# Automatic Modulation Classification using Relation Network with Denoising Autoencoder

Hyeongyun Kim*, Adnan Shahid†, Jaron Fontaine†, Eli De Poorter†, Ingrid Moerman†, Haewoon Nam*
* Department of Electronics and Communication Engineering, Hanyang University, Seoul, Korea
† IDLab, Department of Information Technology, Ghent University - imec, Belgium

***Abstract*—With the advent of deep learning (DL), various automatic modulation classification (AMC) methods using deep learning architectures achieved significant performance improvements compared to conventional algorithms. Aiming to achieve high classification accuracy, DL-based AMC algorithms require numerous annotated training samples for each modulation class to extract salient features, but it is hardly applicable in real-world AMC applications. To tackle the annotated data scarce issue, this paper proposes a novel few-shot learning (FSL) framework, which introduces a relation network with a denoising autoencoder to extract feature representations effectively from a limited dataset. The experimental result demonstrate that the proposed method can achieve higher classification accuracy compared to the conventional FSL algorithm for signal modulation recognition, especially in low signal to noise ratio conditions.**

***Index Terms*—Few-shot learning, modulation recognition, relation network, denoising autoencoder**

## I. Introduction

Automatic modulation classification (AMC) refers to a technology that classifies a modulation type from a received signal without prior knowledge of the communication parameters and plays an important role as the demand of cognitive radio in wireless communications is increasing. Traditionally, AMC has been performed by Likelihood-based (LB) and Feature-based (FB) approaches [1]. For LB algorithms, the probability density function (PDF) of the received signal is utilized to evaluate the likelihood of each possible hypothesis, and the maximum value of the likelihood function gives result in a classified modulation [2]. However, it suffers from high computational complexity and requires knowledge of the signal environment and channel parameters in advance, which results in severe performance degradation problems when the LB algorithm is applied to different wireless channel environments. The alternative approach is FB which extracts the features from the received signal and the decision is made based on their differences [3]. FB approaches have less computational complexity than LB approaches, but still require domain-expertise knowledge.

Deep learning (DL) has shown a great performance increase in the fields of image classification, speech recognition and signal classification. Specifically, a deep neural network (DNN) can automatically extract appropriate features of the received signal, and DNN-based AMC algorithms have shown exceptional classification performance. In [4], the authors introduced the convolutional neural network (CNN)-based approach for modulation classification and achieved outstanding classification performance compared to that of the conventional methods. Although DL-based AMC approaches have shown excellent performance, they face several problems. It requires collecting a training dataset with a huge number of labeled samples in order to extract salient features effectively, which is very expensive and time-consuming. Furthermore, the trained model should be re-trained, if a new class is added or the entire classification class changes. To tackle this issue, transfer learning, a learning technique that reuses a trained model in a data-rich field, has emerged to train a model in a field where training data is scarce. In [5], the authors proposed an adversarial transfer learning architecture reducing the difference between data distributions. However, transfer learning algorithms still requires a number of labeled samples.

Recently, few-shot learning (FSL) algorithm, based on the idea that human beings are good at recognizing and distinguishing objects well even with very few instances, has been in the spotlight. FSL algorithms can train model using a small number of labeled samples and can perform classification tasks on new classes by generalization [6]. The authors in [7], proposed a new FSL framework, Attention Relations Network (ARN), to utilize channel and spatial attention, thereby effectively extracting features from a support set. In addition, a novel network architecture called AMCRN, an AMC algorithm using a relational network, is proposed in [8]. The architecture reached a maximum classification accuracy of 93% and showed a performance improvement of 10 to 50% compared to the existing baseline schemes. Although existing FSL algorithms can be trained with a small number of data and showed excellent classification performance, they have the limitation that a wide range of SNR signal data set is required to obtain good classification performance at low SNR region.

In this paper, motivated by the denoising autoencoder, a DL architecture that can improve robustness to changes in the input by intentionally introducing noise into the signal, we propose a relation network with a denoising autoencoder (RNDAE) to effectively utilize a limited dataset.

