Academic Journal of Engineering and Technology Science
ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

Signal Recognition Model Based on Transformer

Zhe Wang*

Quality and Standards Academy, Shenzhen Technology University, Shenzhen, 518118, China
**Corresponding author: 19935226865@163.com*

Abstract: *With the increasing complexity of the electromagnetic environment, the recognition of electromagnetic signals faces significant challenges. Traditional methods of electromagnetic signal recognition rely on manually designed features, which are insufficient for handling complex signal classification tasks effectively. This paper proposes a novel approach that combines Convolutional Neural Networks (CNN) and Transformer models. The method involves constructing a signal transmission model to generate signals of four modulation types: QAM64, QAM16, 8PSK, and CPFSK. These signals are then demodulated and filtered to obtain I/Q signals, which are converted into two-dimensional grayscale images. A signal recognition model is employed for feature extraction and classification, achieving an accuracy of approximately 90% within a certain Signal-to-Noise Ratio (SNR) range.*

Keywords: *Electromagnetic Signal Recognition, Convolutional Neural Network, Transformer*

1. Introduction

With the increasing complexity of the electromagnetic environment, the types and quantities of radiation sources in the electromagnetic field continue to grow, posing more challenges to the accurate identification of electromagnetic signals. In such an environment, evaluating electromagnetic signal processing models becomes particularly important as it directly relates to the assessment of model performance. Simultaneously, with the significant increase in the number of antennas and the rise in carrier frequencies, the physical characteristics of the electromagnetic radiation field have fundamentally changed. To adapt to this change, traditional far-field plane wave models are no longer suitable for describing signal propagation characteristics at high frequencies and large-scale antenna arrays, necessitating more complex spherical wave models for modeling [1]. Traditional electromagnetic signal identification techniques often struggle to achieve accurate predictions when dealing with complex classification problems. Moreover, these techniques rely on traditional machine learning algorithms, which typically require pre-processing of electromagnetic signals. However, this pre-processing process has significant limitations in extracting deep features of electromagnetic signals, thereby affecting the effectiveness of signal processing [2]. Additionally, the complex electromagnetic environment results in limited available signals and low signal-to-noise ratios, often failing to meet the sample quantity required by traditional training methods, leading to the so-called few-shot recognition problem [3]. Furthermore, the signals received in actual electromagnetic environments include not only known categories but also involve unknown categories, posing the challenge of open set signal recognition. Most existing studies focus on closed set signal recognition, where models often misclassify unknown category signals as known categories, thus impacting subsequent analysis and decision-making in electromagnetic signal systems [4].

Electromagnetic signal identification technology can be mainly divided into two methods: one is based on feature engineering, and the other is based on machine learning [2]. The feature engineering-based method focuses on extracting manually designed features from electromagnetic signals, relying on expertise and experience to select the features most relevant to the classification task. In [5], raw I/Q sample data is collected through a distributed electromagnetic spectrum detection network, and radio frequency fingerprint features are extracted based on I/Q data distribution characteristics. In [6], by analyzing the complexity of received signals and the environmental noise conditions, fractal dimension features of different complexities are used to describe subtle signal characteristics, establishing a feature database and accurately identifying subtle signal features in low signal-to-noise ratio environments using grey relational theory systems to identify different broadcasting stations. Xi et al. used adaptive multi-scale morphological gradient filtering to process electromagnetic environment signals, suppressing noise while preserving signal components reflecting electromagnetic environment characteristics [7]. They

