Enhanced Deep Learning-Based Channel Estimation for 5G NR System Using Lightweight U-Net Architecture

Kamruzzaman Mahedi[1], Mostafa Zaman Chowdhury[1], and Yeong Min Jang[2]
[1]Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh
[2]Department of Electronics Engineering, Kookmin University, Seoul, Korea
Email: mahedi2403579@stud.kuet.ac.bd, mzaman@eee.kuet.ac.bd, yjang@kookmin.ac.kr

***Abstract*—Channel estimation is critical for reliable communication in 5G New Radio (NR) systems, where increased bandwidth and mobility introduce significant challenges. Traditional methods like Least Squares (LS) and Minimum Mean Square Error (MMSE) face challenges with computational complexity and requirement of channel statistics, while recent deep learning approaches using cascaded networks suffer from error propagation and inefficiency. This paper presents a novel lightweight U-Net architecture that jointly performs super-resolution and denoising in a single stage, achieving superior performance with significantly reduced complexity. By treating the time-frequency channel response as a 2D image, our approach employs a single-stage encoder-decoder network with skip connections for joint super-resolution and denoising. The proposed mixed-SNR training strategy enables a single model to operate across the entire SNR range (0-30 dB), eliminating the need for multiple model instances. Extensive simulations on 5G NR TDL channel models demonstrate that the proposed method achieves NMSE improvement over the state-of-the-art cascaded SRCNN+DnCNN approach, while reducing computational complexity, and inference time.**

***Keywords*—5G NR, Channel Estimation, Deep Learning, U-Net, Super-Resolution, TDL Channels.**

I. Introduction

A more intelligent and reliable communication system, as well as a faster data rate and reduced latency, are essential due to growing number of devices associated to the system. One of the most compelling ways to increase data throughput while preserving higher communication reliability for upcoming wireless systems is multiple-input multiple-output (MIMO) with large-scale antenna arrays [1]. In MIMO, orthogonal frequency division multiplexing (OFDM), a type of multi-carrier modulation, has been efficiently used in a extensive range of digital communications systems as a robust transmission method for extremely frequency selective channels [2]. When a signal is transmitted through a channel; channel characteristics typically cause the received signal to be distorted. In MIMO systems, recovering the transmitted signal over a multipath fading channel requires the receiver to estimate the channel characteristics to mitigate their effects. Channel estimation provides information about the distortion experienced by a transmitted signal as it passes through the wireless channel. To compensate the fading and/or co-channel interference, the information extracted from channel estimation is subsequently used by equalizers; enabling reconstruction of the original signal.

In a broader sense, traditional channel estimation techniques can be classified as pilot-based or training based, semi-blind, and blind channel estimation, which are illustrated in Fig. 1. The traditional methods have some advantages but there are some limitations. Training based method like least square (LS) has lower computational complexity but its performance quality is not good, having high bit error rate, and it doesn't use the channel statistics. The minimum mean squared error (MMSE) channel estimation technique has good performance than LS, use the channel statistics but its computational complexity is much higher. For better performance in pilot assisted channel estimation techniques sometimes proper pilot location management becomes challenging; extra pilots are added additional to payload data bit which causes the waste of bandwidth. Moreover, pilot contamination is a severe problem which degrade the performance of training-based channel estimation. These training-based methods come with bandwidth inefficiency incorporated with higher inaccuracy. Whereas, blind channel estimation has higher computational complexity. The third method, semi blind channel estimation, combines these two with a tradeoff between spectral inefficiency and computational complexity [3]. So, the computational complexity has to be reduced, estimation accuracy has to be increased, pilot contamination problem has to be removed as much as possible and moreover the challenges of channel estimation for moving and heterogeneous communication network have to be overcome. Artificial intelligence-based method has already shown huge impact for solving those above-mentioned limitations. Deep learning (DL) and machine learning (ML) have already used for efficient channel estimation in different network structures [4]. DL shows better performance than ML in most of the cases. DL methods have been adopted to pilot design [5], channel estimation [5], [6], [7], [8], [9], signal detection [7] , energy management for unmanned aerial vehicle [10], and CSI feedback [11].