## II. Few-shot learning

In DL, the number of data sample is directly related to the classification accuracy, but in a realistic task, this number is often insufficient. To solve this limited dataset problem, data augmentation methods exist at the data perspective, and Un/Semi-supervised learning, transfer learning, and meta learning methods exist at the network perspective. FSL is

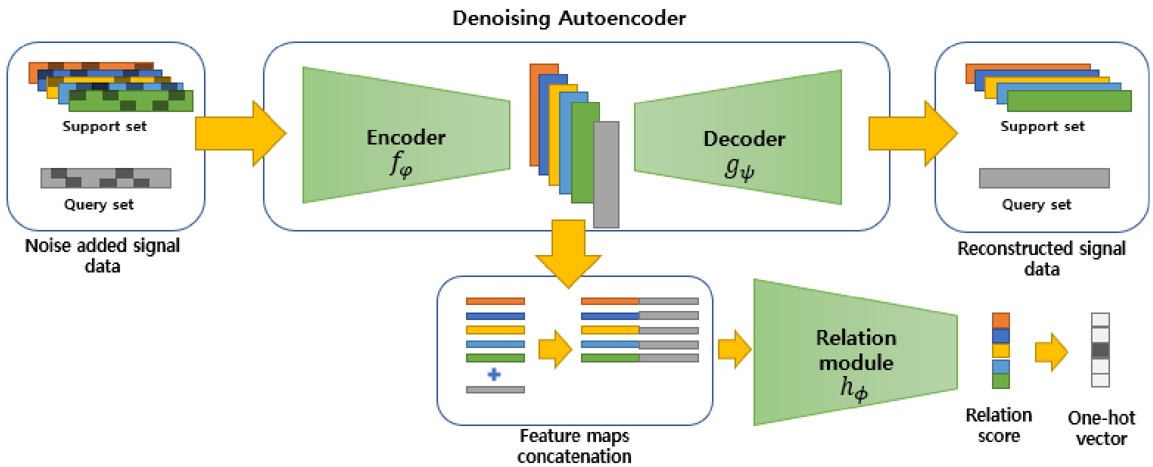


Fig. 1. RNDAE architecture for a 5-way 1-shot problem with one query example. In this case, an I/Q data ($2 \times 128$) of each of the five modulation classes is used as a support set, and another I/Q data out of five classes is used as a query sample.

a method of training a network with a very small number of data samples and is called the 'C-way K-shot' problem. $C$ is the number of classes, and $K$ is the number of data per class. As $K$ increases, the number of instances per class increases, so the performance improves. FSL assumes a very small $K$, and most studies use 5-way 1-shot and 5-way 5-shot as benchmarks. Few shot learning maximizes generalization performance by using episodic training-type meta-learning so that FSL model works well even on new classes which have not been trained previously. Episodic training performs 1 episode with a support set and a query set of C-way K-shot on a training dataset. For each episode, a network is trained using a randomly selected support and query set, and during the test session, a support and query set is also randomly selected from classes that are not used in the training dataset.

## III. Proposed Network

### A. Network Description

The proposed network is mainly composed of two parts, one is a denoising autoencoder and the other is a relation module, as illustrated in Fig.1 and the details of each layer in the RNDAE architecture is described in table I. The denoising autoencoder is used to learn how to reconstruct noise added signal to its original clean signal, so the learned encoder embed the same class signal samples at similar location in latent space, even if noise is added to the signal. The encoder part of the denoising autoencoder consists of four convolution blocks, and each block contains 64 filters of $1{\times}7$ size, followed by batch normalization, ReLU activation, and maxpooling layer. The decoder consists of four transposed convolution blocks to reconstruct denoised signal in the reverse order of the encoder. Noise added samples in support set $S = \{(x_i, y_i)\}_{i=1}^{m}$ $(m = C \times K)$ and query set $Q = \{(x_j, y_j)\}_{j=1}^{n}$ are fed into the denoising autoencoder. The input data format entering into the denoising encoder is $(C + 1) \times 2 \times 128$, one data of each class in support set and one query data with size of $2 \times 128$ each. The encoder extracts the salient feature of the signals and dimensionally reduced feature maps $f_\varphi(x_i)$, $f_\varphi(x_j)$ are obtained. To learn the similarity metric in the relation module $h_\phi$, the feature maps of the support set and query set are concatenated to generate support-query pairs $C(f_\varphi(x_i), f_\varphi(x_j))$ by concatenation operator $C$. Finally, the relation module outputs a relation score $r_{i,j}$ through two convolution blocks and two dense layers. In the $C$-way 1-shot case, the relation score between one query input $x_j$ and support set input $x_i$ passing through the entire network is expressed as

$$r_{i,j} = h_\phi(C(f_\varphi(x_i), f_\varphi(x_j))), \quad i = 1, 2, \ldots, C \tag{1}$$

where relation score $r_{i,j}$ in range of 0 to 1 representing the similarity between $x_i$ and $x_j$. The network is trained to increase the similarity between the same class by reducing the loss between the relation score and the ground truth label.

