$$
$$\otimes [1, e^{j2\pi \text{d} \cos \phi^{\text{ele}}/\lambda}, \ldots, e^{j2\pi(N_2-1)\text{d} \cos \phi^{\text{ele}}/\lambda}]^T, \qquad (3)$$

where $d$ is the antenna spacing, $\lambda$ is the wavelength, and $\bigotimes$ denotes the Kronecker product.

In the uplink pilot transmission, the UEs send the pilot sequences $\mathbf{s}$ to all APs. Considering the hardware complexity and high-power consumption levels constraints at mmWave frequencies, the hybrid structures is used in this paper. So the received signal vector $\mathbf{y} \in \mathbb{R}^{Nt \times Nr}$ at the $m$th AP can be expressed as

$$\mathbf{y}_m = \mathbf{Q}^H \mathbf{H}_{m,k} \mathbf{P} \mathbf{s} + \mathbf{n}, \qquad (4)$$

where $\mathbf{y}_m \in \mathbb{C}^{N_r^{RF} \times 1}$ denotes the received signal, $\mathbf{Q} \in \mathbb{C}^{N_r \times N_r^{RF}}$ denotes the hybrid combining matrix, $\mathbf{P} \in \mathbb{C}^{N_t \times N_t^{RF}}$ denotes the hybrid precoding matrix, and $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_n^2 \mathbf{I}\right)$ denotes a Gaussian noise matrix. Furthermore, we normalize the pilot sequence as $\|\mathbf{s}\|^2 = 1$. The

independence from each AP to the typical UE is natural for spatially separated antennas. And the small-scale fading models a sparse scattering multipath propagation. Furthermore, high antenna correlation levels appear in mmWave communications due to the large tightly-packet antenna array at the APs.

Because of the sparsity characteristics of mmWave channels, the elements of $\mathbf{H}_{m,k}$ are correlated with each other. Moreover, the difference among the neighboring elements are subtle. This feature makes $\mathbf{H}_{m,k}$ very similar to a 2D natural image. Therefore, FFDNet, originated from the problem of image recovery, appears particularly attractive to exploit the channel estimation in the mmWave massive MIMO systems.

## III. FFDNET-BASED CHANNEL ESTIMATION

In the following, we introduce the FFDNet-based channel estimation approach for CF mmWave massive MIMO systems in detail. Fig. 1 illustrates the basic network architecture and executive process of FFDNet-based channel estimator.

### A. FFDNet Network Architecture

We consider the size of received channel matrix as $N_r \times N_t$. The first step is combining the real and imaginary parts of channel matrix into a larger matrix of size $N_r \times 2N_t$. In order to improve the efficiency of the denoiser, the second step is setting a reversible dividing operator which divides a noisy matrix $Y$ of size $N_r \times 2N_t$ into four sub-matrices of size $\frac{N_r}{2} \times N_t$. To make the denoising model flexible enough for different noise levels, we introduce a adjustable noise level map $\mathbf{M}$ as part of input and concatenate $\mathbf{M}$ with same size of the sub-matrices. Consequently, the size of FFDNet input tensor is $\frac{N_r}{2} \times N_t \times 5$.

The FFDNet uses multiple filters (with varying window sizes, e.g. window sizes = $3 \times 3$ in this paper) to obtain multiple features. The idea is to capture the most important feature, one with the highest value for each layer's feature map. These features form the next layer and are passed to a fully connected convolutional layer. Taking the tensor $\widetilde{Y}$ as input, a series of $3 \times 3$ convolution multi-filters layers constitute the following CNN. Among them, one layer is combined by three kinds of operations: "$Conv$", "$ReLU$" and "$BN$", which represent *Convolution, Rectified Linear Units*, and *Batch Normalization*, respectively. In addition, the first convolution layer is "$Conv+ReLU$", the middle convolution layer is "$Conv+BN+ReLU$", and the last convolution layer is "$Conv$". Furthermore, we use zero-padding to make the size of features equal after each convolution layer. The first and the last layer are seen as the input and output operation. Therefore, the analysis and simulation effects on the FFDNet layer are only for the middle layers in the following correspondence. The number of employed multiple filters for the middle layers is the same. Additional, one filter can be expressed as one feature. After the last convolution layer, we use the reverse operator of the dividing operator, applied in the input tensor with the size of $\frac{N_r}{2} \times N_t \times 4$, to produce the estimated noise matrix $\widehat{\Theta}$ with the size of $N_r \times 2N_r$.

