Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 15, No. 3, 2024

Research on radio signal modulation mode identification algorithm based on transformer

Ziyue Yu *, Yu Chen

College of Electronic and Information Engineering, Southwest Minzu University, Chengdu 610041, China

* **Corresponding author**: Ziyue Yu (Email: 1157673324@qq.com)

Abstract: Aiming at the deep learning-based radio signal modulation mode recognition algorithm, a Transformer network model of radio signal modulation mode recognition algorithm is proposed. Firstly, the Transformer network model is built, which contains 12 layers of Transformer encoder and 3 layers of fully connected layers; secondly, the original data is partitioned into sequences with a fixed window size, and the models with different window sizes are compared, and the model with the optimal window size is comprehensively selected for experimental comparisons; lastly, the optimal model is compared with the other benchmark models, and the modulation signal is used to recognize the modulation mode of radio signals using the RaidoML2016.10a international standard dataset. The sample is the RaidoML2016.10a international standard dataset to train the model. The experimental results show that the Transformer model algorithm can achieve recognition accuracy of 94.63% under the condition of signal-to-noise ratio of 16 dB, as well as an average recognition accuracy of 63.18%, and the maximum recognition accuracy and average recognition accuracy are higher than those of the current benchmark models by 2%~11% and 1%~7%, respectively. The model has superior recognition and fast convergence speed.

Keywords: Deep learning; Transformer model; Automatic modulation recognition; Confusion matrix.

1. Introduction

Automatic Modulation Recognition (AMR) is a core technique used for signal analysis in non-collaborative communication environments[1]. By looking at the received signal, AMR is able to identify the modulation used by the transmitter. In recent years, this technology has been widely used in both military and civil applications, covering a wide range of aspects such as cognitive radio, signal monitoring, spectrum management, reconnaissance analysis, and intelligent software-defined radio [2],[3]. With the rapid development of wireless communication technology, especially in the advancement of 5G and future 6G networks, the efficient utilization and management of spectrum resources become more and more important.

In general, automatic modulation recognition (AMR) algorithms can be categorized into two types: likelihood-based (LB) and feature-based (FB). Likelihood-based methods are usually formulated through multiple hypothesis testing, i.e., comparing the likelihood function of an unknown signal with the threshold of a known density function [4]. However, when a large variety of signals are considered, the process of calculating the judgment threshold becomes complex and computationally expensive. Therefore, the LB-AMC method is not suitable for real-time processing and low-cost applications. In contrast, the feature-based (FB) approach is considered an effective alternative to the LB approach because it can significantly reduce the computational complexity. The FB-AMC approach achieves signal recognition through feature extraction and classification. During the feature extraction process, the expert system usually manually selects a series of features, such as wavelet transform-based [5], instantaneous features [6], and statistical features [7]. Although these methods can achieve better performance in simpler environments, they still have limitations when facing complex and changing radio environments. Especially under extreme conditions such as low signal-to-noise ratio (SNR), multipath fading, or frequency offset, the recognition accuracy of traditional methods is easily affected by noise and signal distortion.

In recent years, deep learning (DL) techniques have made significant breakthroughs in radio signal modulation identification, and various algorithms have demonstrated their advantages in modulation identification, using neural networks to extract features of the original signals and identify their modulation types, and powerful automatic feature learning can improve the identification accuracy. In 2016, O'Shea et al [8] proposed automatic modulation mode recognition of 11 types of radio signals using convolutional neural network (CNN), the model is composed of four CNN layers and two fully connected layers, and the experimental results show that the CNN has a recognition accuracy of 73% in the modulation recognition task. Previous DL-AMR models have been successfully implemented as benchmark models for the AMR task by employing various algorithms, including Convolutional Neural Networks (CNN) [9],[10], Recurrent Neural Networks (RNN) [11],[12], and Hybrid Networks[13],[14],[15]. Deep learning, especially Transformer-based models, has gradually become a research hotspot in the field of signal processing due to its excellent ability in dealing with sequential data and long and short-term dependencies. The self-attention mechanism of Transformer models can effectively capture spatio-temporal features in signals, and they are highly parallelized and can handle large-scale data. Compared with traditional convolutional neural networks (CNN) and recurrent neural networks (RNN), Transformer exhibits greater robustness and higher recognition accuracy when processing radio signals.