### B. Objective Function

We introduced two loss functions to improve the training performance and used the mean square error (MSE) for both loss function, one for regressing the relation score $r_{i,j}$ and the other for calculating reconstruction loss of the denoising autoencoder. By adding reconstruction loss, the network can improve robustness to changes in the input. Learning is performed by minimizing the following objective function,

$$L = \lambda L_s + (1 - \lambda) L_r \tag{2}$$

where $L_s$ is regression loss of relation scores and $L_r$ is reconstruction loss of denoising autoencoder. $\lambda$ is weight of loss function and we used $\lambda = 0.8$ for the experiment in this paper. The regression loss $L_s$ is defined as

TABLE I

THE DETAILS OF EACH LAYER IN THE RNDAE ARCHITECTURE

| **Module** | **Layer** | **Output** |
| --- | --- | --- |
| Input Layer | - | (C+1)×2×128 |
| Denoising | Conv(64,1×7,padding=0),BN,Relu, Maxpool(2) | (C+1)×64×1×61 |
| Autoencoder | Conv(64,1×7,padding=0),BN,Relu | (C+1)×64×1×55 |
| | Conv(64,1×7,padding=1),BN,Relu | (C+1)×64×3×51 |
| | Conv(64,1×7,padding=1),BN,Relu | (C+1)×64×5×47 |
| | ConvT(64,1×7,padding=1),Relu | (C+1)×64×3×51 |
| | ConvT(64,1×7,padding=1),Relu | (C+1)×64×1×55 |
| | ConvT(64,1×7,padding=0),Relu | (C+1)×64×1×61 |
| | ConvT(64,2×8,padding=0,stride=2), tanh | (C+1)×2×128 |
| Concatenate | - | C×128×5×47 |
| Relation | Conv(64,1×7,padding=0),BN,Relu, Maxpool(2,4) | C×64×2×10 |
| Network | Conv(64,1×7,padding=0),BN,Relu, Maxpool(2,4) | C×64 |
| | Dense, Relu | C×8 |
| | Dense, Sigmoid | C×1 |

TABLE II

THE DETAILS OF THE SETTINGS FOR THE TRAINING AND TEST DATASETS.

| Parameter | Assignment |
| --- | --- |
| Training set classes | 8PSK,AM-DSB,BPSK,GFSK,PAM4,QAM64 |
| Test set classes | QPSK,AM-SSB,QAM16,CPFSK,WBFM |
| Number of training sets | 6,000 |
| Number of test sets | 100,000 |
| Number of episodes | 100,000 |

$$L_s = \sum_{i=1}^{m}\sum_{j=1}^{n}(r_{i,j} - 1(y_i == y_j))^2 \quad (3)$$

The reconstruction loss $L_r$ is expressed as

$$L_r = \sum_{i=1}^{n} ||x - g_\psi(f_\varphi(x))||^2 \quad (4)$$

## IV. Simulations and Discussions

To evaluate the performance of the proposed RNDAE network, RadioML2016.10a is used as the dataset in our experiment [9]. The dataset contains 11 signal modulations, including BPSK, QPSK, QAM16, QAM64, CPSFSK, 8PSK, GFSK, PAM4, WBFM, AM-SSB, and AM-DSB, each of which consists of $2 \times 128$ size inphase (I) and quadrature (Q) samples. The SNR of the dataset ranges from -20dB to 18dB with an interval of 2dB and there are 1,000 samples for each modulation and SNR. We used only the 18dB dataset when training networks to demonstrate whether the proposed network performs better under limited conditions where only high SNR datasets are available, and used the entire SNR data to test the performance of classification accuracy on each SNR. FSL algorithms basically separates classes for training a network and classes used for test to demonstrate whether the trained network can classify new classes well. Since the experiments were conducted under the conditions of 5-way 1-shot and 5-way 5-shot, we split the entire classes into 6 classes for training set and remaining classes for test set. Detailed settings for the datasets used for training and testing are listed in Table II. To compare the performance of the proposed network, a relation network is used as a benchmark network.

