Deep Learning-Based Anomaly Detection using Hybrid Loss

Michael Onyekwelu, Mingyu Jang, and Dongweon Yoon
Department of Electronic Engineering
Hanyang University, Seoul, Korea
dwyoon@hanyang.ac.kr

***Abstract*—Recently, deep learning-based (DL) automatic modulation classification (AMC) has been extensively studied in cooperative and non-cooperative communication contexts such as cognitive radio and spectrum surveillance. One of the drawbacks of DL-based AMC is its susceptibility to anomalous or interfering signals. In this paper, we propose a DL-based anomaly detection for AMC, utilizing an autoencoder to process the in-phase and quadrature components of a received signal. In order to detect anomalies, we employ a hybrid loss, a combination of the autoencoder's reconstruction loss and the Mahalanobis distance of the latent space embedding of the training vector and each input instance. Through computer simulations, we show that the proposed model has superior detection performance with less computational complexity compared to the conventional DL-based model.**

***Index Terms*—Anomaly Detection, Autoencoder, Automatic Modulation Classification**

I. Introduction

Automatic modulation classification (AMC) is a task that identifies the modulation type of the received signal without any prior knowledge of the communication parameters. AMC plays a significant role in cooperative and non-cooperative communication contexts such as cognitive radio and spectrum surveillance. Recently, deep learning (DL) has been extensively applied in AMC, and it is reported in many places in the literature that DL-based AMC has superior classification performance compared to traditional AMC methods such as feature-based AMC [1]–[3]. However, the classification performance of the DL-based AMC can be degraded by unexpected anomalies. Additionally, anomalies in the communication signals can significantly impact AMC accuracy and may have broader consequences on overall network security.

Therefore, detecting anomalies is crucial to the task of AMC in both cooperative and non-cooperative contexts to ensure better performance and reliable communication while efficiently managing and utilizing the radio frequency spectrum [4], [5]. However, despite the critical concerns raised by the presence of anomalies in communication signals, little research has been conducted on detecting anomalies in this field. In [6], the vector extracted from the final fully connected layer of pre-trained convolutional neural network (CNN) models trained with constellation images were used for anomaly detection. However, due to the overhead involved in converting in-phase and quadrature (IQ) sequences to images, this approach introduces a noticeable increase in computational complexity.

In this paper, we propose a DL-based anomaly detection for AMC by utilizing a hybrid loss that incorporates the reconstruction loss and the Mahalanobis distance within the latent space embedding of an autoencoder. As inputs to the model, we use the received IQ sequence. Through computer simulations, we show that the model outperforms the conventional model in terms of detection metrics while also exhibiting significantly lower computational complexity, validating its effectiveness for practical applications.

The rest of this paper is organized as follows: Sections II and III describe the DL-based anomaly detection and simulation results, respectively, and Section IV concludes the paper.

II. DL-based Anomaly Detection

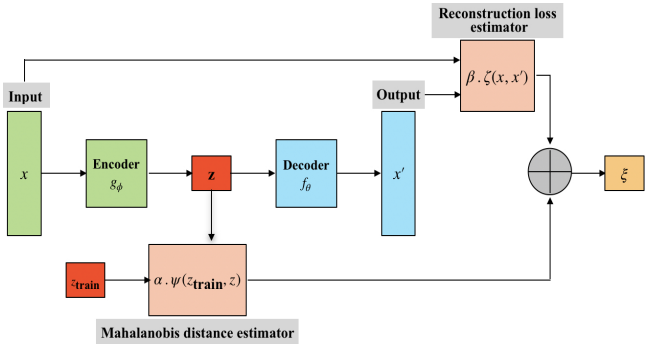


Fig. 1: Mahalanobis-based Autoencoder Architecture

Figure 1 depicts an illustration of the Mahalanobis-based autoencoder (MAE) model's architecture. The model consists of two primary components: an autoencoder and the Mahalanobis distance estimator. The autoencoder comprises three parts, namely the encoder, bottleneck layer, and decoder, and is designed to learn the task of data compression while simultaneously minimizing reconstruction loss. This allows it to map the input data into a lower-dimensional latent space at the bottleneck layer and then reconstruct it back to its original form. The encoder, decoder, and reconstruction loss of the autoencoder can be expressed mathematically as follows

$$z = g_{1\phi}(g_{2\phi}(g_{3\phi}(\cdots g_{n\phi}(x)))) \tag{1}$$

$$x' = f_{1\theta}(f_{2\theta}(f_{3\theta}(\cdots f_{n\theta}(z)))) \tag{2}$$

$$\zeta(x, x') = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)} - x'^{(i)})^2 \tag{3}$$

where $x, x', z, N, g_{i\phi}$, and $f_{i\theta}$ are the input, output (reconstructed input), bottleneck layer vector, number of samples, encoder, and decoder weight matrices, respectively. On the other hand, the Mahalanobis distance estimator utilizes the Mahalanobis distance metric to measure the dissimilarity between the encoded data representations and a learned distribution of normal and anomalous data instances. The Mahalanobis distance can be expressed as