This study aims to explore the Transformer-based radio signal modulation mode recognition algorithm to address the limitations of traditional methods in complex environments by designing a model architecture suitable for radio signals. Specifically, the original radio signal is segmented into sequences with a fixed window size for automatic feature extraction, the models with different window sizes are compared, and the model with the optimal window size is comprehensively selected for experimental comparison. Through experimental validation, the recognition recognition accuracy of different models and algorithms under various signal-to-noise ratios as well as the average recognition accuracy are compared; the number of parameters, the training time, the memory size, and the highest recognition accuracy of different models are compared, and the model complexity and confusion matrix are analyzed.

2. Signal model and proposed system model

2.1. Signal Model

We mainly consider single-input single-output communication systems, where the radio signal r(t) received by the communication receiver can be expressed as:

$$r(t) = s(t) * h(t) + n(t) \tag{1}$$

The modulation signal s(t) can be expressed as:

$$s(t) = s_I(t) + js_Q(t) \cdot e^{j2\pi f_c t} \tag{2}$$

In equation (1), s(t) is the modulated signal from the transmitter, h(t) is the impulse response of the transmission channel, and n(t) denotes additive Gaussian white noise (AWGN). The received signal r(t) is sampled n times at rate f_s = 1/T_s to produce the discrete-time observation signal r(t). The signal r(t) received by the receiver is stored as a discrete in-phase and quadrature (IQ) component. Equation 2 represents the complex envelope form of the signal.

The communication channel impulse response h(t) can be expressed as:

$$h(t) = a * \delta(t - \tau) \tag{3}$$

In equation (3), a is the amplitude of the signal, δ is the impulse function, t is the time interval, and τ is the channel multipath delay.

Overall, automatic modulation identification is essentially a classification task aimed at inferring the modulation mode employed by the original transmitted signal s(t) from the received signal r(t), where M represents the total number of possible modulation types. The core of the process lies in dissecting the characteristics of the received signal and accurately determining the modulation class to which it belongs. The key to evaluating the performance of a modulation classification algorithm is its classification speed and recognition accuracy.

2.2. Transformer network model

The Transformer network model proposed in this paper is shown in Figure 1. Firstly, the original data is preprocessed, the original data is 2×128, and the sequence is segmented into N identical sequences with a fixed window size, secondly, the data is fed into a 12-layer encoder for automatic feature extraction, which includes layer normalization, multi-attention mechanism, and feed-forward neural network, and lastly, the classification module consists of three fully-connected layers, and the eleven modulation signals are classified through Softmax recognize and classify them.

