Analysis of Deep Learning-based MIMO Detectors

1st Sangbu Yun
dept. Electrical Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Korea
sb9767@postech.ac.kr

2nd Youngjoo Lee
dept. Electrical Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Korea
youngjoo.lee@postech.ac.kr

***Abstract*—Multiple-Input Multiple-Output (MIMO) communication systems have become a fundamental technology in modern wireless networks due to their ability to enhance data rates and system capacity. However, traditional MIMO detection algorithms face significant challenges, including increasing complexity and performance degradation with growing system dimensions. Deep learning has shown great promise in various domains in recent years, leading researchers to explore its potential in addressing MIMO detection capability. This paper provides a comprehensive overview of deep-learning-based MIMO detection techniques, presenting an extensive literature review and taxonomy of approaches. We conduct a comparative analysis with conventional techniques and evaluate the detection performance of deep learning-based approaches. The paper also compares the required number of FLOPs operations to identify the potential of each detection method.**

***Index Terms*—MIMO detection, Deep-learning**

I. Introduction

Multiple-Input Multiple-Output (MIMO) communication systems have emerged as a key technology for enhancing data rates and improving overall system performance in modern wireless communication networks [1]. MIMO leverages multiple antennas at both the transmitter and receiver to enable spatial diversity and multiplexing gains, thereby significantly increasing spectral efficiency and system capacity. However, MIMO systems also face formidable challenges, particularly in the area of accurate signal detection and decoding.

Traditional MIMO detection algorithms, such as Maximum Likelihood (ML), Minimum Mean Squared Error (MMSE), and Zero-Forcing (ZF), have proven effective but suffer from increasing computational complexity and performance degradation as the number of antennas and users grows. As a result, there has been a growing interest in exploring alternative detection techniques that can address these limitations and provide more efficient and reliable solutions.

This work was supported in part by Korea Research Institute for defense Technology planning and advancement - Grant funded by Defense Acquisition Program Administration(DAPA) (KRIT-CT-22-078), in part by the Institute of Information communications Technology Planning Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2021-0-00165, Development of 5G+ Intelligent Basestation Software Modem), and in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) under Grant 2022R1A2C2092521. *(Corresponding Author: Youngjoo Lee)*

Authors are with the Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, South Korea. (e-mail: sb9767@postech.ac.kr; youngjoo.lee@postech.ac.kr)

Among the various emerging techniques, deep learning has gained remarkable attention in recent years for its ability to learn complex representations from large-scale data automatically. The success of deep learning in diverse domains, such as image recognition, natural language processing, and speech recognition, has motivated researchers to explore its potential applications in wireless communication systems, including MIMO detection [2].

This paper aims to provide a comprehensive overview of the current state-of-the-art in deep-learning-based MIMO detection techniques. By presenting an extensive review of the literature, we aim to shed light on the advancements, challenges, and potential opportunities of integrating deep learning into MIMO detection schemes. Our work encompasses a diverse set of research papers, publications, and technical reports, showcasing the broad scope of this rapidly evolving field.

II. System Model

Let N_t and N_r denote the number of the transmitting and receiving antennas in a MIMO system, respectively. Suppose the transmitted symbols are from the symbol constellation, and the set of the constellation symbols is denoted by $\tilde{\Theta}$. The transmitted symbol vector can be represented by a $N_t \times 1$ vector $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_{N_t}]$ with each component $\tilde{x}_i \in \tilde{\Theta}$. For given channel matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$, the received signal vector $\tilde{\mathbf{y}}$ is given by

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{1}$$

where $\tilde{\mathbf{n}}$ denotes the additive white Gaussian noise (AWGN) vector such that $\tilde{n_i} \sim \mathcal{CN}(0, \tilde{\sigma}^2), \forall i \in \{1, \cdots, N_r\}$. The complex-valued system in Eq. (1) can be reformulated into an equivalent real-valued system by real value decomposition as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2}$$