To estimate the channel, cascaded DL architecture is well known; for example, in [12], time-frequency response between single Tx and Rx antenna is considered as 2D image for known pilot positions. Two cascaded DL models (SRCNN and DnCNN) are then used for improving image resolution and removing noise. Inspired by this work, we present a single U-Net based architecture which provides less complexity, less inference time, and improved accuracy. The main contributions of this work can be summarized as follows:

1) A lightweight U-Net that jointly performs super-resolution and denoising in a single stage is proposed.

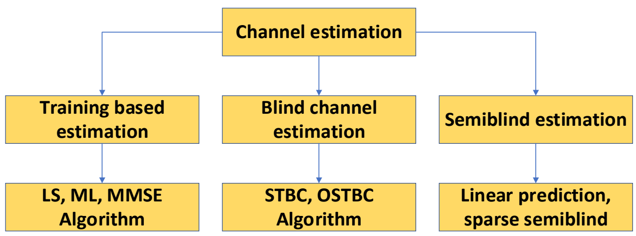

Fig. 1. Classifications of Traditional Channel Estimation Methods

2) Theoretical analysis proving that a single network can simultaneously learn interpolation and noise suppression through multi-scale feature extraction and fusion is provided.
3) Evaluation on 5G NR TDL channel models (TDL-A, TDL-C) as specified by 3GPP TR 38.901, along with legacy models (VehA, SUI5) is presented.

The rest of this work is organized as follows: Section II demonstrates overview of channel estimation methods. A lightweight U-Net model that jointly performs super-resolution and denoising in a single stage is proposed in section III. In section IV, performance analysis of proposed model is presented and finally section V concludes our research work.

II. Overview

A. Conventional Estimation

Data aided or training-based channel estimations use the known pilot data to estimate the channel. To overcome multipath fading channels, OFDM with cyclic prefix is applied. Now, for a single-antenna user, let $\mathbf{x} \in \mathbb{C}^{P\times 1}$ represents the transmitted signal vector and $\mathbf{y} \in \mathbb{C}^{Q\times 1}$ denotes the received signal vector. Hence, the received signal vector in the frequency domain can be obtained:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{Q\times P}$ is the matrix that denotes the channel response, $\mathbf{N} \in \mathbb{C}^{Q\times 1}$ is the noise vector. In case of OFDM system having subframe size of (Q×P) for single user, the tth time slot and nth subcarrier, where t ranges from 0 to (P-1), n ranges from 0 to (Q-1), the received signal for particular time-frequency slot can be represented as:

$$\mathbf{Y}_{nt} = \mathbf{H}_{nt}\mathbf{X}_{nt} + \mathbf{N}_{nt} \tag{2}$$

The channels at the pilot positions are estimated using the LS method. The estimated channel matrix $\hat{\mathbf{H}}_p$ is derived by minimizing the following cost function:

$$J(\hat{\mathbf{H}}_p) = \mathrm{E}\{\mathbf{y}_p - \mathbf{H}_p\mathbf{x}_p\} \tag{3}$$

where $J(.)$ represents the cost function, $\mathrm{E}\{.\}$ indicates expected value of the function, $\boldsymbol{x}_{\mathrm{p}}$ contains the transmitted pilot symbols, and $\boldsymbol{y}_{\mathrm{p}}$ denotes corresponding received signal. The least squares solution at pilot positions is obtained by taking the derivative of the cost function with respect to $\boldsymbol{H}_{\mathrm{p}}$ and setting it to zero, resulting in:

$$\hat{\mathbf{H}}_p = \left(\mathbf{x}_p^H\mathbf{x}_p\right)^{-1}\mathbf{x}_p^H\mathbf{y}_p \tag{4}$$

The estimated channel response vector is subsequently applied for equalization through interpolation, where the received signal is divided by the estimated channel response to compensate the channel's impact. Though it has less complexity, but it is poor at accuracy.