$$\psi(z_{\text{train}}, z) = \sqrt{(z - \hat{\mu}_{z_{train}})^T \hat{\Sigma}^{-1}_{z_{train}} (z - \hat{\mu}_{z_{train}})} \tag{4}$$

where $\hat{\mu}_{z_{train}}$ and $\hat{\Sigma}_{z_{train}}$ are the mean vector and covariance matrix of the encoded training dataset distribution, respectively. A high Mahalanobis distance between a point and the distribution indicates a higher likelihood of the point being an anomaly. Therefore, in this paper, to effectively detect anomalies, we combine the autoencoder's reconstruction loss with the Mahalanobis distance between the encoded training dataset distribution and each input data instance, forming a hybrid loss, which can be mathematically expressed as

$$\xi = \alpha.\psi(z_{\text{train}}, z) + \beta.\zeta(x, x') \tag{5}$$

where α and β are parameters estimated using a validation set of in-distribution samples. α and β can be mathematically expressed as

$$\alpha = \frac{1}{\sigma(\psi(z_{\text{train}}, z_{\text{valid}}))} \quad \beta = \frac{1}{\sigma(\zeta(x_{\text{valid}}, x'_{\text{valid}}))} \tag{6}$$

where $\sigma(\psi(z_{\text{train}}, z_{\text{valid}}))$ and $\sigma(\zeta(x_{\text{valid}}, x'_{\text{valid}}))$ denotes the standard deviation of the Mahalanobis distance between the encoded training and validation dataset distribution and the standard deviation of the reconstruction loss on the validation dataset. This normalization prevents either of the components constituting the hybrid loss from dominating the overall hybrid loss. With an IQ sequence as input to the MAE model, we anticipate that anomalous and normal datasets will result in high and low hybrid loss values, respectively. Therefore, we set a threshold γ to detect anomalous data. When the hybrid loss value is higher than γ, the input data is declared an anomaly, and vice versa when the loss is lower.

III. Simulation Results

In this section, we show the anomaly detection performance through computer simulations. The dataset for anomaly detection is subdivided into two: the normal and anomalous datasets. In this paper, we consider the normal dataset which comprises 3 modulation schemes, including binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), and 8-phase shift keying (8PSK), and the anomalous dataset which comprises 2 modulation schemes, including 16-quadrature amplitude modulation (16QAM) and 64-quadrature amplitude modulation (64QAM). For both datasets, we consider an additive white Gaussian noise (AWGN) channel with a signal-to-noise ratio (SNR) range of -5 dB to 10 dB considering a 1 dB interval. We generate the dataset comprising 1000 samples for each SNR per modulation scheme, resulting in a dimension of (48000×1000) for the normal dataset and (32000×1000) for the anomalous dataset, respectively.

To achieve higher convergence and overall performance while training deep learning (DL) models, the input dataset must be balanced. Therefore, to prepare our dataset for training, we first split it along its real and imaginary parts and then concatenate it along its row, resulting in the normal and anomalous datasets of dimensions (48000×2000) and (32000×2000), respectively. Next, we scale the concatenated dataset using a min-max scaler with a minimum and maximum value of 0 and 1, respectively. After scaling, the normal dataset was divided into training, validation, and test datasets with dimensions of (32000×2000), (6000×2000), and (10000×2000), respectively. These datasets serve as inputs to the MAE model.

Simulations were conducted on an NVIDIA GeForce RTX A6000 GPU with 48GB of VRAM using Python 3.9.13, PyTorch 2.0.1+ Cuda117. The tabulated architecture and hyperparameter details for our model are provided in Table I and Table II, respectively. As seen in Table I, the model is a very simple model with only 3 layers at the encoder and decoder.

TABLE I: Model Architecture

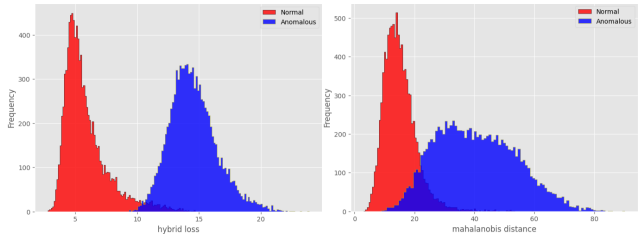
Layer Name	Layer	Size
Encoder	Input Layer	1×2000
	Hidden Layer 1	1×1240
	Hidden Layer 2	1×245
	Hidden Layer 3	1×84
Decoder	Hidden Layer 4	1×84
	Hidden Layer 5	1×245
	Hidden Layer 6	1×1240
	Output Layer	1×2000
Bottleneck	Bottleneck Layer	1×12

