Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 3, 2024

RT-DETR-Based Signal Modulation Recognition with AIFI-Dattention

Minghao Cao[1], Peng Chu[1,*], Hongjie Guo[2], Wei Xing[3]

[1]School of Electronic Information, Xijing University, Xian, China
[2]Xi'anVocational University of Information, Xian, China
[3]Unit 93117, Nanjing, China

[*]Corresponding author

Abstract: In response to the issues of high computational complexity, low accuracy, and cumbersome manual feature extraction steps in traditional machine learning algorithms for communication signal modulation recognition, a communication signal modulation recognition model based on deep learning is proposed. This model can directly recognize the category of communication signals after sampling and is characterized by high recognition accuracy, strong generalization capability, good noise resistance, and a simplified processing flow. It effectively addresses the limitations of traditional algorithms in automatic feature extraction. Through extensive experiments and accurate analysis of communication signal features, an end-to-end model based on the Transformer RT-Detr model is adopted, achieving high recognition accuracy.

Keywords: Feature extraction networks. Signal modulation recognition. Deep learning.

1. Introduction

With the rapid development of wireless communication and signal processing technologies, signal modulation recognition plays a crucial role in fields such as military communication, radio monitoring, and cognitive radio. The task of signal modulation recognition is to extract modulation information from the received signals to determine the signal's modulation type. Traditional signal modulation recognition methods often rely on manually designed feature extraction and classification algorithms, but these methods struggle to cope in complex electromagnetic environments and have limited recognition accuracy[1].

In recent years, with the rise of deep learning, researchers have begun exploring the use of deep neural networks for signal modulation recognition. Deep learning methods can automatically learn and extract features from signals, demonstrating greater robustness and higher recognition accuracy in complex environments. However, existing deep learning-based modulation recognition methods still face trade-offs between model complexity, computational efficiency, and recognition accuracy. Therefore, designing a deep learning model that can maintain high recognition accuracy while reducing computational complexity has become an important research direction.

In this study, we employ a novel RT-DETR model for signal modulation recognition. The RT-DETR model, as a lightweight object detection model, offers high computational efficiency and strong generalization ability. To further enhance the model's performance, we selected ResNet-18 (R18) as the backbone and integrated the AIFI-DAttention module. These improvements aim to enhance the model's feature extraction capabilities and optimize feature fusion efficiency, thereby improving recognition accuracy while keeping the model lightweight.

The main contributions of this study include:

We propose a lightweight signal modulation recognition method based on RT-DETR and improve the model's feature extraction and fusion capabilities through the AIFI-DAttention module.

We designed a comprehensive experimental plan to validate the effectiveness of the proposed method in recognizing different modulation types and conducted a detailed comparative analysis with traditional methods.

The structure of this paper is as follows: Section 2 introduces related work, Section 3 provides a detailed description of our proposed method, and Section 4 presents experimental results and analysis.

2. Related Research

2.1. Traditional Signal Modulation Recognition Methods

Signal modulation recognition plays a crucial role in wireless communication. Early research primarily relied on traditional handcrafted feature extraction and classification algorithms. Common methods include recognition based on statistical features, instantaneous feature analysis, and cyclostationary feature extraction. These methods typically extract frequency-domain, time-domain, or instantaneous features of the signal, such as amplitude, phase, frequency, and instantaneous power, to distinguish modulation types. For instance, methods based on higher-order cumulants (HOC) have performed well in modulation recognition, especially in high signal-to-noise ratio (SNR) scenarios. However, these methods heavily rely on the precision and robustness of feature extraction, and their recognition performance tends to degrade in complex and dynamic electromagnetic environments.

Machine learning methods, such as Support Vector Machines (SVM) and decision trees, have also been widely applied to modulation recognition. These methods classify signals by inputting handcrafted features into the classifiers. However, handcrafted feature extraction has significant limitations, especially when the signals are affected by noise or multipath effects, causing the performance of traditional methods to deteriorate significantly.

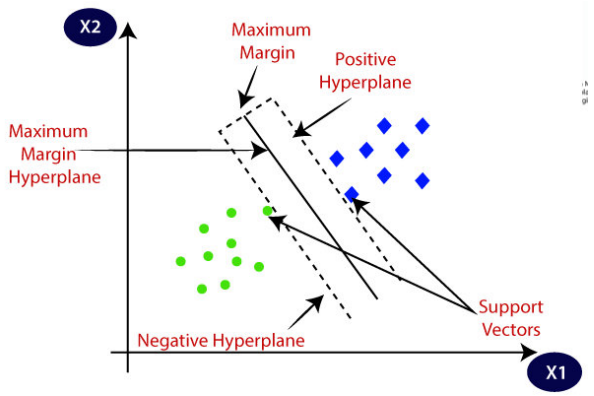

Figure 1. Principle of SVM.