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} \Re(\tilde{\mathbf{H}}) & -\Im(\tilde{\mathbf{H}}) \\ \Im(\tilde{\mathbf{H}}) & \Re(\tilde{\mathbf{H}}) \end{bmatrix}, \\ \mathbf{y} &= \begin{bmatrix} \Re(\tilde{\mathbf{y}}) \\ \Im(\tilde{\mathbf{y}}) \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \Re(\tilde{\mathbf{x}}) \\ \Im(\tilde{\mathbf{x}}) \end{bmatrix}, \mathbf{n} = \begin{bmatrix} \Re(\tilde{\mathbf{n}}) \\ \Im(\tilde{\mathbf{n}}) \end{bmatrix}, \end{aligned} \tag{3}$$

in which $\mathbf{y} \in \mathbb{R}^N$, $\mathbf{H} \in \mathbb{R}^{N \times K}$, and $\mathbf{x} \in \Theta^K$, where $N = 2N_r, K = 2N_t$, and Θ is the set of the alphabet for the real and imaginary parts of the symbol constellation $\tilde{\Theta}$.

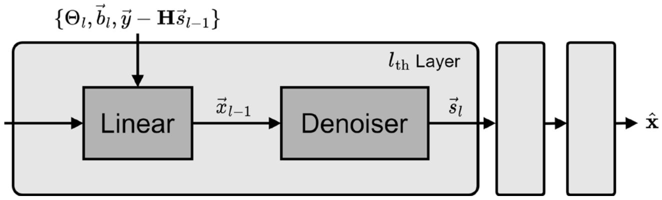

Fig. 1. An architecture overview of an iterative detector.

III. Learning-Based MIMO Detection Networks

Deep-unfolding architecture is to design trainable architectures by parameterizing iterative optimizers with a fixed number of iterations, which is illustrated in Fig. 1. The basic idea is to unfold the iterations of a specific algorithm into a layered architecture, which iteratively estimates the solution of MLD. The general form of these iterative algorithms comprises the following updates:

$$\begin{aligned}\hat{\mathbf{x}}^{[l]} &= \hat{\mathbf{s}}^{[l-1]} + \mathbf{\Theta}^{[l]}\left(\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{[l-1]}\right) + \mathbf{b}^{[l]} \\ \hat{\mathbf{s}}^{[l]} &= \Phi_t(\hat{\mathbf{x}}^{[l]}),\end{aligned} \tag{4}$$

where the first step computes residual error to compute intermediate signal $\hat{\mathbf{s}}^{[l]}$, then applies denoiser Φ_t to produce new estimation $\hat{\mathbf{s}}^{[l]}$.

1) *DetNet*: Inspired by iterative projected gradient descent optimization, DetNet [3] updates $\hat{\mathbf{s}}$ with L layers as

$$\begin{aligned}\hat{\mathbf{s}}^{[l]} &= \Pi\left[\mathbf{s} - \theta^{[l]}\frac{\partial\|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2}{\partial\mathbf{s}}\right]_{\mathbf{s}=\hat{\mathbf{s}}^{[l-1]}} \\ &= \Pi\left[\hat{\mathbf{s}}^{[l-1]} - \theta^{[l]}\mathbf{H}^\mathsf{T}\mathbf{y} + \theta^{[l]}\mathbf{H}^\mathsf{T}\mathbf{H}\hat{\mathbf{s}}^{[l-1]}\right],\end{aligned} \tag{5}$$

where $\Pi[\cdot]$ denotes a nonlinear projection operator and $\theta^{[l]}$ is a gradient step size. Each iteration in the DetNet architecture is carried out by a single layer which is composed of a linear combination of $\hat{\mathbf{s}}^{[l]}$, $\mathbf{H}^\mathsf{T}\mathbf{y}$ and $\mathbf{H}^T\hat{\mathbf{x}}_k$ and non-linear projection. The performance can be enhanced by training step size θ_k at each layer. Thus, the DetNet is described as

