PRVNet: Variational Autoencoders for Massive MIMO CSI Feedback

Mostafa Hussien
École de technologie supérieure
Université du Québec
Montreal, Canada
Email: mohussei@partner.ciena.com

Kim Khoa Nguyen
École de technologie supérieure
Université du Québec
Montreal, Canada
Email: KimKhoa.Nguyen@etsmtl.ca

Mohamed Cheriet
École de technologie supérieure
Université du Québec
Montreal, Canada
Email: Mohamed.Cheriet@etsmtl.ca

Abstract—In a frequency division duplexing multiple-input multiple-output (FDD-MIMO) system, the user equipment (UE) send the downlink channel state information (CSI) to the base station for performance improvement. However, with the growing complexity of MIMO systems, this feedback becomes expensive and has a negative impact on the bandwidth. Although this problem has been largely studied in the literature, the noisy nature of the feedback channel is less considered. In this paper, we introduce PRVNet, a neural architecture based on variational autoencoders (VAE). VAE gained large attention in many fields (e.g., image processing, language models, or recommendation system). However, it received less attention in the communication domain generally and in CSI feedback problem specifically. We also introduce a different regularization parameter for the learning objective, which proved to be crucial for achieving competitive performance. In addition, we provide an efficient way to tune this parameter using KL-annealing. Empirically, we show that the proposed model significantly outperforms state-of-the-art, including two neural network approaches. The proposed model is also proved to be more robust against different levels of noise.

I. INTRODUCTION

The massive-input massive-output (MIMO) system is considered as one of the main enabling technology for the fifth-generation, 5G, wireless communication networks. Accordingly, both academia and industry have prioritized the research on this direction. One of the rich and rapidly developing research points is channel state information (CSI) feedback. In MIMO systems, a base station (BS) might be equipped by hundreds of antennas. This can help to reduce the multiuser interference and increasing the cell throughput. However, this can mainly require BS to perform precoding at its side which requires an access to CSI at BS. The uplink CSI can be acquired by means of channel estimation while the downlink CSI should be sent to the BS from the user equipment (UE). Unfortunately, in the massive MIMO systems, this feedback is huge and the bandwidth incurred for sending this feedback is unacceptable.

To overcome this challenge, the UE should compress the CSI matrix before sending it back to the BS. The compressed form of the CSI matrix should preserve enough information about the original CSI matrix. Accordingly, the BS can reconstruct the original CSI matrix with high level of accuracy. Having an inaccurate reconstruction for the CSI matrix at the

BS results in negative consequences on the system performance. The problem then is how to compress the CSI matrix in a way that preserves most of its features and information. In addition, the compression and decompression operations should be completed in a real-time. The encoder part also should not consume much space and power since it resides on the UE side which might has limited space and power resources.

The problem of CSI feedback has been extensively studied in the literature. Traditional methods [1], [2], [3] considered compressed sensing (CS) technique to alleviate the problem. This technique requires the CSI matrix to maintain a high degree of sparsity which is not always fulfilled in the practical systems. In addition, most CS algorithms are iterative; thus, they suffer from slow performance.

On the other hand, artificial intelligence (AI) and deep learning (DL) have shown great power in addressing some nonlinear wireless communication problems [4]. A line of work has utilized AI and DL techniques to solve the CSI feedback problem. The authors in [5] opened the door for applying DL techniques in CSI feedback problem. They presented *CsiNet*, a convolutional neural network architecture with skip connections in the decoder part. The superiority of *CsiNet* performance has been proven against traditional CS-based techniques. However, *CsiNet* is a point estimation model which means that for each dimension in the codeword, the model learns one value. This implies that any noise in this value, can largely hurt the reconstruction quality at the BS. In contrast with *CsiNet*, our model estimates a distribution parameters for each dimension, namely, mean and variance for a Gaussian distribution. Then, we sample from this distribution, as shown in the next sections. This makes our codewords more robust against the noise and the decoder has the capacity to reconstruct the received codewords even in the presence of certain level of noise.