2.2. Deep Learning-Based Signal Modulation Recognition

With the rapid development of deep learning, researchers have begun exploring the use of deep neural networks to automatically extract features from signals, addressing the shortcomings of traditional handcrafted feature extraction. Convolutional Neural Networks (CNNs), known for their excellent performance in image processing, were introduced into signal modulation recognition[2]. O'Shea et al. first proposed using CNNs to directly learn features from the I/Q data of signals and perform modulation classification. This method achieved remarkable results on multiple public datasets.

Subsequently, Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory networks (LSTMs), were introduced to capture temporal information in signals[9]. By modeling the temporal dependencies of the signals, these methods have demonstrated strong robustness in low SNR environments. However, RNN-based models are time-consuming to train and are prone to gradient vanishing or exploding issues.

In recent years, hybrid architectures combining CNNs and RNNs, such as CNN-RNN and CNN-LSTM models, have gained attention. These models improve modulation recognition accuracy by fusing spatial and temporal features. Although these deep learning methods outperform traditional methods, they often require substantial computational resources and training time, posing challenges for real-time applications.

2.3. Lightweight Networks and Transformer Applications

To address the high computational complexity of deep learning models, lightweight network structures have been widely researched. Lightweight CNN models, such as MobileNet and SqueezeNet, reduce the number of parameters and computation, enabling efficient performance on resource-constrained platforms like mobile devices. While these models perform well in specific application scenarios, they still struggle with accuracy when handling complex signal modulation recognition tasks.

The Transformer structure, known for its success in natural language processing, has garnered attention and has been gradually introduced into visual tasks. Its self-attention mechanism can capture global features without relying on temporal sequences, providing new approaches to feature extraction. RT-DETR, a lightweight object detection model that integrates the advantages of the Transformer, can extract rich feature information while maintaining computational efficiency[3], making it highly promising for signal modulation recognition tasks.

2.4. Model Enhancement and Feature Fusion Techniques

To further enhance the performance of deep learning models in signal modulation recognition, researchers have proposed various enhancement and optimization strategies. Attention mechanisms have been widely applied in deep learning, particularly when dealing with signals with complex patterns, as they effectively enhance the expression of important features. AIFI-DAttention, a dynamic attention mechanism, can adaptively adjust attention weights according to the characteristics of the input signal, improving the model's feature extraction capabilities[4].

Moreover, feature fusion techniques have been applied in multimodal signal processing and complex scene recognition. The AIFI-EfficientAdditive module improves traditional additive feature fusion methods, enhancing the synergy between features while maintaining computational efficiency, thereby improving modulation recognition accuracy.

2.5. Contributions of This Research

Building upon the aforementioned research, this study proposes a signal modulation recognition method based on the RT-DETR model. By combining ResNet-18 (R18) as the backbone and introducing the AIFI-DAttention module, our model significantly improves the accuracy and efficiency of signal modulation recognition while maintaining a lightweight structure. We conducted comprehensive testing of the model in various complex channel environments, and the experimental results indicate that our method outperforms traditional methods in terms of recognition accuracy and computational efficiency.

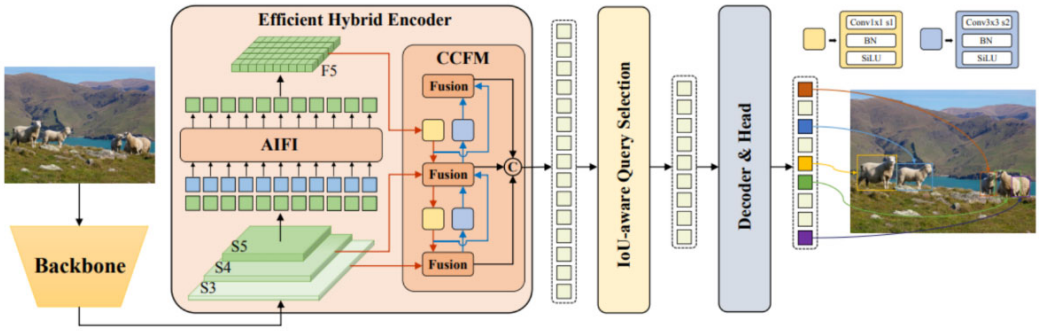

Figure 2. Overview of RT-Detra

3. Model Structure and Dataset

3.1. Model Overview

In this study, we employ the RT-DETR model as the foundation for signal modulation recognition. RT-DETR is a lightweight object detection model, and its Transformer-based architecture effectively captures the global features of input signals, making it well-suited for the task of complex signal modulation recognition. To further enhance model performance, we made the following improvements based on RT-DETR:

Backbone Selection: We chose ResNet-18 (R18) as the backbone of the model. ResNet-18 is a lightweight convolutional neural network architecture that uses residual connections to mitigate the vanishing gradient problem. This enables effective extraction of deep features while keeping the model lightweight.

Dynamic Attention Mechanism (AIFI-DAttention): To improve the model's sensitivity to complex signal features, we integrated the AIFI-DAttention module. This module, based on dynamic attention mechanisms, adaptively adjusts the attention weights of different feature maps, thereby improving feature extraction precision and robustness.

3.2. Data Preprocessing

Before performing signal modulation recognition, we preprocess the input signal data using the following methods:

Normalization: All input signal data is normalized to the [0,1] range to eliminate amplitude differences between signals, reducing their impact on the model[10].

Signal Framing and Overlapping: To capture richer temporal information, the input signal is divided into multiple frames, with a certain overlap between frames. Each frame of the signal is treated as an independent input sample.

Data Augmentation: To increase the diversity of the training data, we applied data augmentation techniques such as random noise addition and frequency shifting[11]. These augmentations help the model generalize better to different channel environments.

3.3. Network Architecture Design

The overall structure of the RT-DETR model is composed of the following parts:

Backbone Network: ResNet-18 serves as the backbone, extracting primary features from the input signal[6]. The network consists of several convolutional layers and residual blocks, with each residual block using skip connections to enhance gradient flow, ensuring stable training of the deep network.

Transformer Encoder: After extracting primary features, the feature maps are input into the Transformer encoder. The self-attention mechanism of the Transformer allows it to effectively capture global information in the signal, enhancing recognition performance[13].

AIFI-DAttention Module: The AIFI-DAttention module is integrated into the network to apply weighted processing to the feature maps output by the Transformer encoder. Its dynamic attention mechanism allows the network to adaptively adjust attention distribution, better capturing key features

Classifier: The final feature map is input into a fully connected layer for classification. This layer outputs the probability distribution corresponding to various modulation signals, with the predicted modulation type being the one with the highest probability.

3.4. Model Training and Optimization

We use the cross-entropy loss function as the optimization objective of the model. The model training process uses the Adam optimizer with the following hyperparameters:

Learning Rate: The initial learning rate is set to 0.001, and a cosine annealing strategy is used to dynamically adjust the learning rate to ensure stability and fast convergence during training.

Batch Size: The batch size is set to 32, striking a good balance between training speed and model performance.

Epochs: The model was trained for 100 epochs. After each epoch, the model's performance was evaluated on the validation set, and the model with the highest validation accuracy was saved as the final model.

Weight Decay: To prevent overfitting, a weight decay strategy with a decay rate of 0.0001 was employed.

Gradient Clipping: To avoid the gradient explosion problem, a gradient clipping strategy was used with a threshold set at 1.0.

Additionally, we applied an early stopping strategy during training. If the validation loss did not decrease for 10 consecutive epochs, the training process was terminated early[12].

3.5. Experimental Setup

To validate the effectiveness of the proposed model, we designed a series of experiments. These experiments were conducted on multiple public datasets, covering different signal-to-noise ratios (SNR) and various modulation types. We compared the performance of models using different backbones (such as ResNet-18) and evaluated the models with and without the AIFI-DAttention. The experimental results were evaluated using metrics such as accuracy, precision, and recall.

3.6. Dataset Generation

The dataset for this study was generated through MATLAB simulations, with the signal-to-noise ratio (SNR) for all signals set at 30 dB. This SNR level represents a low-noise environment, suitable for testing the model's recognition ability under ideal conditions. The dataset generation process is as follows:

Modulation Types: The dataset includes multiple common digital modulation types, including BPSK, QPSK, 8PSK, 16QAM, and 64QAM[5].

Signal Length and Sampling Rate: Each signal length is set to 1024 sample points, with a sampling rate of 1 MHz. The generated signals are stored in complex I/Q data format for subsequent processing and feature extraction.

Data Augmentation: Although all signals have an SNR set at 30 dB, other simulation conditions such as frequency offset and phase shift were introduced through data augmentation, increasing the diversity and complexity of the data.