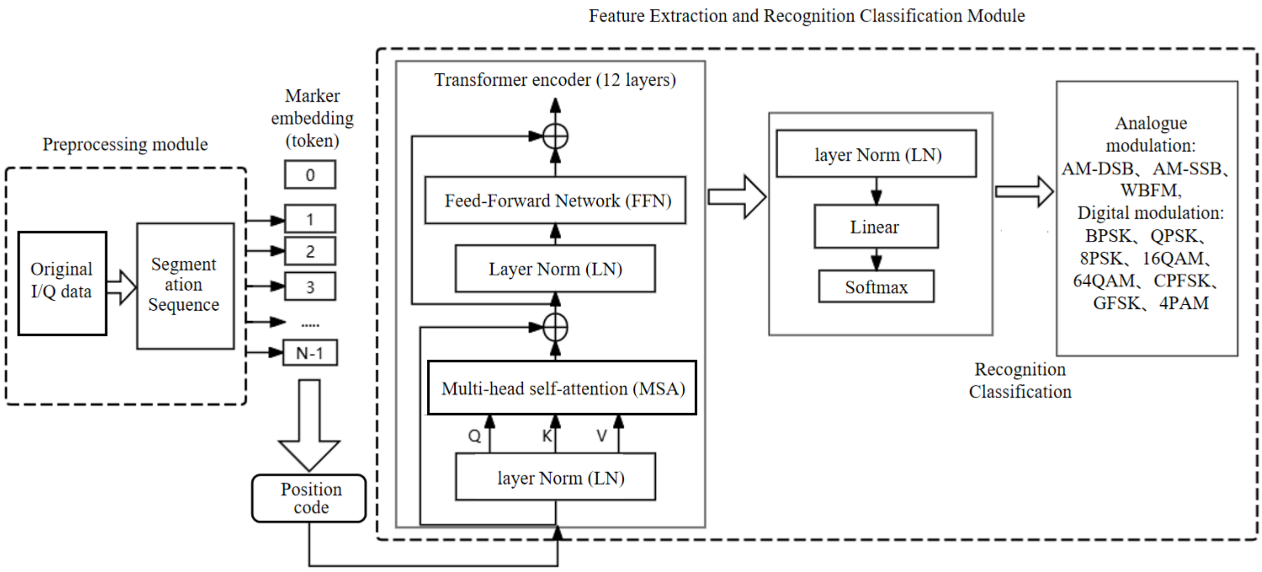

Figure 1. Structure of the Transformer network model

The following activation functions are commonly used as shown in equations (4) (5) (6)

$$\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{4}$$

$$\mathrm{Softmax}(x) = \frac{e^{x_i}}{\sum_{j=0}^{k} e^{x_k}} \tag{5}$$

$$\mathrm{SeLu}(x) = \lambda \begin{cases} x, & x > 0, \\ \alpha e^{x} - \alpha, & x \leq 0. \end{cases} \tag{6}$$

$$\mathrm{ReLu}(x) = \begin{cases} 1, x > 0, \\ 0, x \leq 0. \end{cases} \tag{7}$$

Batch Normalization Layer: this is a method of normalizing each batch of features. The purpose of batch normalization (BN) is to reduce the covariance bias, i.e., features continue to change distribution during training.

Pooling Layer: the pooling layer reduces the spatial size of the representation in order to reduce the number of parameters and computations in the network and to allow feature mapping to run independently. The pooling layer can be either maximum pooling or average pooling.

Dropout Layer: this layer is usually used to reduce overfitting in neural networks. When the dropout technique is used, neurons in the hidden or visible layer are randomly removed or discarded.

2.3. Dataset and implementation details

A. Dataset

Table 1. RML2016.10A dataset

RML2016.10A dataset Modulation type	Analog modulation: AM-DSB, AM-SSB, WBFM, Digital modulation: BPSK, QPSK, 8PSK, 16QAM, 64QAM, CPFSK, GFSK, 4PAM
Data format	IQ (In-phase and Quadrature) data format:2×128
Sample size Signal-to-Noise Ratio Range	220000(1000*20*11) [-20, +18] dB
Sampling rate	200kHz
Maximum carrier offset and its standard deviation	500Hz, 0.01Hz
Maximum offset of the sampling rate and its standard deviation	50Hz, 0.01Hz
Channel model	Rice channel
Channel environment	Additive Gaussian white noise, selective fading, center frequency shift, sample rate shift

This experiment uses the open-source benchmark dataset RadioML2016.10a. The dataset contains 220,000 samples of modulated signals with signal-to-noise ratios (SNRs) ranging from -20dB to 18dB in 2dB steps; there are 1,000 samples at each SNR value, for a total of 220k samples, and the length of each sample is 128. The dataset covers 11 common modulated signal The dataset covers 11 common modulated signal types including WBFM, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, 4-PAM, 16-QAM, 64-QAM, QPSK, and 8PSK. each signal sample consists of two I/Q sequences, each containing 128 feature points. To better approximate the real-world environment, the signals in the dataset are corrupted by a variety of channel disturbances, including additive white noise (AWGN), multipath fading, sampling rate offsets, and center frequency offsets. We divide the dataset into training, validation and test sets in the ratio of 6:2:2.