Specifically, the FFDNet is formulated as $X = \mathcal{F}(Y, \mathbf{M}, \Theta)$, where $M$ represents the noise level map as part of input. The model parameters $\Theta$ are unchanged to different noise levels. We assume that the noise level $\sigma$ is known from the channel matrix. In order to make the training process easier, we have SNR $= \frac{|\mathbf{H}|^2}{\sigma^2}$ as a known parameter. According to the simulation results, the neural network can obtain the structure of noise matrix. In other words, the noise level $\sigma$ does not need to be very accurate during the testing due to the great robustness of the FFDNet. For additive white Gaussian noise (AWGN) with different noise levels, $\mathbf{M}$ is set to an noise map where each element being $\sigma$. Thus, FFDNet
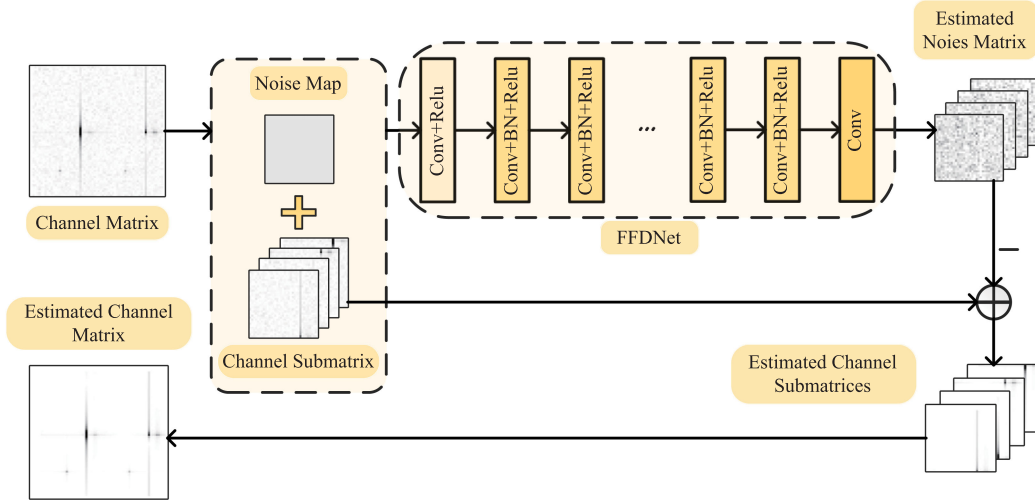
Fig. 1. Architecture of FFDNet-based channel estimator.

can provide a flexible way to deal with a extensive range of noise levels with only one network.

### B. Comparison of CNN and Residual Learning

The success of DnCNN for channel estimation inspired by image denoising is mainly due to its great capacity of modeling in network architecture and training. Generally, the aim of channel estimation is constructing a mapping relationship $X = \mathcal{F}(Y, \Theta_\sigma)$ between the received noisy matrix $Y$ and the expected output $X$. We trained the model parameters $\Theta_\sigma$ for noisy matrix because $\Theta_\sigma$ is fixed with the noise level $\sigma$. Consequently, DnCNN-based channel estimation methods are constrained in its noise flexibility, and the learned model is normally limited by the fixed noise level. Therefore, the trained model is difficult to be directly applied to channel matrices with other noise levels.