$$\begin{aligned}\mathbf{q}^{[l]} &= \hat{\mathbf{s}}^{[l-1]} - \theta_1^{[l]}\mathbf{H}^\mathsf{T}\mathbf{y} + \theta_2^{[l]}\mathbf{H}^\mathsf{T}\mathbf{H}\hat{\mathbf{s}}^{[l-1]}, \\ \mathbf{x}^{[l]} &= \left[\mathbf{v}^{[l-1]}, \mathbf{q}^{[l]}\right]^\mathsf{T}, \\ \mathbf{z}^{[l]} &= \sigma(\mathbf{W}_1^{[l]}\mathbf{x}^{[l]} + \mathbf{b}_1^{[l]}), \\ \hat{\mathbf{s}}^{[l]} &= \psi_t(\mathbf{W}_2^{[l]}\mathbf{x}^{[l]} + \mathbf{b}_2^{[l]}), \\ \mathbf{v}^{[l]} &= \mathbf{W}_3^{[l]}\mathbf{z}^{[l]} + \mathbf{b}_3^{[l]},\end{aligned} \tag{6}$$

where $\psi_t(x) = -q + \frac{1}{|t|}\sum_{i\in\Omega}[\sigma(x+i+t) - \sigma(x+i-t)]$ with $q = 1, \Omega = \{0\}$ for QPSK, and σ denotes $\mathrm{ReLU}(x)$. Many research efforts have been conducted to enhance the DetNet architecture, due to its impractical computational complexity.

2) *ScNet:* The ScNet [4] simplifies DetNet by removing $\mathbf{v}^{[l]}$ in Eq. (6), and input and output of each layer are directly connected element-wise. With the proposed simplifications, the complexity of the unfolded layer is reduced from $\mathcal{O}(64K^2)$ to $\mathcal{O}(3K)$ when compared to that of DetNet. The output of each layer is updated as

$$\begin{aligned}\mathbf{x}^{[l]} &= \left[\hat{\mathbf{s}}^{[l-1]}, \mathbf{H}^\mathsf{T}\mathbf{y}, \mathbf{H}^\mathsf{T}\mathbf{H}\hat{\mathbf{s}}^{[l-1]}\right]^\mathsf{T} \in \mathbb{R}^{3K\times 1}, \\ \hat{\mathbf{s}}^{[l]} &= \psi_t(\mathbf{w}^{[l]} \odot \mathbf{x}^{[l]} + \mathbf{b}^{[l]}),\end{aligned} \tag{7}$$

where $\odot$ denotes an element-wise multiplication. Even if the method dramatically reduces computational complexity, it still suffers from inferior detection capability.

3) *FS-Net:* The FS-Net [5] also follows updating process of Eq. (5), and it can be rewritten as

$$\hat{\mathbf{s}}^{[l]} = \Pi\left[\hat{\mathbf{s}}^{[l-1]} + \theta^{[l]}(\mathbf{H}^\mathsf{T}\mathbf{H}\hat{\mathbf{s}}^{[l-1]} - \mathbf{H}^\mathsf{T}\mathbf{y})\right], \tag{8}$$

which implies that the elements at the same position of $\mathbf{H}^\mathsf{T}\mathbf{H}\hat{\mathbf{s}}^{[l]}$ and $\mathbf{H}^\mathsf{T}\mathbf{y}$ can be updated with same parameters. Therefore, FS-Net sets the input vector of $(l+1)$th layer to

$$\hat{\mathbf{x}}^{[l]} = \left[\hat{\mathbf{s}}^{[l-1]}, \mathbf{H}^\mathsf{T}\mathbf{H}\hat{\mathbf{s}}^{[l-1]} - \mathbf{H}^\mathsf{T}\mathbf{y}\right]^\mathsf{T} \in \mathbb{R}^{2K\times 1}, \tag{9}$$

which reduces the input and, eventually, the internal size of the network to 2/3 of ScNet.

4) *OAMPNet2:* OAMPNet2 [6] learns tuning parameters per iteration to the OAMP algorithm [7] as