B. Realization details

We set up two experiments to evaluate the Transformer model. The first experiment explores the performance of comparing the segmentation of sequences with different window sizes, and the model with the optimal window size is comprehensively selected for experimental comparison. In the second experiment, the current state-of-the-art AMR models are used to provide benchmark comparisons, including PET-CGDNN [14], MCLDNN [15], LSTM [16], ResNet [17], and CNN [18]. We compared the Transformer algorithm with these six SoA algorithms.

All models were evaluated on the same dataset, using categorical cross-entropy as the loss function and optimized using the Adam optimizer. The initial learning rate was set to 0.001, and if the validation loss failed to decrease within 5 consecutive training cycles, the learning rate was multiplied by 0.8 to improve the training efficiency. The training period (Epoch) is set to 150 and the batch size (Batch size) is 256 to avoid falling into local optimal solutions. When the validation loss does not show improvement within 50 cycles, we terminate the training and use the model that achieved the lowest loss on the validation set to predict the modulation type of each signal in the test set. Experiments were performed using the TensorFlow framework (version 2.6.0), Keras (version 2.6.0), and NVIDIA GeForce RTX 3060 Laptop GPUs, utilizing TensorFlow-gpu accelerated computation.

3. Experimental results and discussion

3.1. 3.1 Window selection experiment

In the Transformer network model, the size of the signal block is closely related to the length of the effective input sequence. Specifically, the smaller the signal block is, the longer the effective input sequence is; conversely, the shorter the effective input sequence is. Therefore, this section first explores the problem of choosing the window size and further analyzes the impact of the signal block size or the length of the input sequence on the classification accuracy of the model. Considering that the original input sequence size is 128x2, different window sizes are selected for training and testing in this experiment. The window size is denoted as S, and its specific values are 32x2, 16x2, 8x2, 4x2 and 1x2. The experimental results are shown in Figure 2.

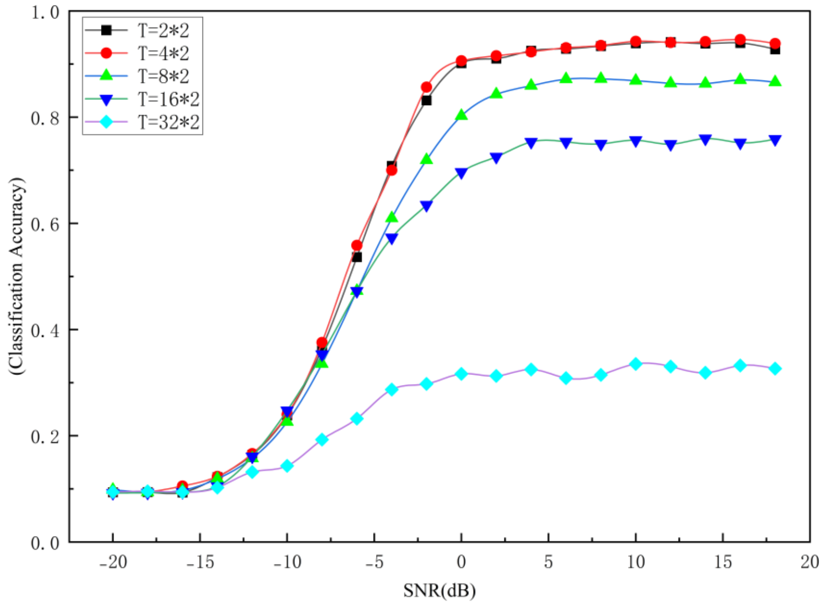

Figure 2. Variation of recognition accuracy with signal-to-noise ratio for different window sizes

Table 2. Classification accuracies of different models at different signal-to-noise ratios and average classification accuracies