Based on the experience of the residual learning strategy in image denoising [19], the residual network is much easier to train for better performance than the original network. Then the final step is subtraction to get the residuals of $\widetilde{Y}$ and $\widehat{\Theta}$. Note that the batch normalization and residual learning technique can stabilize and improve the denoising performance together. Batch normalization offers some merits for residual learning, such as alleviating internal covariate shift problem. Moreover, the residual learning can offset the adverse effect to the convergence of training caused by batch normalization. So the combination of residual learning and batch normalization can accelerate the training speed and improve the denoising performance. The advantage of combination of residual learning and batch normalization is demonstrated in simulation results. It is worth mentioned that we utilize the orthogonal initialization method within the convolution filters to effectively decide the balance between the operating efficiency and noise reduction.

## IV. SIMULATION RESULTS

In this section, the performance of FFDNet-based channel estimator for CF mmWave massive MIMO systems is illustrated. Moreover, we present simulation results for other channel estimators as benchmarks. The training rate is set as 0.001, and the training process stops if the training error keeps fixed in 5 sequential epochs. The mini-batch size is set as 125, and the rotation and flip based data augmentation is also adopted during training. The training, validation, and testing sets include 16000, 6000, and 8000 samples, respectively. Moreover, we
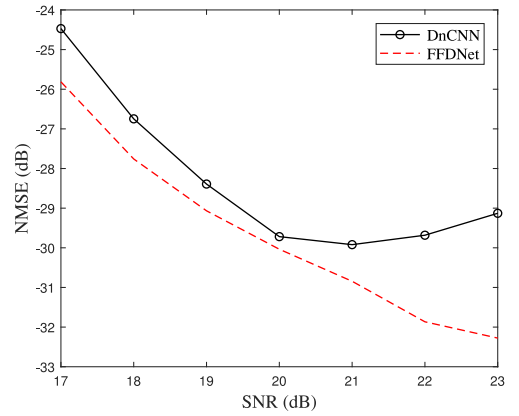


Fig. 2. NMSE performance comparison of FFDNet and DnCNN with one single model. $N_r = 64$, $N_t = 32$, layer $= 10$, feature $= 96$.

scale the data to the range of [0, 100] for training. The FFDNet network is trained by using the stochastic gradient descent (SGD) method.

Fig. 2 compares the NMSE performance of the FFDNet and the DnCNN with only one single model. The antenna number of APs and UEs are set 64 and 32, respectively. Node that the number of layers and features for both FFDNet and DnCNN is 10 and 96, respectively. The [17, 23] dB range of SNR of input received signal matrices for FFDNet and 20 dB SNR for DnCNN, because DnCNN can only accept a single noise level. Neural network parameter settings are designed to ensure adequate complex capacity redundancy.

As expected, it shows that FFDNet can achieve better NMSE performance for the SNR in the range of [18, 22] dB. DnCNN achieves the similar NMSE performance only at 20 dB of SNR, while DnCNN performs worse than FFDNet in other SNR ranges because FFDNet has a larger noise perception range.