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

then compressed the signals using non-negative matrix factorization techniques and calculated a set of feature parameters used for classifying electromagnetic environment signals. These methods typically involve feature extraction from the collected signals, selecting feature vectors with high separability for signal identification. In this process, the choice of classifier mainly relies on machine learning techniques, leveraging their pattern recognition capabilities to improve identification accuracy. Yang et al. represented input signal feature parameters as matrices and constructed variance matrices by calculating the variance of each parameter [8]. They then input feature parameters of unknown modulation type signals into a KNN classifier, which identified signal types by calculating the Euclidean distance to the k-nearest signals of known types. In [9], a weighted optimized SVM algorithm is used to classify multidimensional feature vectors of radar signals, improving recognition accuracy under different signal-to-noise ratio conditions. In [10], various types of pulse repetition interval modulation radar signals are simulated, and feature values are calculated and input into the C4.5 decision tree algorithm to construct a decision tree for signal type recognition. Although these methods have made some progress in feature extraction and signal classification, electromagnetic signal features depend on expert knowledge and experience, which may result in suboptimal performance when dealing with unknown or varying signal types and may not meet the demand for real-time processing of large amounts of data. Although these methods have made some progress in feature extraction and signal classification, several limitations remain. They rely on expert knowledge and experience, making the feature extraction process complex and time-consuming, which hinders their ability to automatically adapt to new or changing signal types. Additionally, when processing large volumes of data, the computational burden is substantial, limiting real-time processing capabilities. Moreover, the performance of these methods may degrade significantly when dealing with unknown or varying signal types, indicating a lack of robustness and generalizability.

Neural network technology, a computational model that mimics the structure and function of human brain neurons, is primarily used for tasks such as pattern recognition, data classification, and prediction [11]. Neural networks consist of multiple layers, each comprising numerous interconnected "neurons" or nodes. Each neuron receives inputs, performs a weighted sum, and determines the output through an activation function [12]. This structure makes neural networks particularly suitable for solving signal identification problems in complex electromagnetic environments. Chen et al. proposed a method combining short-time Fourier transform and CenterNet-based deep learning network to detect and estimate frequency hopping signals in complex electromagnetic interference environments by converting signals into spectrograms and using deep learning for shape and position recognition to achieve high-accuracy parameter estimation [13]. Tu et al., using the RadioML 2016.10 A dataset, conducted a comparative study between complex-valued and real-valued neural network building blocks, including complex convolutional layers and complex dense layers, to verify the superiority of complex-valued networks over real-valued networks in automatic modulation classification [14]. Hou et al. used the improved Cohen class time-frequency distribution (CTFD) to convert radar signals into time-frequency images (TFI) and employed convolutional denoising autoencoders (CDAE) for filtering to reduce noise effects [15]. Signal separation and identification were achieved using three semantic segmentation networks—fully convolutional network (FCN-8s), U-Net, and DeepLab V3+—and multiple radar signals were classified using color threshold filtering. In [16], Choi-Williams time-frequency distribution was used to obtain two-dimensional time-frequency images of signals, achieving high-precision classification of known signals through convolutional neural networks and extracting corresponding feature vectors, which were then input into the SVDD algorithm to construct a high-dimensional hypersphere for detecting whether the signal to be identified belonged to a known category. It is worth noting that different signal recognition and classification tasks require different network architectures. Furthermore, in cases of insufficient or unevenly distributed data, neural networks are prone to overfitting the training data. Additionally, the decision-making process of neural networks lacks transparency and interpretability, which may lead to suboptimal performance in certain application scenarios.

In [17], the GNU Radio was used to generate a dataset of electromagnetic signals covering eight modulation modes such as BPSK and QPSK, and these signals were converted into two-dimensional images by generating grayscale images from the real and imaginary parts. A convolutional neural network with two convolutional layers and two fully connected layers was then used for modulation recognition, and the method was validated on 10,400 test samples, demonstrating a high accuracy in recognizing different modulation modes with a single signal recognition time of approximately 0.1 milliseconds. Building upon the network architecture in this paper, I incorporated a Transformer [18], a method that captures long-range dependencies in sequential data using self-attention mechanisms. Under the channel model constructed in this study, the classification accuracy for QAM64, QAM16, 8PSK, and CPFSK was compared. It was found that the classification accuracy of the model before improvement was only around 75%, whereas the improved model achieved a classification accuracy of over 90%. This

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

demonstrates that the combination of Convolutional Neural Networks and Transformers has a stronger generalization capability for different types of signals.