In MMSE estimation, which outperforms LS, the channel is estimated based on the minimum mean-square error criterion. The corresponding cost function is expressed as:

$$J(\hat{\mathbf{H}}) = E\{\mathbf{y} - \mathbf{x}\hat{\mathbf{H}}\,\mathbf{H}_p\} \tag{5}$$

The mean-squared error criterion is applied by minimizing $J(\hat{\mathbf{H}}_p)$ with respect to $\hat{\mathbf{H}}_p$. Taking the partial derivative and setting it to zero yields the MMSE solution:

$$\hat{\mathbf{H}}_{MMSE} = \mathbf{R}_{hy}\mathbf{R}_{yy}^{-1}y \tag{6}$$

where $\boldsymbol{R}_{hy}$ and $\boldsymbol{R}_{yy}$ can be defined as:

$$\mathbf{R}_{hy} = \mathbf{R}_{HH}\mathbf{y}, \tag{7}$$

$$\mathbf{R}_{yy} = \mathbf{x}\mathbf{R}_{HH}\mathbf{x}^H + \sigma^2 I \tag{8}$$

where $\boldsymbol{R}_{hy}$ denotes the cross-covariance matrix between the channel matrix H and the received data matrix **Y**, $\boldsymbol{R}_{yy}$ represents the autocovariance matrix of the received data matrix **Y**, $\boldsymbol{R}_{HH}$ is the autocovariance matrix of the channel matrix **H**, σ^2 is the noise variance. Ideal MMSE shows the best result with full knowledge of channel statistics, which is unfeasible in most of the scenarios.

B. ChannelNet

Focus is placed on the individual propagation link between a single pair of Tx and Rx antenna while in motion. To recover the full time–frequency channel response from pilot symbols, they adopt a lattice-type pilot pattern similar to LTE. The least-squares estimates at pilot positions, $\mathbf{h}_p^{LS}$, serve as a noisy, low-resolution representation of the channel. A two-stage learning framework comprising of Convolutional Neural Network (CNN) is then employed to reconstruct the complete channel image.

In the first stage, a super-resolution (SR) network maps the vectorized real and imaginary components of $\mathbf{h}_p^{LS}$ to a high-resolution channel estimate. Utilizing the SRCNN architecture, which initially up samples the input through interpolation and then refines it using a three-layer CNN. In the second stage, a denoising network (DnCNN) is cascaded with the SR module to mitigate noise artifacts.

III. Lightweight U-Net Architecture

LS estimation has low complexity but its accuracy is poor. On the other hand, MMSE has a great improvement than LS but then here, for proper channel estimation, full channel statistics should be known; which is impractical in most of the cases. A deep learning-based model is proposed in [12] consisting of two cascaded CNN for super-resolution and denoising which is obviously better choice than that of LS and ALMMSE. But using cascaded network increases not only the complexity but also the inference time. A single network can be used for better performance in estimating the channel by using U-Net. Both super-resolution and denoising are inverse problems solved through learned regularization. The U-Net learns a single unified regularizer that exploits spatial correlation for interpolation while simultaneously filtering noise based on learned signal statistics. The multi-scale architecture with skip connections enables different network depths to specialize in different aspects: some scales capture global structure for SR, while other scales preserve local details for denoising. Therefore, a single U-Net having only 3 encoder-decoder layers to outperform the state of the art is the main aspect of this work.

The proposed U-Net architecture addresses the channel estimation problem by treating it as a joint image super-resolution and denoising task. Unlike cascaded approaches where SR and denoising are performed sequentially by separate networks, our single-stage U-Net architecture performs both tasks simultaneously through a unified learning framework. Fig. 2 illustrates the complete U-Net Lite architecture, consisting of three main parts: (1) contracting path (encoder), (2) bottleneck with channel attention, and (3) expanding path (decoder) with skip connections. Each encoder performs a 2-layer convolution operation with 3×3 kernel followed by ReLU activation and then max-pooling is accomplished, reduces the image shape half to the input. The output from the max-pooling operation is the input to the next encoder stage. For the first encoder stage, input dimension is 72×14 with two channels; one is for real part and another is for imaginary part. Filter size for first encoder stage is 32 and is doubled in each next step. After 3rd encoder stage bottleneck is reached having input size 9×1 with 256 filters in convolution. Output of the bottleneck is then upsampled and fed to the first decoder stage. Each decoder stage consists of concatenated output from the corresponding encoder, 2 convolutional layers with 3×3 kernel followed by ReLU activation. Output of each decoder except the last one is upsampled by the size 2×2 and fed to the next decoder. Finally, a 1×1 convolution operation is performed to the last decoder output to get the desired estimated channel response.