$$\begin{aligned}\hat{\mathbf{x}}^{[l]} &= \hat{\mathbf{s}}^{[l-1]} + \theta_l^{(1)}\mathbf{H}^\top\left(v_l^2\mathbf{H}\mathbf{H}^\top + \sigma^2\mathbf{I}\right)^{-1}(\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{[l-1]}), \\ \hat{\mathbf{s}}^{[l]} &= \eta_t\left(\hat{\mathbf{x}}^{[l]}; \sigma_l^2\right), \\ v_l^2 &= \frac{\left\|\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{[l-1]}\right\|_2^2 - \mathrm{tr}\left(\mathbf{R_{nn}}\right)}{\mathrm{tr}\left(\mathbf{H}^\top\mathbf{H}\right)}, \\ \sigma_l^2 &= \frac{\theta_l^{(2)}}{N_t}\left(\frac{\left\|\mathbf{I} - \theta_l^{(1)}\mathbf{H}^\top\mathbf{H}\right\|_F^2}{\|\mathbf{H}\|_F^2}\left[\left\|\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{[l-1]}\right\|_2^2 - N_r\sigma^2\right]_+ \right. \\ &\qquad \left. + \frac{\left\|\theta_l^{(1)}\mathbf{H}^\top\right\|_F^2}{\|\mathbf{H}\|_F^2}\sigma^2\right),\end{aligned} \tag{10}$$

where η_t denotes the optimal denoiser for Gaussian noise and $\mathbf{R_{nn}}$ is the noise covariant matrix. For n_{th} element of vector $\mathbf{z}$, $\eta_{t,n}$ is computed as

$$\begin{aligned}\eta_{t,n}\left(z_n; \sigma_t^2\right) &= \frac{1}{Z}\sum_{x_i\in\tilde{\Theta}} x_i \exp\left(-\frac{\|z_n - x_i\|^2}{\sigma_l^2}\right), \\ Z &= \sum_{x_j\in\tilde{\Theta}} \exp\left(-\frac{\|z_n - x_j\|^2}{\sigma_l^2}\right).\end{aligned} \tag{11}$$

5) *MMNet:* MMNet [8] has two different versions based on the target channel condition. The internal estimates of each algorithm are

$$\begin{aligned}\hat{\mathbf{x}}^{[l]} &= \begin{cases}\hat{\mathbf{s}}^{[l-1]} + \theta_l^{(1)}\mathbf{H}^\top(\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{[l-1]}), & \text{i.i.d} \\ \hat{\mathbf{s}}^{[l-1]} + \mathbf{\Theta}_l^{(1)}(\mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{[l-1]}), & \text{Otherwise,}\end{cases} \\ \hat{\mathbf{s}}^{[l+1]} &= \begin{cases}\eta_t\left(\hat{\mathbf{x}}^{[l]}; \sigma_l^2\right), & \text{i.i.d} \\ \eta_t\left(\hat{\mathbf{x}}^{[l]}; \boldsymbol{\sigma}_l^2\right), & \text{Otherwise,}\end{cases}\end{aligned} \tag{12}$$

and their noise variance at the input of the denoiser is given as scalar σ_t^2 and vector $\boldsymbol{\sigma}_t^2$ for each case.

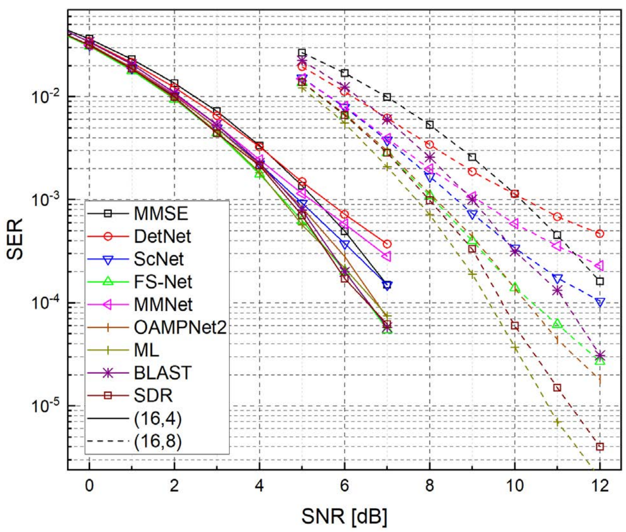

Fig. 2. SER simulation result when $(N_r, N_t) = \{(16, 4), (16, 8)\}$.