2. The basic funamental of BP neural network

A complete signal transmission model encompasses processes ranging from source coding, modulation, channel transmission, demodulation, to signal recovery. In the channel model constructed in this paper, data from text files are first converted into binary form. Electromagnetic signals are then generated using four modulation techniques. During the modulation process, the generated baseband signals undergo upconversion to produce modulated signals, which are subsequently subjected to noise addition and multipath effect simulation. The resulting signals are then passed through a lowpass filter for demodulation to extract the I/Q signals. Finally, the processed signal data and labels are segmented and saved in suitable formats for subsequent signal recognition algorithm training and testing.

2.1 Source Coding

During the encoding phase, the characters in the text file are first read and converted into binary sequences, with each character's ASCII value being transformed into the corresponding 8-bit binary number. These binary sequences are then mapped to I/Q symbols based on four modulation schemes: QAM64, QAM16, 8PSK, and CPFSK. Taking 64QAM modulation as an example, every six binary bits are grouped together, and their mapping rules are as follows:

$$64\text{QAM Mapping} = \begin{cases} (000000) \to -7+7j \\ (000001) \to -5+7j \\ (000011) \to -3+7j \\ (000010) \to -1+7j \\ (000110) \to 1+7j \\ (000111) \to 3+7j \\ (000101) \to 5+7j \\ (000100) \to 7+7j \\ \vdots \\ (111111) \to 7-7j \end{cases} \tag{1}$$

Through the aforementioned mapping rules, the binary sequences are successfully mapped to the corresponding I/Q data, ready for the modulation and channel processing stages.

2.2 Addition of Gaussian Noise

To simulate a Gaussian noise environment, random Gaussian noise is added to each generated signal in this experiment. First, the power of the original signal is calculated as follows:

$$P_{\text{signal}} = \frac{1}{N}\sum_{n=0}^{N-1} |x[n]|^2 \tag{2}$$

where $x[n]$ represents the discrete-time signal, and N is the number of samples in the signal.

To ensure that the added noise conforms to a specified Signal-to-Noise Ratio, the noise power P_{noise} is calculated based on the signal power and SNR. The SNR, typically expressed in decibels (dB), is the ratio of signal power to noise power:

$$\text{SNR} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{3}$$

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

Thus, the noise power P_{noise} can be expressed as:

$$P_{\text{noise}} = \frac{P_{\text{signal}}}{10^{\text{SNR}/10}} \tag{4}$$

Gaussian noise is then generated based on the calculated noise power $n[n]$:

$$n[n] \sim \mathcal{N}\left(0, \sqrt{P_{\text{noise}}}\right) \tag{5}$$

where $\mathcal{N}$ denotes a normal distribution.

Finally, the generated Gaussian noise is added to the original signal, yielding the noisy signal $y[n]$:

$$y[n] = x[n] + n[n] \tag{6}$$

2.3 Simulation of Multipath Effects

To simulate multipath effects in the channel, the experiment involves processing the signal with multiple delay, attenuation, and phase shift parameters. Multipath effects refer to the phenomenon where the received signal is composed of multiple signals with different delays, attenuations, and phase shifts due to reflection, scattering, and refraction during propagation. In a multipath environment, the received signal $r(t)$ is a superposition of signals from multiple paths, expressed as:

$$r(t) = \sum_{i=0}^{N-1} \alpha_i s(t - \tau_i) e^{j\phi_i} \tag{7}$$

where $s(t)$ is the original transmitted signal, α_i is the attenuation factor of the $i\text{-th}$ path, τ_i is the delay time of the $i\text{-th}$ path, ϕ_i is the phase shift of the $i\text{-th}$ path, and N is the number of multipath components.