First encoder performs denoising capturing local noise patterns and pilot positions. Encoder 2, 3, and bottleneck stages execute SR by capturing medium-range correlations Doppler shift, and delay spread structure. Concatenation preserves clean high-frequency details from encoder for de-noising while decoder provides interpolated structure from bottleneck. The resize operations in encoder 3 and bottleneck are necessary because the input dimension 72×14 is not a power of 2. Max-Pooling with stride 2 causes: 14→7→3→1 by floor division. Up-sampling produces: 1→2, 3→6, 7→14. The resize procedure using bilinear interpolation regulates up-sampled features to match skip connection sizes before concatenation, preserving spatial information without loss.

Our properly designed neural network can learn unified regularizer *R(H)* that simultaneously handles spatial smoothness and noise suppression. The joint problem can be defined as:

$$\mathbf{A}_{joint}(\mathbf{H}) = \mathbf{M} \odot (\mathbf{H} + \mathbf{N}) \tag{9}$$

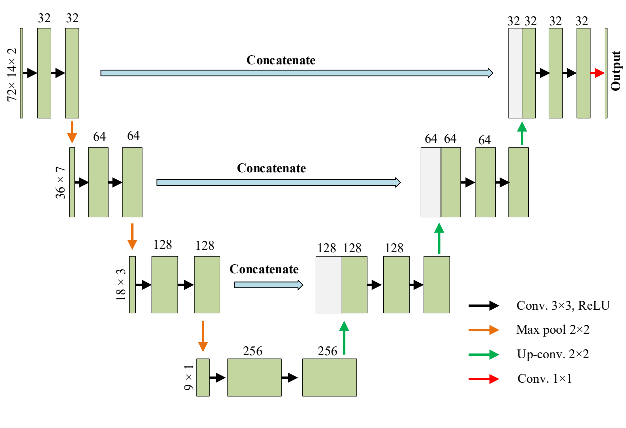

Fig. 2. Architecture of Proposed U-Net Model.

where $\mathbf{M} \in \{0,1\}^{Q \times P}$ is the pilot mask, $\odot$ is the element wise multiplication, and $\mathbf{A}_{joint}(\mathbf{H})$ is forward operator for combined SR and denoising techniques. Let $p = \{(n_p, t_p)\}_{p=1}^{N_p}$ denote the set of pilot positions, $\boldsymbol{M}$ be the pilot mask where $\mathbf{M}_{nt} = 1$ if (n, t) ∈ p and 0 otherwise, and $N_p = 48$ pilots per subframe.

$$\mathbf{A}_{SR}(\mathbf{H}) = \mathbf{M} \odot \mathbf{H}, \tag{10}$$

$$\mathbf{A}_{DN}(\mathbf{H}) = \mathbf{H} + \mathbf{N} \tag{11}$$

where $\mathbf{A}_{SR}(\mathbf{H})$ is the forward operator for SR and $\mathbf{A}_{DN}(\mathbf{H})$ is for noise suppression and both SR and denoising operations can be performed by (9) as it is the combination of (10) and (11). Channel response for proposed model is then:

$$\hat{\mathbf{H}}_{UNet} = E\| y - \mathbf{A}_{joint}(\mathbf{H}) \|^2 + \lambda R(\mathbf{H}) \tag{12}$$

where $\lambda = 0.1$ for best generalization. At this value, data fits reasonably well at pilots and smooth interpolation is obtained between pilots.

To enable joint learning of super-resolution and denoising, we design a multi-objective loss function.

$$L_{total} = L_{MSE} + \lambda_f L_f + \lambda_p L_p \tag{13}$$

where L_{MSE} represents spatial domain MSE loss, L_f is the frequency domain loss, and L_p is the weighted pilot loss. L_{MSE} ensures overall estimation accuracy and affects both SR and denoising uniformly which is defined as:

$$L_{MSE} = \frac{1}{P \times Q} \sum_{n=1}^{Q} \sum_{t=1}^{P} |\mathbf{H}_{nt} - \hat{\mathbf{H}}_{nt}|^2 \tag{14}$$