The authors in [6] exploited the temporal and frequency correlations of wireless channels. They presented a model called, *CSINet-LSTM*, which extends *CsiNet* with long short term memory (LSTM) network. LSTM is a classic type of recurrent neural network that is capable of learning long time-dependencies, temporal correlation, between the input samples.

In [7], the authors proposed a neural network architecture, called *CRNet*, for multi-resolution CSI feedback in massive MIMO. The model is shown to have an improved performance against classic CS-based techniques as well as *CsiNet*. Another extension to *CsiNet* model called *CsiNet+* is introduced in [8]. However, the floating point operations (flops) in *CsiNet+* is much larger than the *CsiNet* which can argue that the improvements come at the cost of the complexity.

We can summarize the limitations in the literature work as: a) most of the work assumed error-free control channel and less attention has been paid to CSI feedback problem in the presence of feedback errors. b) Although their proved power in many domains, no prior work has investigated the power of generative models, especially variational autoencoders (VAE) [9], in CSI feedback problem.

a) *Contribution*: The main contributions of this work can be summarized as follows:

- We introduce a novel partially regularized VAE model, named PRVNet, for CSI feedback problem with new objective function. The new objective function is proved to outperform that of classic VAE.
- We introduce a practical and efficient way to tune the additional parameter introduced in the objective function inspired by Kullback-Leibler (KL)-annealing.
- We introduce the first work, to the best of the authors knowledge, that adapt generative models (i.e., VAE) for CSI feedback problem.
- We consider the CSI feedback in the presence of feedback errors, unlike most of the work in literature that assume ideal control channel. Since we adopt a distribution estimation model, our proposed model is capable of reconstructing the CSI matrices with high accuracy even under noisy control channels.
- The experiment results demonstrate that PRVNet outperforms both classical CS-based techniques and recent DL-based techniques.

The rest of this paper is organized as follows: in section II we present a detailed system model. Section III presents a detailed description for the proposed PRVNet and its architecture and training procedure. The results of the proposed model along with a comparison against other works in the literature are presented in section IV.

II. SYSTEM MODEL

We consider a simple single-cell downlink massive MIMO system with $N_t \gg 1$ transmit antennas at a BS and a single receiver antenna at a UE. The system is operated in OFDM over $\tilde{N}_c$ subcarriers. The received signal at the n_{th} subcarrier, y_n , is given by the following equation:

$$y_n = \tilde{h}_n^H v_n x_n + z_n, \quad (1)$$

where $\tilde{h}_n \in \mathbb{C}^{N_t \times 1}$, $\mathbf{v}_n \in \mathbb{C}^{N_t \times 1}$, $x_n \in \mathbb{C}$, and $z_n \in \mathbb{C}$ denote the channel vector, precoding vector, data-bearing symbol, and additive noise of the n_{th} subcarrier, respectively. Also, assume $\tilde{H} = [\tilde{h}_1 \dots \tilde{h}_{\tilde{N}_c}] \in \mathbb{C}^{\tilde{N}_c \times N_t}$ be the CSI stacked in the spatial

frequency domain. The BS can design the precoding vectors $\{\mathbf{v}_n, n = 1, \dots, \tilde{N}_c\}$ once it receives $\tilde{H}$ feedback.

In the FDD systems, the BS continually receives the channel matrix, $\tilde{H}$, through feedback links. The total number of feedback parameters is $N_t \tilde{N}_c$ which is not allowed for limited feedback links. Although the channel estimation process is a challenging task, we assume that perfect CSI has been acquired through pilot-based training [10] and we are going to focus on the feedback scheme.

To reduce feedback overhead, we propose that $\tilde{H}$ can be sparsified in the angular-delay domain using a 2D discrete Fourier transform (DFT) as follows:

$$H = F_d \tilde{H} F_a^H, \quad (2)$$

where F_d and F_a are $\tilde{N}_c \times \tilde{N}_c$ and $N_t \times N_t$ DFT matrices, respectively. Only small fraction of the elements of H are large components, and the remainder is close to zero. In the delay domain, only the first N_a rows of H contain values because the time delay between multipath arrivals lies within a limited period. Therefore, we can retain the first N_a of H and ignore the remaining rows. We will use H_a to denote $N_a \times N_t$ truncated matrix. The total number of feedback parameters can be reduced to $2N_a N_t$ which remains a large number in the massive MIMO regime. For classical CS-based methods, H_a is sparse enough when $N_t \rightarrow \infty$ which makes H_a does not meet the sparsity requirement with the limited N_t .