Dataset Splitting: The generated signal data is split into training, validation, and test sets in a 70%, 15%, and 15% ratio, ensuring an even distribution of modulation types across different datasets.

4. Experimental Process and Results Analysis

4.1. Training Process

In the training process, the data features of communication signals are first extracted using convolution windows in the convolutional neural network (CNN). The categorical cross-entropy loss function is employed, and the loss function $L_i$is defined as follows

$$L_i = -\sum_j t_{i,j} \log\left(p_{i,j}\right) \quad (1)$$

where t represents the true labels, i represents the input data, j represents the categories, and p represents the predicted results. This loss function is commonly used in multi-class classification tasks, such as when using the softmax function as the final output. The optimizer continuously reduces the loss function to update the parameters of the hidden layers. In this case, we selected the Adam optimizer, which is currently one of the most widely used optimizers[7]. According to the literature[16], this optimizer offers advantages such as low memory requirements, simple implementation, and high computational efficiency. The learning rate for the optimizer was set to a fixed value of 0.001, and a dropout rate of 0.5 was used to randomly deactivate neurons, preventing overfitting due to the large number of parameters in the fully connected layer[14].

The confusion matrix in Figure 3 shows the classification performance at a signal-to-noise ratio (SNR) of 30dB. The horizontal axis represents the nine predicted modulation categories, while the vertical axis represents the actual modulation categories. This confusion matrix provides a visual representation of how well the model correctly classifies each modulation type under the given conditions[8].

This process, combining the loss function, Adam optimizer, and dropout techniques, helped achieve better generalization and avoided overfitting during training.

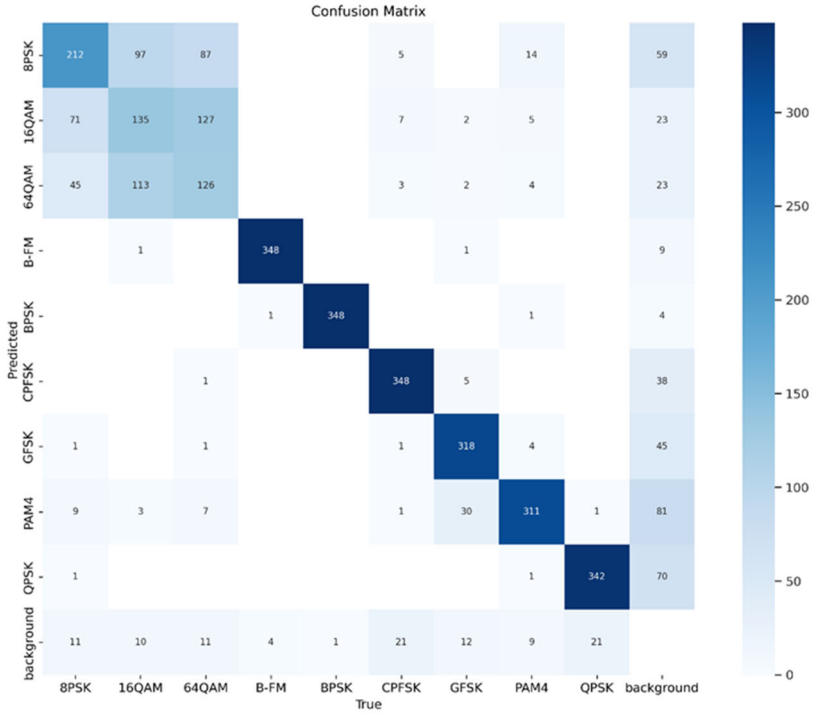

Figure 3. The confusion matrix at a signal-to-noise ratio (SNR) of 30 dB.

4.2. Results Analysis

The neural network was trained for 100 epochs using the training data, while the test data was used to evaluate the model after each epoch. Figure 4 shows the loss reduction curve, where the loss decreases to around 0.2 at its lowest point. Figure 5 illustrates the training accuracy and test accuracy of the model. The training accuracy of the deep neural network reaches a maximum of 71.7%, and the test accuracy reaches a maximum of 72%. The test accuracy increases alongside the training accuracy until both curves stabilize, with no signs of extreme divergence. There were no issues with overfitting or underfitting, and the neural network successfully learned the features of the training data. The test results indicate that the trained model exhibits strong generalization capabilities.