L_f preserves spectral characteristics and improves interpolation quality, can be defined as:

$$L_f = \frac{1}{P \times Q} \sum_{n=1}^{Q} \sum_{t=1}^{P} |F(\mathbf{H}_{nt}) - F(\hat{\mathbf{H}}_{nt})| \tag{15}$$

where F denotes 2D Fast Fourier Transform.

$$L_p = \frac{1}{N_P} \sum_{(n,t) \in P} \omega_{nt} |\mathbf{H}_{nt} - \hat{\mathbf{H}}_{nt}|^2 \tag{16}$$

where $\omega_{nt} = 2$ for all pilot positions. This ensures pilot

positions are accurately estimated first, providing anchor points for interpolation (denoising emphasis). We set $\lambda_f = 0.1$ and $\lambda_p = 2$ based on validation set performance. The frequency loss weight is kept small to avoid overemphasis on spectral matching, while the pilot loss weight is higher to ensure reliable anchor points. We use normalized mean squared error (NMSE) as our performance metrics, defined as:

$$NMSE(dB) = 10\log_{10}\left(\frac{1}{P \times Q} \sum_{n,t} |\mathbf{H}_{nt} - \hat{\mathbf{H}}_{nt}|^2\right) \tag{17}$$

IV. PERFORMANCE ANALYSIS

In this segment we train our model and evaluate the NMSE over the uniform SNR ranges from 0 to 30 dB and collate the results with the state of the arts. All the simulations are performed in pyhton following 5G NR TDL channel models (TDL-A, TDL-C) as specified by 3GPP TR 38.901, along with legacy models (VehA, SUI5), demonstrating superior performance in challenging high delay spread scenarios where conventional methods degrade significantly. Following 5G NR specifications, we employ a lattice-type pilot pattern where pilots are distributed in a diamond-shaped arrangement. Traditional approaches require separate models for different SNR ranges. A mixed-SNR training strategy that enables a single model to handle the entire SNR spectrum is proposed by this work. The network learns to automatically adjust its processing based on input noise level. At low SNR, it emphasizes spatial smoothing (denoising), while at high SNR, it preserves sharp features (SR). Keras and Tensorflow along with GPU backend are used to evaluate our model. For training, testing and validation we use 30000, 5000, and 5000 samples, respectively. For optimization, we use Adam optimizer with initial learning rate of 0.0001, batch size of 128, and maximum 250 epoch. Fig. 3 depicts NMSE over a range of SNR for VehA channel. Our proposed method outperforms SRCNN+DnCNN model for all SNR values. For higher SNR values ranges from 20-30 dB, SRCNN+DnCNN exhibits poor performance than ALMMSE whereas better adjust its processing based on input noise level. At low SNR, it emphasizes spatial smoothing (denoising), while at high SNR, it preserves sharp features (SR). Keras and Tensorflow along with GPU backend are used to evaluate our model. For training, testing and validation we use 30000, 5000, and 5000 samples, respectively. For optimization, we use Adam optimizer with initial learning rate of 0.0001, batch size of 128, and maximum 250 epoch. Fig. 3 depicts NMSE over a range of SNR for VehA channel. Our proposed method outperforms SRCNN+DnCNN model for all SNR values. For higher SNR values ranges from 20-30 dB, SRCNN+DnCNN exhibits poor performance than ALMMSE whereas better performance is obtained from our proposed one but not very close to ideal MMSE.

In Fig. 4 NMSE versus SNR for SUI5 channel is illustrated. Here, again our proposed method shows better performance than conventional methods but it is observed that here performance of our work degrades due to the higher complexity of SUI5 channel than that of VehA channel.

As for 5G NR TDL channel models specified by 3GPP TR 38.901 has TDL-A to TDL-E channel models having different delay spread. Among them we use TDL-A and TDL-C as we get moderate and high delay spread for them, respectively. Due to the higher delay spread performance in TDL-C is worse than TDL-A. The performance of our proposed method for TDL-A is better than SRCNN+DnCNN method which can be clearly seen in Fig. 5. More interestingly, for TDL-C it shows slightly better performance though having less complexity is illustrated in Fig. 6.