TF model	Learned Parameters	Training time (second/epoch)	Memory (MB)	Average accuracy (%)	Highest accuracy (%)
T=2*2	11,535,820	89	44.3	62.67%	94.15%
T=4*2	4,469,719	37	17.4	63.18%	94.63%
T=8*2	2,734,396	35	10.7	57.55%	87.16%
T=16*2	2,292,964	32	9.06	51.41%	75.98%
T=32*2	2,182,120	30	8.64	24.46%	33.53%

Figure 2 shows the classification accuracy of the model with different window settings, as the window size setting increases, the memory size of the model decreases, the number of parameters and the training time decreases, and the classification accuracy decreases significantly. It can be concluded that the window size affects the classification accuracy of the model. After comprehensive consideration, we finally choose to set the window size to 4×2. This model is the optimal algorithm, which is used for the second experiment to compare with other benchmark models.

As shown in Table 2, the classification accuracies as well as the average classification accuracies of different network models under different signal-to-noise ratio conditions are compared. The results show that without affecting the recognition accuracy as well as the model size, the model with a window size of 4×2 is selected for comprehensive consideration. Therefore, the Transformer network model with a window of 4×2 is selected as the optimal modeling algorithm for the second experiment for comparison with other benchmark models.

As shown in Fig. 3, the confusion matrix of the Transformer model with a window size of 4 × 2 at a signal-to-noise ratio of 18 dB is demonstrated. As can be seen from the figure, the confusion between WBFM and AM-DSB is easy to occur mainly because they both belong to the continuous modulation type and may be modulated using the same audio source (e.g., speech or music) in practical applications. Due to the similarity of the audio sources, the modulated signals are often audibly indistinguishable, thus increasing the likelihood of confusion. In addition, from a spectral analysis point of view, both WBFM and AM-DSB modulated signals have a certain bandwidth. Although the bandwidth of WBFM is usually larger than that of AM-DSB, in some cases, their spectral characteristics may be similar, leading to difficulties in direct differentiation in spectrograms. Another common confusion issue is the misclassification between 16-QAM and 64-QAM. The constellation diagrams of these two modulations are similar and the presence of overlapping constellation points in the digital domain leads to easy confusion between 16-QAM and 64-QAM. Although this problem still exists in the Transformer model, it has significantly improved in solving this problem compared to other current deep learning frameworks.

As shown in Fig. 4, the convergence curve of the loss function of the Transformer model is demonstrated, from which it can be seen that the loss curve represents the loss function curve of the model, which is roughly stabilized and converged at ground 150 Epochs.