We are interested in designing the encoder:

$$s = F_{enc}(H_a), \quad (3)$$

which can transform the channel matrix into an M -dimensional vector (codeword), where $M < N$ and $N = 2N_a N_t$. In this case, we can define the data compression ratio, $\gamma = M/N$. In addition, we have to design the inverse transformation (decoder) from the codeword to the original channel such that:

$$H_a = F_{dec}(s). \quad (4)$$

The CSI feedback approach works as follows. Once the channel matrix $\tilde{H}$ is acquired at the UE side, we perform 2D DFT in (2) to obtain the truncated matrix H_a and then use the encoder (3) to generate a codeword s . The generated code word, s , is returned to the BS which uses the decoder (4) to obtain an approximation for the truncated channel matrix $\hat{H}_a$. The final channel matrix in the spatial-frequency domain can be obtained by performing inverse DFT.

III. PROPOSED PRVNET FOR CSI FEEDBACK

In the following sections, we refer to a set of CSI channel matrices as a dataset X consisting of C different CSI matrices indexed by $c \in \{1, 2, \dots, C\}$.

A. Variational Autoencoders (VAE)

As the name autoencoder implies, VAE consists of two models, namely encoder and decoder models. These models

Fig. 1. Schematic diagram for the PRVNet for CSI feedback.

are trained jointly to minimize the standard VAE objective in (5).

$$\mathbb{E}_{z \sim q(z|x)} [\log p_\theta(x|z)] - KL(q_\phi(z|x)||p(z)). \quad (5)$$

The encoder model, also known as the inference model, computes the function

$$f_\phi(x_c) \equiv [\mu_\phi(x_c), \sigma_\phi(x_c)] \in \mathbb{R}^{2K} \quad (6)$$

where the non-linear function $f_\alpha(\cdot)$ is a neural network architecture (i.e., multi layer perceptron in our work) with parameters α . While $\mu_\phi(x_c)$ and $\sigma_\phi(x_c)$ both are K -dimensional vector representing the mean and variance of a Gaussian distribution. The latent representation, code word, z_c is a K -dimensional vector sampled from this distribution:

$$q_\phi(z_c|x_c) = \mathcal{N}(\mu_\phi(x_c), \text{diag}\{\sigma_\phi^2(x_c)\}). \quad (7)$$

That is, for each CSI matrix x_c in the dataset, the inference model outputs the corresponding variational parameters of variational distribution $q_\phi(z_c|x_c)$ which, when optimized, approximates the intractable posterior $p(z_c|x_c)$.

The sampled codeword then go through the decoder model which is also known as the generative model. The generative model uses the latent representation, z_c , to reconstruct the original input x_c using nonlinear function $p_\theta(x_c|z_c)$. The model then trained to optimize the function given by (5).

The first term in (5) represents the reconstruction loss between the original data point and the reconstruction. While the second term is the KL divergence between the encoder's distribution $q_\phi(z|x)$ and the true distribution $p(z)$. This divergence measures how much information is lost when using q to represent a prior over z and encourages its values to be Gaussian. Since the function in (5) is a lower bound for the log marginal likelihood, it is referred to as the evidence lower bound (ELBO) function. We can note that ELBO is a function in both ϕ and θ .

B. The proposed model (PRVNet)

While the classical VAE trained with the loss function given by (5) is a real powerful generative model, we can argue if we really need all of it's statistical properties for solving the CSI feedback problem. We can find a room for improvement if we are willing to scarify the ability to perform ancestral sampling for a degree. As discussed in subsection III-A, the second term in the loss function (5) introduces a compromise between how close the approximate posterior stays to the prior during learning and our ability to reconstruct the original data from the codeword. To this end, we propose introducing a parameter β , where $\beta \neq 1$. Note that using this parameter, we are no longer optimizing a lower bound on the log marginal likelihood.