Fig. 3 compares the NMSE performance of different channel estimation methods. The number of layers and features for both FFDNet and DnCNN is 9 and 64, respectively. We can see FFDNet and DnCNN outperform conventional channel estimation methods (LS and MMSE) by a large margin, especially when SNR is lower than 20 dB. This illustrates that the FFDNet achieves much better NMSE performance than LS and MMSE. Moreover, as the SNR increases, the advanced
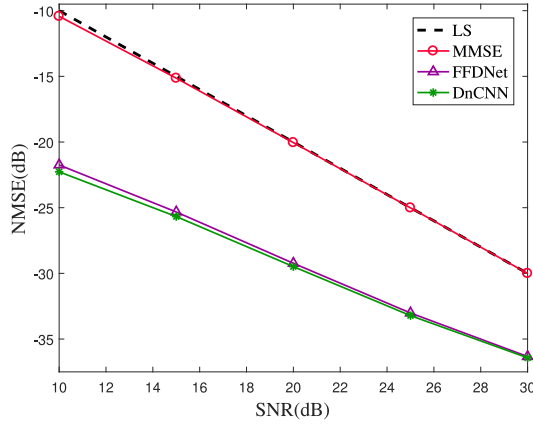
Fig. 3. Comparison of NMSE performance between the FFDNet, DnCNN, LS, and MMSE. $N_r = 64$, $N_t = 32$, layer $= 9$, feature $= 64$.
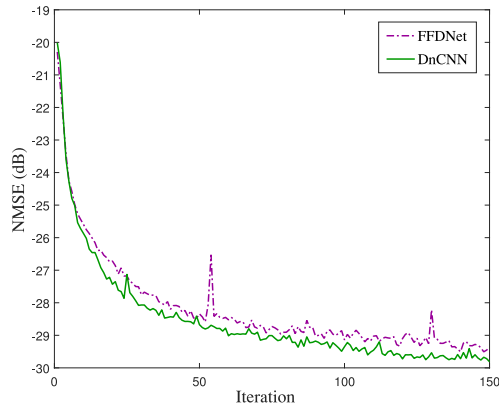


Fig. 4. Convergence of FFDNet and DnCNN. $N_r = 64$, $N_t = 32$, layer $= 9$, feature $= 96$.
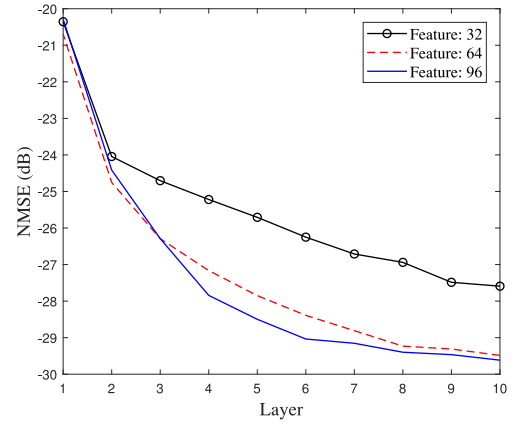


Fig. 5. NMSE performance comparison of FFDNet with different features and layers. $N_r = 64$, $N_t = 32$.
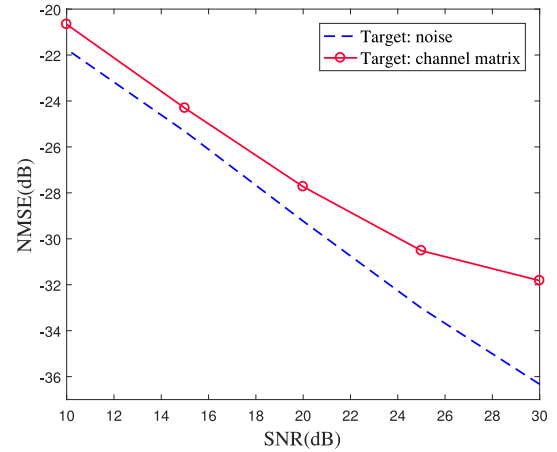


Fig. 6. NMSE performance comparison of FFDNet with different target: Noise and channel matrix. $N_r = 64$, $N_t = 32$, layer $= 9$, feature $= 64$.

performance of FFDNet is gradually reduced compared to LS and MMSE. For example, when SNR is 15 dB, the NMSE performance of the FFDNet obtains 10.33 dB gain over LS and obtains 8.02 dB gain when SNR is 25 dB. Furthermore, FFDNet is slightly inferior compared with DnCNN when the scale of neural network (Layer $= 9$, Feature $= 64$) is small, but the gap is gradually narrowing with the increasing of SNR level. For example, when SNR is 15 dB, the NMSE performance of the DnCNN surpass the FFDNet by 0.35 dB and when SNR is 25 dB, the DnCNN obtains 0.20 dB gain over the FFDNet. Overall, FFDNet produces the best quality of channel estimation than DnCNN and conventional methods.