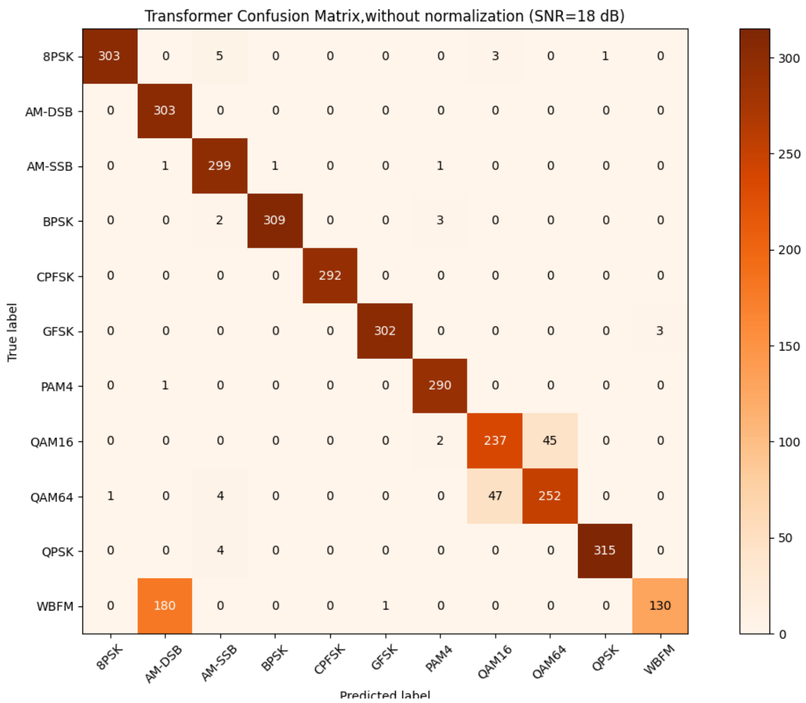

Figure 3. Confusion matrix for the Transformer model with a window size of 4×2 at a signal-to-noise ratio of 18 Db

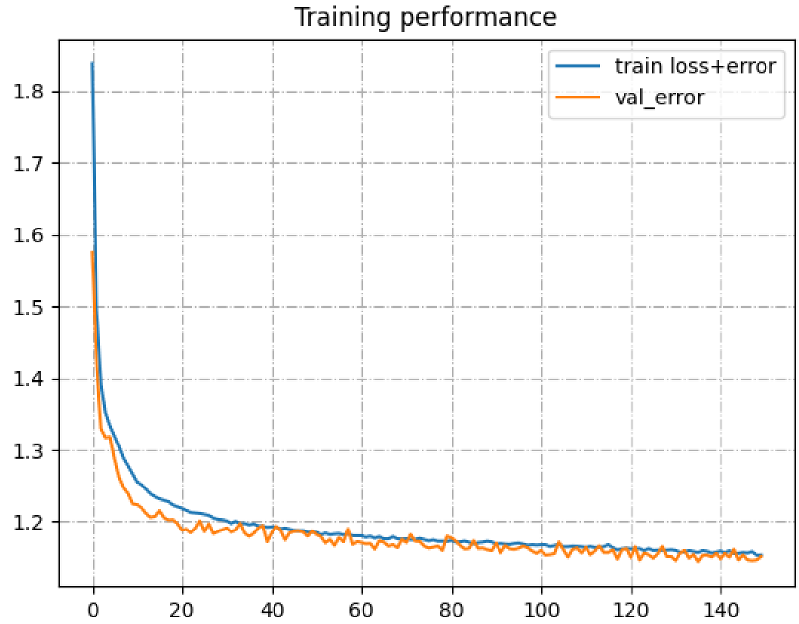

Figure 4. Convergence curve of the loss function for the Transformer model

3.2. Comparison of different AMR models

As shown in Fig. 3, the overall recognition accuracies of different network models under different signal-to-noise ratio conditions are compared. The results show that when the signal-to-noise ratio is -20dB~18dB, the performance of our proposed Transformer model is significantly higher than the recognition accuracies of the other benchmark models compared to the other network models. When the signal-to-noise ratio is 0dB~18dB, its maximum accuracy reaches 93.21%. It is 9%~11%, 10%~11%, 1%~2%, and 0%~1% higher than CNN, ResNet, LSTM, and MCLDNN, respectively. When the signal-to-noise ratio is 16 dB, the classification accuracy of our model reaches a maximum of 93.13%. When the signal-to-noise ratio is 16 dB, the classification accuracy of our model reaches a maximum of 94.63%, which is 1.68% higher than the benchmark model MCLDNN. Overall, our model is 1% to 2% higher than the current optimal deep learning framework MCLDNN at high signal-to-noise ratios, but at low signal-to-noise ratios there is basically no improvement compared to several other deep learning frameworks, but there is a significant improvement in the overall recognition accuracy. Therefore, it can be concluded that the Transformer network model has the advantage of high recognition accuracy in radio signal modulation recognition.

As shown in Table 3, several metrics were selected for performance comparison and complexity analysis, including the number of parameters, average training time, memory space (weight size), and the highest recognition accuracy.PET-CGDNN has a significant advantage in terms of the number of parameters, the average training time, and the memory space because the model was lightened using the pruning principle, which reduced the number of parameters as well as weight sizes . Taken together, our proposed model has significant advantages in several aspects, and the highest recognition accuracy reaches 94.63%.