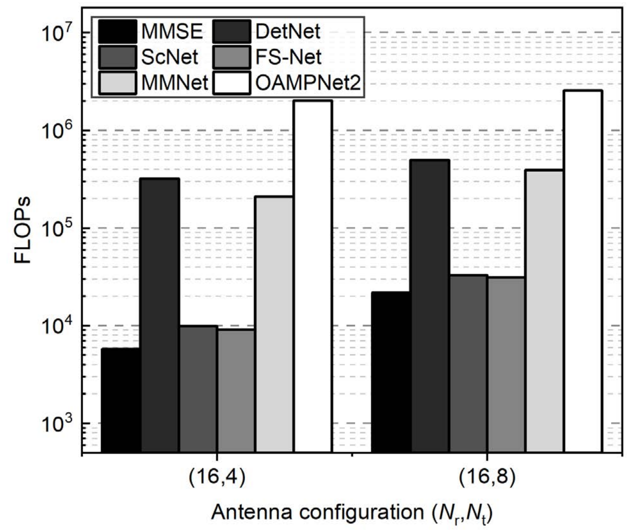

Fig. 3. FLOPs comparison between baseline MMSE and other DL-detectors.

IV. Result Analysis

A. Symbol Error Rate Simulation

We conducted simulations to assess the SER performance of various deep learning models compared to conventional detection techniques. In the simulations, we consider two antenna configurations: $(N_r, N_t) \in \{(16, 4), (16, 8)\}$ combined with QPSK modulation and the independent and identically distributed (i.i.d) Rayleigh fading channel. The SER performance of the proposed network when $N_r = 16$; $N_t = \{4, 8\}$ is illustrated in Fig. 2. Our results consistently demonstrated that deep learning models exhibited competitive SER performance across different Signal-to-Noise Ratio (SNR) ranges. OAMPNet2 detector shows the best detection performance among DL-driven methods, while FS-Net also showed comparable performance for both configurations. In particular, the performance gap between classic MMSE and other DL detectors become larger when the number of transmit antenna increases.

B. Computational Complexity Analysis

To evaluate the computational complexity of deep learning models, we performed FLOPs analysis across a range of architectures and system configurations. Our analysis highlighted that while deep learning models generally entail higher computational requirements compared to traditional algorithms for MIMO detection, the improvements in SER performance often justify the increased computational load. Notably, some prior works, such as ScNet and FS-Net architecture, have shown reasonable computational overhead with significant improvement in SER performance. On the other hand, DL detectors with an optimal denoiser OAMPNet and MMNet requires high computational cost due to the inherent expectation matching operation, which can be represented by $\mathrm{SoftMax}$ operator.

In conclusion, our results underscore the potential of deep learning to enhance MIMO detection performance with moderate complexity overhead. While considering computational cost remains essential, the adaptability of deep learning architectures to varied SNR conditions reinforces their utility in real-world wireless communication scenarios. Careful selection and optimization of deep learning models are crucial to harness their advantages for MIMO detection.

V. Conclusion

This paper delves into the confluence of deep learning and MIMO detection, illuminating the promising path ahead for enhancing wireless communication systems. Our analysis underscores the trade-offs between computational complexity and performance advantages, emphasizing the need for tailored solutions in diverse MIMO scenarios. Positioned at the forefront of wireless technology evolution, this paper serves as a guiding compass for researchers and practitioners, charting the course for innovative advancements in MIMO detection through the lens of deep learning.

References

[1] I. Update, "Ericsson mobility report," *Ericsson: Stockholm, Sweden*, 2018.

[2] M. A. Albreem, A. H. Alhabbash, S. Shahabuddin, and M. Juntti, "Deep learning for massive MIMO uplink detectors," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 741–766, Dec. 2021.

[3] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.

[4] G. Gao, C. Dong, and K. Niu, "Sparsely connected neural network for massive MIMO detection," in *2018 IEEE 4th Int. Conf. Comput. Commun. (ICCC)*. IEEE, Dec. 2018, pp. 397–402.

[5] N. T. Nguyen and K. Lee, "Deep learning-aided tabu search detection for large MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4262–4275, June 2020.

[6] H. He *et al.*, "Model-driven deep learning for mimo detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, Feb. 2020.

[7] A. M. Tulino, S. Verdú *et al.*, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[8] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive mimo," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5635–5648, 2020.