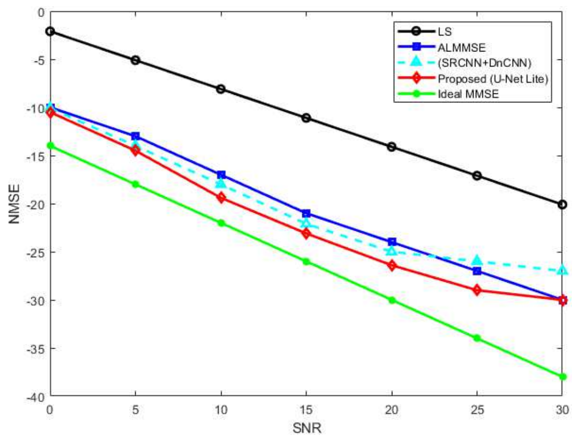

Fig. 3. NMSE vs SNR for VehA Channel.

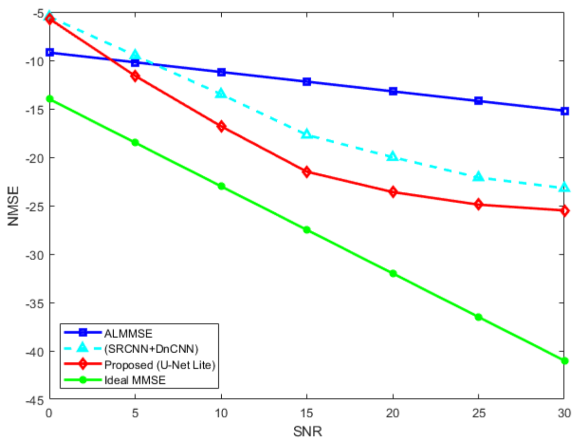

Fig. 4. NMSE vs SNR for SUI5 Channel.

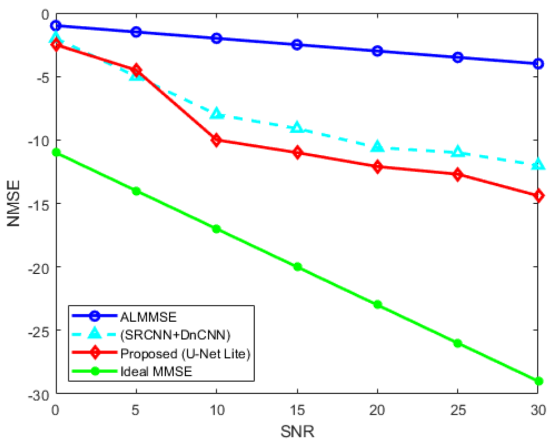

Fig. 5. NMSE vs SNR for TDL-A Channel.

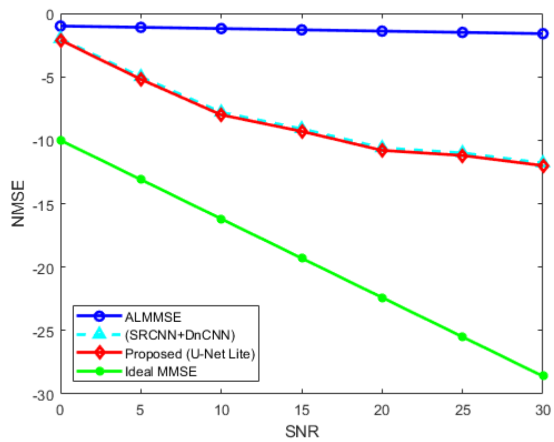

Fig. 6. NMSE vs SNR for TDL-C Channel.

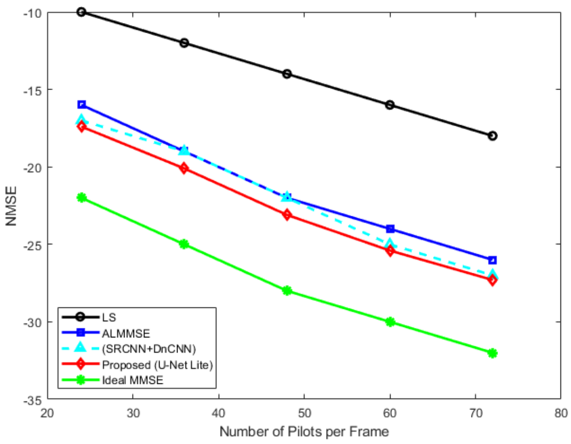

Fig. 7. NMSE vs No. of Pilots per Frame.