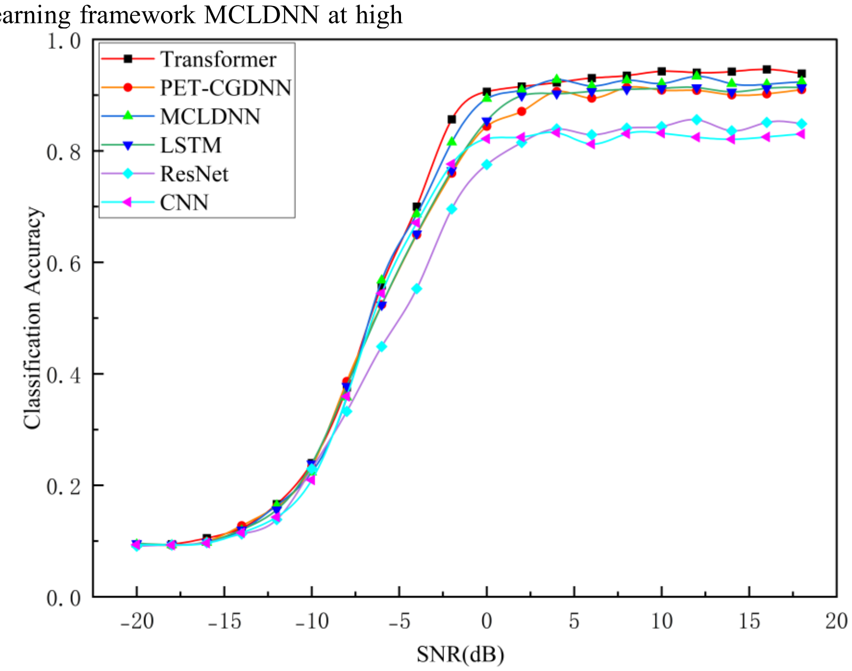

Figure 5. Variation of different AMR models with signal-to-noise ratio

Table 3. Classification accuracies of different models at different signal-to-noise ratios and average classification accuracies

Model	-20dB	-10dB	-4dB	0dB	18dB	Average
Transformer	9.45	24.01	70.02	90.63	93.87	63.18
PET-CGDNN	9.32	23.32	64.95	84.41	91.01	60.44
MCLDNN	9.41	22.36	68.64	89.36	92.41	62.06
LSTM	9.59	23.86	65.14	85.41	91.36	60.76
ResNet	9.09	22.95	55.27	77.55	83.86	55.45
CNN	9.36	20.95	67.09	82.18	83.05	56.78

Table 4. Number of parameters, training time, memory size and maximum recognition accuracy for different models

Model Framework	Learned Parameters	Training time (second/epoch)	Memory (KB)	Highest accuracy (%)
Transformer	4469719	37	174000	94.63
PET-CGDNN	71871	6	846	91.36
MCLDNN	406199	17	4859	92.95
LSTM	201099	11	2818	91.41
ResNet	3098283	32	36370	84.36
CNN	858123	20	10126	83.09

As shown in Table 4, the different models are compared and analyzed, including the recognition accuracies at different signal-to-noise ratios and the overall average recognition accuracy. All models are evaluated on the same dataset, adjusting the input and output layers to fit the data dimensions, and therefore the number of parameters varies accordingly, as shown in the table. It is clear that the proposed Transformer model has the highest overall recognition accuracy and outperforms the other AMR models. The recognition accuracy is also better than other benchmark models at 0dB~18dB.From all the signal-to-noise ratios, the overall average recognition accuracy of our model is 63. 18%, which is 1.68% higher than the current optimal model MCLDNN.