Setting $\beta < 1$ means that we push the model to learn better data reconstruction and pays less attention to the prior constraint $\frac{1}{C} \sum_{c=0}^C q(z|x_c) \approx p(z) \approx \mathcal{N}(z; 0, I_K)$. This, in turns, implies that a model trained with $\beta < 1$ will be less able to generate novel CSI matrices by ancestral sampling. In the same time, setting $\beta > 1$ emphasis the importance of the prior distribution constraint than the ability to reconstruct the input from the codeword. It is worth to note that setting β to zero eliminates the prior distribution constraint and reduces the loss function to that of the classical point estimate autoencoders.

Advantageously, our goal is to make good reconstruction at the BS side not to generate novel imagined CSI matrices. Treating β as a free parameter, with $\beta < 1$, therefore can significantly improve the reconstruction results without any additional cost in terms of time or number of model parameters. Accordingly, we propose using the objective function in (8). Since we can interpret the second term as a regularization term, we coin a model trained with (8) by partially regularized VAE network (PRVNet).

$$-\mathbb{E}_{z \sim q(z|x)} [\log p_\theta(x|z)] + \beta \cdot KL(q_\phi(z|x)||p(z)) \quad (8)$$

Fig. 2. Variational autoencoder with reparameterization trick versus classical point estimate autoencoders.

1) *Selecting a value for β* : In this section, we propose an algorithm for selecting the best value of β . In the beginning of training phase, we set the value of $\beta = 0$ and gradually increase the value of β to 1. We linearly anneal the KL term slowly over a large number of gradient updates to ϕ and θ and record the best value of β when the performance reaches the peak [11]. After figuring the best value of β , which we denote here as β_* , we retrain the model with values of β starting from 0 to β_* . If we have limited computation power, we can stop increasing β once we notice a degradation in the validation metric. In this way, training our model does not incur any additional cost than training the traditional VAE model.

2) *Training PRVNet*: Recall that the proposed model optimizes the function in (8) while VAE is trained to optimize the standard ELBO function given in (5). We can obtain an unbiased estimate of (8) by sampling $z_c \sim q_\phi$ and optimize it by stochastic gradient descent. However, the challenge is that we cannot trivially take gradients with respect to ϕ through this sampling. The reparameterization trick [9] eliminates this challenge by sampling $\epsilon \sim \mathcal{N}(0, I_k)$ and reparametrize $z_c = \mu_\phi(x_c) + \epsilon \odot \sigma_\phi(x_c)$. This way, the stochasticity in the sampling process is eliminated and the gradient with respect to ϕ now can be back-propagated through the sampled latent code z_c . The detailed description of the training process is shown in Algorithm 1.

C. Taxonomy of Autoencoders

Variational autoencoders are generative models that learn latent representation for input data. While classic autoencoders assume a deterministic latent space (i.e., point estimate for each dimension in the latent space), in a VAE the latent variable is stochastic samples from a tractable distribution (i.e., usually assumed to be Gaussian distribution).

Maximum-likelihood estimation in a regular autoencoder takes the following form:

$$\begin{aligned} \theta^{AE}, \phi^{AE} &= \arg \max_{\theta, \phi} \sum_c \mathbb{E}_\delta(z_c - g_\phi(x_c)) [\log(p_\theta)(x_c|z_c)] \\ &= \arg \max_{\theta, \phi} \sum_c \log(p_\theta)(x_c|g_\phi(x_c)) \end{aligned} \quad (9)$$

Algorithm 1: VAE-SGD Training PRVNet for CSI feedback with stochastic gradient descent.