Fig. 7 depicts effects of number of pilots per frame on NMSE for VehA model, when the SNR value is fixed at 20 dB. Our proposed model shows better result not only than ALMMSE but also than SRCNN+DnCNN.

V. Conclusion

This paper proposes a lightweight U-Net architecture for enhanced channel estimation which meets superior performance, lower complexity, better generalization, and 5G NR compatibility as we include TDL-A and TDL-C channel models along with legacy models. By treating the time-frequency channel response as a 2D image, our approach performs super-resolution and denoising jointly in a single stage through multi-scale feature extraction and fusion. The encoder-decoder structure with skip connections enables efficient learning of both interpolation and noise suppression without error propagation inherent in cascaded methods. The results reveal that the proposed U-Net outperforms SRCNN+DnCNN strategy with less complexity and close to the performance of ideal MMSE for each and every channel model. We use mixed-SNR training strategy eliminating the needs for multiple model instances, making it practical for real-world deployment. This work focuses on individual propagation links between single Tx-Rx antenna pairs, can be extended to massive MIMO.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF), South Korea grant funded by the Korea government (MSIT) (No. 2022R1A2C1007884).

References

[1] Lu Lu, Ye li, A. Lee Swindlehurst, Alexei Ashikhmin, and Rui Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing,* vol. 8, no. 5, pp. 742-758, Oct. 2014.

[2] G. Auer and E. Karipidis, "Pilot aided channel estimation for OFDM: a separated approach for smoothing and interpolation," in *Proc. of IEEE International Conference on Communications*, Seoul, Korea (South), Vol. 4, pp. 2173-2178, 2005.

[3] H. Kaur, M. Khosla and R. K. Sarin, "Channel estimation in MIMO-OFDM system: a review," in *Proc. of 2018 Second International Conference on Electronics,* Communication *and Aerospace Technology (ICECA)*, pp. 974-980, 2018.

[4] Al-Imran, M. Z. Chowdhury, R. B. Mofidul, and Y. M. Jang, "Machine learning and deep learning in FSO communication: a comprehensive survey," *ICT Express,* 2025.

[5] H. Fu, W. Si and I. -M. Kim, "Deep learning-based joint pilot design and channel estimation for OFDM systems," *IEEE Transactions on Communications,* vol. 71, no. 8, pp. 4577-4590, Aug. 2023.

[6] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga and A. Behboodi, "Deep learning-based channel estimation for wideband hybrid mmWave massive MIMO," *IEEE Transactions on Communications,* vol. 71, no. 6, pp. 3679-3693, June 2023.

[7] J. Li, Z. Zhang, Y. Wang, B. He, W. Zheng and M. Li, "Deep learning-assisted OFDM channel estimation and signal detection technology," *IEEE Communications Letters,* vol. 27, no. 5, pp. 1347-1351, May 2023.

[8] Al-Imran, M. S. Nazim, H. Nguyen, and Y. M. Jang, "Channel estimation of massive MIMO FSO communication system using deep attention residual U-Net," *ICT* Express*,* vol. 11, no. 2, pp. 287-292, Apr. 2025.

[9] R. B. Mofidul, Al-Imran, M. S. Nazim, and Y. M. Jang, "Hybrid CNN-Swin transformer denoising network for channel estimation in optical IRS-assisted indoor MIMO VLC system," *ICT Express,* 2025.

[10] M. S. Nazim, M. M. Rahman, M. I. Joha, and Y. M. Jang, "Optimizing energy efficiency in hybrid UAVs using DQN-based energy management system," *ICT Express,* 2025.

[11] J. Wang, X. Zhang, Z. Lu, J. Tan and H. Zhang, "Practical deployment for deep learning-based CSI feedback systems: generalization challenges and enabling techniques," *IEEE Communications Magazine,* pp. 1-7, March 2025.

[12] M. Soltani, V. Pourahmadi, A. Mirzaei and H. Sheikhzadeh, "Deep learning-based channel estimation," in *Proc. of IEEE Communications Letters*, vol. 23, no. 4, pp. 652-655, April 2019.