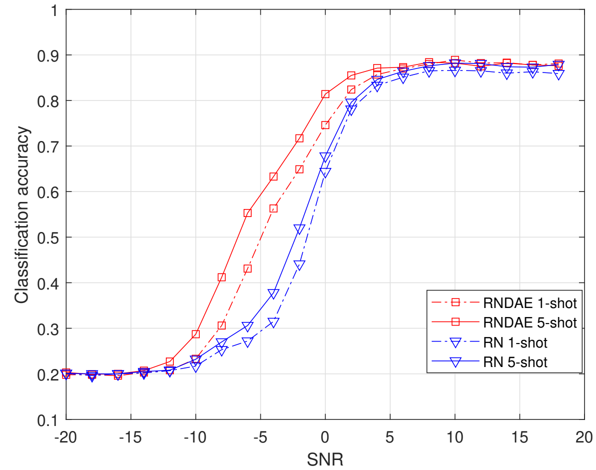


Fig. 2. Performance evaluation of the proposed RNDAE network and relation network under 5-way 1-shot and 5-way 5-shot condition

The performance of proposed RNDAE and relation network is plotted in fig 2. RNDAE and RN trained with 5-way 1-shot and 5-way 5-shot settings similarly converge to 88% classification accuracy when SNR exceeds 4dB. However, in the low SNR region between -10dB and 2dB, there is a performance gap in classification accuracy of 2% to maximum 25%. Additionally, 5-way 5-shot case is trained with 5 times as many support dataset compared to 5-way 1-shot, which improves classification accuracy by 3-10% at low SNR region. Since the denoising autoencoder learns to restore a randomly noise-added signal to a clean original signal, the learned encoder maps signals to a close position in the latent space even if noise is added to the signal, which results in an improved classification performance in low SNR region.

## V. Conclusion

In this paper, a novel network architecture, RNDAE was proposed to improve the performance of AMC utilizing FSL architecture under the condition when only a few high SNR signals in the dataset are available. The experimental results show that the proposed scheme can obtain a classification performance improvement of 2% to maximum 25% compared with baseline scheme particularly in the low SNR region between -12dB and 4dB.

## Acknowledgment

This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008745) supervised by the Korea Institute for Advancement of Technology (KIAT), and also supported in part by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-00964), and also supported by the Fund for Scientific Research Flanders, Belgium, FWO-Vlaanderen, FWO-SB, under Grant 1SB7619N.

## REFERENCES

[1] O. A. Dobre, "Signal identification for emerging intelligent radios: Classical problems and new challenges," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 2, pp. 11–18, 2015.

[2] A. Bouzegzi, P. Ciblat, and P. Jallon, "Maximum likelihood based methods for ofdm intercarrier spacing characterization," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, IEEE, 2008.

[3] O. A. Dobre, R. Venkatesan, D. C. Popescu, *et al.*, "Second-order cyclostationarity of mobile wimax and lte ofdm signals and application to spectrum awareness in cognitive radio systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 26–42, 2011.

[4] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[5] K. Bu, Y. He, X. Jing, and J. Han, "Adversarial transfer learning for deep learning based automatic modulation classification," *IEEE Signal Processing Letters*, vol. 27, pp. 880–884, 2020.

[6] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[7] Z. Zhang, Y. Li, and M. Gao, "Few-shot learning of signal modulation recognition based on attention relation network," in *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 1372–1376, IEEE, 2021.

[8] Q. Zhou, R. Zhang, J. Mu, H. Zhang, F. Zhang, and X. Jing, "Amcrn: Few-shot learning for automatic modulation classification," *IEEE Communications Letters*, vol. 26, no. 3, pp. 542–546, 2021.

[9] T. J. O'shea and N. West, "Radio machine learning dataset generation with gnu radio," in *Proceedings of the GNU Radio Conference*, vol. 1, 2016.