Input: Dataset X consisting of C CSI matrices

Randomly initialize θ and ϕ ;

while not converged do

 Sample a batch of CSI channels $\mathcal{B}$;

forall $c \in \mathcal{B}$ **do**

 Sample $\epsilon \in \mathcal{N}(0, I)$;

 Compute z_c using the reparameterization trick;

 Compute noisy gradient $\nabla_\theta \mathcal{L}$ and $\nabla_\phi \mathcal{L}$ using the sampled z_c ;

end

 Average noisy gradient for a batch;

 Update θ and ϕ using stochastic gradient descent;

end

Return θ and ϕ

We can note from (9) that classical autoencoder effectively optimizes the first term in the VAE objective using a delta variational distribution. This means that $q_\phi(z_c|x_c) = \delta(z_c - g_\phi(x_c))$, and hence it does not regularize $q_\phi(z_c|x_c)$ toward any distribution like VAE. We can also note that $\delta(z_c - g_\phi(x_c))$ is a delta distribution with mass only at the output $g_\phi(x_c)$. Contrast this to what happens in VAE, where the learning is done using a variational distribution (i.e., $g_\phi(x_c)$ generates the parameters of a certain tractable distribution, the mean and variance in the case of Gaussian distribution). This implies that VAE has the ability to capture per-data-point variances in the latent state z_c . One of the main concerns in autoencoders is the high possibility of overfitting since the network learns to put all the probability mass to the non-zero entries in x_c . By introducing dropout [12] at the input layer, the classical autoencoder is less prone to overfitting. Fig. 2 shows the main difference between point estimate autoencoders and VAE.

IV. SIMULATION RESULTS AND ANALYSIS

A. Experiment Setup

We considered two types of scenarios as given in [5]: the outdoor scenario at 300MHz and the indoor scenario at 5.3GHz. The channels are generated following the default settings of COST2000 [13]. At the BS, a uniform linear array (ULA) with $N_t = 32$ is considered. For the FDD system, we set $N_c = 1024$ in frequency domain and $N_\alpha = 32$ in angular domain. The dataset contains 150,000 independently generated channels divided into three parts. The train, validation, and testing parts consist of 100,000, 30,000 and 20,000 channel matrices. We set the batch size to 128. The model weights are initialized according to He initialization [14]. We optimize the model using Adam optimizer [15] with 0.1 learning rate for 1000 epochs. The function proposed in (8) is used as a loss function at the output layer.

B. Performance of PRVNet

We compare the performance of PRVNet with three state-of-the-art CS-based methods, namely, Lasso L_1 -solver [16],

TABLE I
COMPARISON OF NMSE (dB) FOR DIFFERENT METHODS.

CR	Methods	NMSE (db)	
		Indoor	Outdoor
1/4	LASSO	-7.59	-5.08
	BM3D-AMP	-4.33	-1.33
	TVAL3	-14.87	-6.90
	CSINet	-17.36	-8.75
	CRNet	-26.99	-12.71
	PRVNet (our)	-27.7	-13.9
1/16	LASSO	-2.72	-1.01
	BM3D-AMP	0.26	0.55
	TVAL3	-2.61	-0.43
	CSINet	-8.65	-4.51
	CRNet	-11.35	-5.44
	PRVNet (our)	-13	-6.1
1/32	LASSO	-1.03	-0.24
	BM3D-AMP	24.72	22.66
	TVAL3	-0.27	0.46
	CSINet	-6.24	-2.81
	CRNet	-8.93	-3.51
	PRVNet (our)	-9.52	-4.23
1/64	LASSO	-0.14	-0.06
	BM3D-AMP	0.22	25.45
	TVAL3	0.63	0.76
	CSINet	-5.84	-1.93
	CRNet	-6.49	-2.22
	PRVNet (our)	-6.9	-2.53

TABLE II
THE EFFECT OF ANNEALING β TO DIFFERENT VALUES ON INDOOR SCENARIO FOR A COMPRESSION RATIO 1/4.