4. Conclusion

In this paper, we have proposed a Transformer network model which has good recognition performance and fast convergence. The experimental results show that the Transformer model algorithm can achieve a recognition accuracy of 94.63% under the condition of a signal-to-noise ratio of 16 dB, as well as an average recognition accuracy of 63.18%, with the highest recognition accuracy and average recognition accuracy higher than the current benchmark model by 2%~11% and 1%~7%, respectively. The recognition performance is better than other benchmark models, but the algorithm still has further improvement: ① The algorithm has poor recognition effect when the SNR is ≤0dB, and in the future, we can consider noise reduction of the signal at the input of the deep learning network first. ② The algorithm has been validated at the theoretical level, but it has not been ported to the corresponding hardware devices. The next step can be to deploy the algorithm on hardware devices and test its performance in real environments.

References

[1] Wang, Yan et al. "Communication Modulation Signal Recognition Based on the Deep Multi-Hop Neural Network",Journal of the Franklin Institute 358.12 (2021): 6368-6384.

[2] Kumar, Satish et al. "Automatic Modulation Recognition: an FPGA Implementation",IEEE Communications Letters 26.9 (2022): 2062-2066.

[3] Huynh-The, Thien et al. "Efficient Convolutional Networks for Robust Automatic Modulation Classification in OFDM-Based Wireless Systems",IEEE systems journal 17.1 (2022): 964-975.

[4] Zhang, Tingping et al. "Deep Learning for Robust Automatic Modulation Recognition Method for IoT Applications.",IEEE Access 8 (2020): 117689-117697.

[5] Ho, Ka Mun et al. "A Wavelet-Based Method for Classification of Binary Digitally Modulated Signals",IEEE Sarnoff Symposium (2009): 144-148.

[6] Moser, Elliott et al. "Automatic modulation classification via instantaneous features",National Aerospace and Electronics Conference (2015): 218-223.

[7] Park, Myung Chul, and Dong Seog Han. "Deep Learning-Based Automatic Modulation Classification with Blind OFDM Parameter Estimation.",IEEE Access 9 (2021): 108305-108317.

[8] O'Shea, Timothy J., and Nathan West. "Radio Machine Learning Dataset Generation with GNU Radio",openalex 1.1 (2016)

[9] A. P. Hermawan, R. R. Ginanjar, D.-S. Kim, and J.-M. Lee, "CNN-based automatic modulation classifification for beyond 5G communications," IEEE Commun. Lett., vol. 24, no. 5, pp. 1038–1041, May 2020.

[10] Mendis, Gihan Janith et al. "Deep Learning Based Radio-Signal Identification with Hardware Design",IEEE Transactions on Aerospace and Electronic Systems 55.5 (2019): 2516-2531.

[11] Ratendran, Sreeraj et al. "Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors",IEEE Transactions on Cognitive Communications and Networking 4.3 (2018): 433-445.

[12] Hong, Dehua et al. "Automatic Modulation Classification Using Recurrent Neural Networks",IEEE International Conference Computer and Communications (2017): 695-700.

[13] Xu, Jialang et al. "A Spatiotemporal Multi-Channel Learning Framework for Automatic Modulation Recognition",IEEE Wireless Communications Letters 9.10 (2020): 1629-1632.

[14] Zhang, Fuxin et al. "An Efficient Deep Learning Model for Automatic Modulation Recognition Based on Parameter Estimation and Transformation",IEEE Communications Letters 25.10 (2021): 3287-3290.

[15] Xu, Jialang et al. "A Spatiotemporal Multi-Channel Learning Framework for Automatic Modulation Recognition", IEEE Wireless Communications Letters 9.10 (2020): 1629-1632.

[16] Ke, Ziqi, and Haris Vikalo. "Real-Time Radio Technology and Modulation Classification Via an LSTM Auto-Encoder", IEEE Transactions on Wireless Communications 21.1 (2021): 370-382.

[17] Liu, Xiaoyu et al. "Deep Neural Network Architectures for Modulation Classification", Computing Research Repository (2017)

[18] Tekbıyık, Kürşat et al. "Robust and Fast Automatic Modulation Classification with CNN under Multipath Fading Channels", Vehicular Technology Conference (2020): 1-6.