Fig. 4 investigates the NMSE performance of FFDNet and DnCNN models against the number of iterations. The SNR of received signal is set to 20 dB. If the number of layers and features of FFDNet and DnCNN are 9 and 96, respectively, the NMSE performance of DnCNN is better than FFDNet. However, the gap between the two CNN schemes is always within 1 dB. Additional, DnCNN and FFDNet both converge within 150 iterations.

The NMSE performance of the FFDNet-based channel estimation for CF mmWave massive MIMO is further investigated in Fig. 5. We set the SNR to 20 dB. In all cases, the FFDNet achieves the better NMSE performance as the layer of neural network increases. Moreover, when the number of neural network features is 64 and 96, the NMSE performance converges within eight layers. Limited by complex capacity (e.g., layer $< 3$) of neural networks, the FFDNet with different

features like 32, 64 and 96 achieves similar NMSE performance; while layer $\geq 3$, the FFDNet with 64 or 96 features has a large NMSE performance margin compared with 32 features. Consequently, the main challenge in accurately describing noise is the lack of observational dimensions and modeling capabilities of neural networks, such as features and layers. From the experimental results in Fig. 5, we can increase the number of features and layers as appropriate to improve the denoising performance of channel estimation. By considering the balance of time-consuming and performance, the layer and feature of FFDNet can be set to 8 and 64, respectively. It is worth noting that the simulation results show the SNR (within at least 20 dB) has limited effect on the denoising performance of FFDNet, which further validates the robustness of FFDNet.

Fig. 6 shows the NMSE performance of FFDNet with and without using residual learning. If FFDNet applies residual learning, the target of FFDNet should be noise matrix; if not, the target should be the mmWave channel matrix. In Fig. 6, it shows that the residual learning model can result in the more stable convergence and better performance than the original one, especially when SNR $< 20$ dB.

Finally, the evaluation was performed in PyCharm Community Edition (Python 3.6 environment) on a computer with Intel Core i5-7600K CPU @ 3.8 GHz, 8 GB of RAM and an Nvidia GeForce GTX 1660Ti GPU. The average running times (in seconds) of FFDNet and DnCNN

are 0.004073 and 0.004603, respectively. Impressively, the running time of FFDNet is less than that of DnCNN.

## V. CONCLUSION

In this correspondence, we introduce a FFDNet-based channel estimation method for the CF mmWave massive MIMO systems. More specifically, several techniques have been utilized in the training and testing design of FFDNet, for example the utilization of noise level map as input, residual learning, and denoising in the space of sub-images. Our results show that FFDNet can achieve improved NMSE performance for different noise levels with only one training model. Thanks to its effectiveness, flexibility and efficiency, FFDNet can be used as a practical solution for channel estimation in CF mmWave massive MIMO systems.

## REFERENCES

[1] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. New York, NY, USA: Cambridge Univ Press, 2017.

[2] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[3] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2019.2925562.

[4] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[5] J. Zhang, Y. Wei, E. Björnson, Y. Han, and S. Jin, "Performance analysis and power control of cell-free massive MIMO systems with hardware impairments," *IEEE Access*, vol. 6, pp. 55 302–55 314, 2018.

[6] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.

[7] W. Fan, J. Zhang, E. Bjornson, S. Chen, and Z. Zhong, "Performance analysis of cell-free massive MIMO over spatially correlated fading channels," in *Proc. IEEE Int. Conf. Commun.*, 2019, pp. 1–6.

[8] J. Zhang, J. Zhang, J. Zheng, S. Jin, and B. Ai, "Expanded compute-and-forward for backhaul-limited cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2019, pp. 1–6.

[9] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 651–663, Sep. 2019.

[10] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44596–44612.

[11] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.

[12] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 62, no. 9, pp. 5117–5144, Sep. 2016.

[13] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[14] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?," 2019, *arXiv:1902.02647*.

[15] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[16] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[17] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[18] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[19] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.