TABLE II: Hyperparameters

Hyperparameter	Value
Learning Rate	$3.33e^{-5}$
Batch Size	128
Number of Epochs	35
Optimizer	Adam
Activation function	LeakyReLU
Loss	MSE

Figure 2 depicts the histogram distributions of the normal and anomalous datasets for the models. The histogram for the MAE and ResNet50-based CNN model in [6] are given in Figure 2a and 2b, respectively. And Figure 3 depicts the AUROC of the MAE and ResNet5-based CNN in [6]. To obtain the decision metrics, we first generate the hybrid loss distribution histograms from normal and anomalous test datasets as shown in Figure 2, estimating the AUROC as

shown in Figure 3. Subsequently, we measured the distance between the neutral line (the diagonal line connecting [0, 0] and [1, 1] points) and the coordinates of the true positive rate (TPR) and false positive rate (FPR) obtained from the AUROC analysis. We obtain the optimal value of the threshold γ when the distance attains the highest value. Using this γ, we compute other detection metrics to compare the performance between the proposed MAE and the ResNet50-based CNN in [6]. Table III presents various detection metrics including TPR, true negative rate (TNR), positive prediction value (PPV), false negative rate (FNR), and F1-score. As seen in Table III, the detection performance of the proposed model outperforms that of the ResNet50-based CNN in [6] for all detection metrics. The high detection accuracy of the proposed model can be attributed to the use of the compressed vector at the bottleneck layer, which contains salient features of the dataset for loss estimation.



(a) MAE Model (b) ResNet50-based CNN Model

Fig. 2: Histogram distributions of the normal and anomalous datasets for the Models.

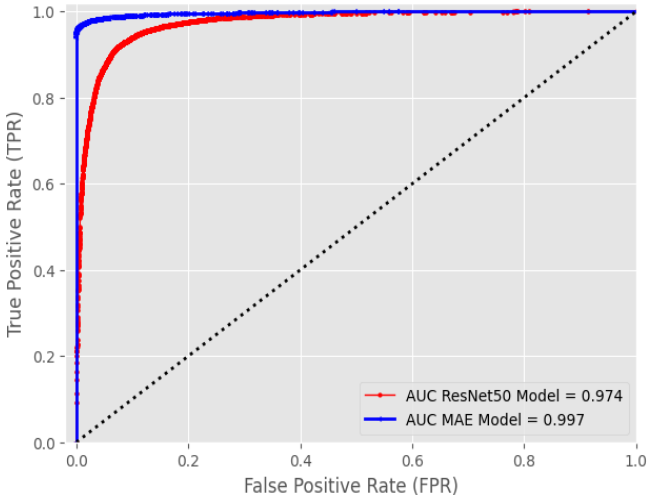


Fig. 3: AUROC of MAE and ResNet50-based CNN Models

TABLE III: Model Comparison using Detection Metrics

Model	TPR	TNR	PPV	FNR	F1-Score	AUROC
MAE	0.971	0.988	0.988	0.029	0.979	0.997
ResNet50	0.957	0.867	0.852	0.043	0.902	0.974

We further compare the computational complexity of the ResNet50-based CNN in [6] and the MAE model. The ResNet50 model comprises three processes that mainly constitute its complexity: image generation, PCA computation, and Mahalanobis distance computation. On the other hand, the proposed MAE basically comprises only one process that constitutes its complexity, which is the computation of the Mahalanobis distance. Therefore, the proposed MAE model has much less computational complexity than the ResNet50-based CNN model in [6].

IV. Conclusion

In this paper, we proposed a DL-based anomaly detection for AMC. The DL model combines the Mahalanobis distance within the latent space embedding of an autoencoder, along with the autoencoder reconstruction loss, to efficiently detect anomalies. Through computer simulations, we showed that the proposed model outperformed the conventional DL-based method across all performance metrics, including TPR, TNR, PPV, FNR, F1-score, and AUROC, while maintaining a lower level of computational complexity.

References

[1] J. Lee, J. Kim, B. Kim, D. Yoon, and J. Choi, "Robust automatic modulation classification technique for fading channels via deep neural network," *Entropy*, vol. 19, no. 9, p. 454, Aug. 2017.

[2] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 718-727, Mar. 2019.

[3] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, "Deep learning-based signal modulation identification in OFDM systems," *IEEE Access*, vol. 7, pp. 114631-114638, 2019.

[4] Y. -J. Tang, Q. -Y. Zhang, and W. Lin, "Artificial neural network based spectrum sensing method for cognitive radio," in *Proc. WiCOM*, Chengdu, China, 2010, pp. 1-4.

[5] D. B. Schuster, "International telecommunication union — 150 years of history: adaptation to change and the opportunity for reform," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 10-15, Jun. 2015.

[6] M. A. Conn and D. Josyula, "Radio frequency classification and anomaly detection using convolutional neural networks," in *Proc. RadarConf*, Boston, MA, USA, 2019, pp. 1-6.