β annealing strategy	NMSE (db)
No annealing	-25.83
Annealing β to the maximum ($\beta=1$)	-26.32
Annealing β to 0.3	-27.7

TVAL-3 [17], and BM3D-AMP [18]. In addition, we compare the performance of our proposed model to two recent deep learning-based methods, namely, CsiNet [5] and CRNet [7]. In order to evaluate the performance of different methods, we measure the distance between the original CSI matrix, H_a , and the reconstruction image, $\hat{H}_a$, by means of normalized mean square error:

$$NMSE = \mathbb{E} \left(\left\| H_a - \hat{H}_a \right\|_2^2 / \|H_a\|_2^2 \right) \quad (10)$$

Table I shows the performance of the proposed PRVNet against the state-of-the-art methods. We can see that PRVNet outperforms all classical CS-based methods as well as the state-of-the-art deep learning-based methods. PRVNet with the proposed loss function is capable of capturing CSI features to increase the reconstruction accuracy at the BS.

The effect of β annealing is demonstrated in table II. We can see that the model achieved the highest NMSE when no β -annealing is applied. Under the same dataset and compression ratio, the model achieved lower NMSE β being annealed to 1. The best NMSE has been achieved with annealing β from 0 to 0.3 and complete the training epoch without further increase in the value of β . Although, this value might be sub-optimal

TABLE III
THE ROBUSTNESS OF THE PROPOSED PRVNET UNDER DIFFERENT NOISE CONDITIONS.

Different values of σ	NMSE (db)
$\sigma = 0$	-27.7
$\sigma = 0.05$	-26.56
$\sigma = 0.1$	-25.81
$\sigma = 0.15$	-25.6
$\sigma = 0.2$	-24.95

compared to a thorough grid search. However, the proposed algorithm is much more efficient, and gives us competitive empirical performance.

To further evaluate the robustness of the proposed method under different noise conditions, we simulate the noisy control channel by adding a random Gaussian noise to the codeword and pass it through the decoder network.

$$z = z + \epsilon, \quad (11)$$

where $\epsilon \sim \mathcal{N}(0, \sigma_n)$ and $\sigma_n \in \{0.05, 0.1, 0.15, 0.2\}$. In each case, we observe the degradation of the NMSE. The results are shown on table III in which we can see that the proposed PRVNet is robust against different noise-levels. We notice a slow increase in the NMSE with the noise variance which indicate that the codewords generated by the proposed PRVNet can still convey the information of the original CSI matrix under noisy control channels. This can be attributed to the fact that PRVNet, unlike other models in the literature, learns a distribution for each dimension in the codeword. This makes the effect of the noise much less than point estimation models because even with noise, a value in the codeword may still look as been sampled from the learnt distribution.

V. CONCLUSION

In this paper, a novel deep learning model named PRVNet was proposed for downlink channel state information (CSI) feedback in massive MIMO-FDD systems. The PRVNet modifies the variational autoencoder objective to fit the special characteristics of CSI feedback problem. The codewords generated by PRVNet are proved to be robust against noise. The performance of the proposed model has been evaluated and proved to outperform state-of-the-art models. Also, extensive experiments showed the robustness of the proposed model in different noise conditions.

REFERENCES

- [1] P.-H. Kuo, H. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2012, pp. 492–497.
- [2] L. Lu, G. Y. Li, D. Qiao, and W. Han, "Sparsity-enhancing basis for compressive sensing based channel feedback in massive MIMO systems," in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.
- [3] W. Shen, L. Dai, Y. Shi, X. Zhu, and Z. Wang, "Compressive sensing-based differential channel feedback for massive MIMO," *Electronics Letters*, vol. 51, no. 22, pp. 1824–1826, 2015.
- [4] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, 2017.

- [5] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.
- [6] C. Lu, W. Xu, H. Shen, J. Zhu, and K. Wang, "MIMO channel information feedback using deep recurrent network," *IEEE Communications Letters*, vol. 23, no. 1, pp. 188–191, 2019.
- [7] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.
- [8] "Convolutional neural network-based multiple-rate compressive sensing for massive mimo csi feedback: design, simulation, and analysis."
- [9] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [10] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, 2014.
- [11] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara, "Variational autoencoders for collaborative filtering," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 689–698.
- [12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [13] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 92–99, 2012.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on Imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
- [15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [16] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [17] L. Chengbo, W. Yin, and Y. Zhang, "TV minimization by augmented Lagrangian and alternating direction algorithms," 2009.
- [18] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.