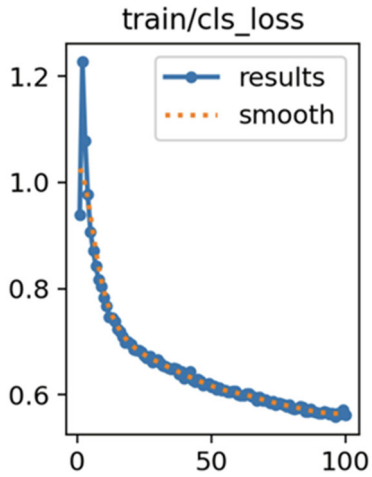

Figure 4. The variation curve of the classification loss.

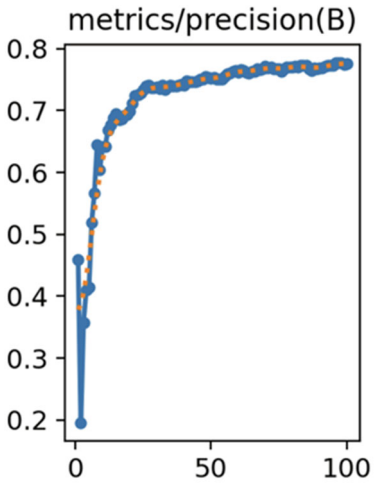

Figure 5. The accuracy of the model during training.

5. Conclusions

This paper conducted extensive experiments to test various hyperparameters, and while continuously optimizing the network structure, a new deep neural network model was redesigned. The experiments have demonstrated the effectiveness of this approach. Compared to traditional methods, there has been a significant improvement in recognizing various modulation types. Most importantly, this approach solves the problem of end-to-end signal recognition and eliminates the cumbersome process of manual feature extraction. The test accuracy is very high, and the model exhibits good generalization capability. It is believed that with continued research by more scholars and the ongoing development of deep learning, there will undoubtedly be further breakthroughs in modulation recognition methods for signal and information processing, and accuracy will continue to improve. However, there are still many areas for improvement in this experiment, such as the limited amount of data, as well as room for further enhancement in network structure and hyperparameters. Future work will involve continuous attempts and refinements.

References

[1] Maganioti, A.E., Chrissanthi, H.D., Charalabos, P.C., Andreas, R.D., George, P.N. and Christos, C.N. Cointegration of Event-Related Potential (ERP) Signals in Experiments with Different Electromagnetic Field (EMF) Conditions. Health, (2010) 2, 400-406.

[2] Bootorabi, F., Haapasalo, J., Smith, E., Haapasalo, H. and Parkkila, S. Carbonic Anhydrase VII—A Potential Prognostic Marker in Gliomas. Health, (2011) 3, 6-12.

[3] Shafik, R. A., Rahman, S., & Supangkat, S. H. Automatic Modulation Classification for Cognitive Radios using Cyclostationary Features. IEEE Communications Surveys & Tutorials, (2012) 14(1), 105-117.

[4] O'Shea, T. J., & West, N. Radio Machine Learning Dataset Generation with GNU Radio. Proceedings of the GNU Radio Conference, (2017) 1(1), 1-9.

[5] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. End-to-End Object Detection with Transformers. Proceedings of the European Conference on Computer Vision, (2020) 213-229.

[6] Hu, J., Shen, L., & Sun, G. Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2018) 7132-7141.

[7] Amini, A., Shirani-Mehr, H., & Karbasi, A. Digital Modulation Classification: A Deep Learning Approach. IEEE Transactions on Communications, (2017) 65(11), 4658-4668.

[8] He, K., Zhang, X., Ren, S., & Sun, J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2016) 770-778.

[9] Kingma, D. P., & Ba, J. Adam: A Method for Stochastic Optimization. Proceedings of the International Conference on Learning Representations(2015) (ICLR).

[10] Powers, D. M. W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, (2011) 2(1), 37-63.

[11] Hochreiter, S., & Schmidhuber, J. Long Short-Term Memory. Neural Computation, (1997) 9(8), 1735-1780.

[12] LeCun, Y., Bottou, L., Orr, G. B., & Müller, K.-R. (2012). Efficient BackProp. In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), Neural Networks: Tricks of the Trade (pp. 9-48). Springer.

[13] Shorten, C., & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(60), (2019). 1-48.

[14] Prechelt, L. (1998). Early Stopping - But When? In G. B. Orr & K.-R. Müller (Eds.), Neural Networks: Tricks of the Trade (pp. 55-69). Springer.

[15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Proceedings of the International Conference on Learning Representations (ICLR).

[16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15(1), (2014). 1929-1958.