2.4 Demodulation and Filtering

During the demodulation process, the original signal is recovered by separating and extracting the in-phase (I) and quadrature (Q) components. To demodulate the I component, the received signal is multiplied by the cosine component of the carrier signal:

$$\text{I}(t) = r(t) \cdot \cos(2\pi f_c t) \tag{8}$$

where $r(t)$ is the received signal, f_c is the carrier frequency, and t is time.

To demodulate the Q component, the received signal is multiplied by the sine component of the carrier signal:

$$\text{Q}(t) = -r(t) \cdot \sin(2\pi f_c t) \tag{9}$$

A low-pass filter is then applied to the demodulated signals to eliminate high-frequency noise. In this experiment, a first-order recursive filter is used, with the recursive formula given by:

$$y[n] = \alpha x[n] + (1 - \alpha) y[n-1] \tag{10}$$

where $y[n]$ is the filtered signal, $x[n]$ is the input signal, and α is the filter coefficient.

The filter coefficient α is derived from the bandwidth parameter and sampling rate:

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

$$\tau = \frac{1}{2\pi \ \text{Bandwidth}} \tag{11}$$

$$\alpha = \frac{1}{1 + \tau f_s} \tag{12}$$

where τ is the time constant, Bandwidth represents the bandwidth of the low-pass filter, and f_s is sampling rate.

In practical applications, clock offset and frequency offset are common issues. Clock offset refers to the desynchronization between the sampling clocks of the transmitter and receiver, leading to signal waveform distortion, which can be corrected through interpolation or resampling. Frequency offset refers to the desynchronization of carrier frequencies, causing signal spectrum shift, which can be resolved by estimating the frequency offset through spectral analysis and applying corresponding frequency compensation. Through these steps, the demodulation function extracts the I/Q components of the received signal and corrects for clock and frequency offsets, restoring the original baseband signal.

2.5 Generation of Grayscale Images

The I/Q signals, processed through demodulation and filtering, are divided into fixed-length batches. Each batch of I/Q data is repeated multiple times to form a suitably sized two-dimensional matrix. These data can be further converted into two-dimensional grayscale images, serving as input data for signal classification and modulation recognition.

3. Network Construction and Training

In this study, we designed a model combining Convolutional Neural Networks and Transformer architecture to process and classify two-dimensional signal data. The model consists of a one-dimensional convolutional layer to extract local features, followed by a Transformer to capture global dependencies, and finally a fully connected layer for classification, utilizing a Softmax layer to output probability distributions. The model accepts input data with the shape (batch_size, 1, 128, 2), where batch_size represents the batch size, 128 is the signal length, and 2 is the number of signal channels. During forward propagation, the input data shape is transformed to (batch_size, 128, 2) through a squeeze operation, then permuted to (batch_size, 2, 128) to fit the input format of the one-dimensional convolutional layer. The first layer of the model is a one-dimensional convolutional layer with an input channel size of 2, an output channel size of 64, a kernel size of 32, a stride of 16, and no padding. This convolutional layer performs local feature extraction on the input signal, resulting in an output shape of (batch_size, 64, 7).

To accommodate the input format of the Transformer, the output is permuted to (7, batch_size, 64). The Transformer has an input feature dimension of 64, 8 multi-head attention mechanisms, 2 encoder layers, and 2 decoder layers. The feedforward network dimension is 256, with a dropout rate of 0.1 and ReLU as the activation function. Within the Transformer, the signal undergoes processing through self-attention mechanisms and feedforward networks, capturing global dependencies. The processed data shape is then permuted to (batch_size, 7, 64).Next, the data is flattened to (batch_size, 64 * 7) through a flatten layer, and a fully connected layer transforms the input dimension from 64 * 7 to 32. A ReLU activation function is used for nonlinear transformation, followed by another fully connected layer that converts the output dimension from 32 to 4, corresponding to four classification labels. Finally, a Softmax layer outputs the classification probability distribution, with an output shape of (batch_size, 4). The network architecture is illustrated in Figure 1.

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

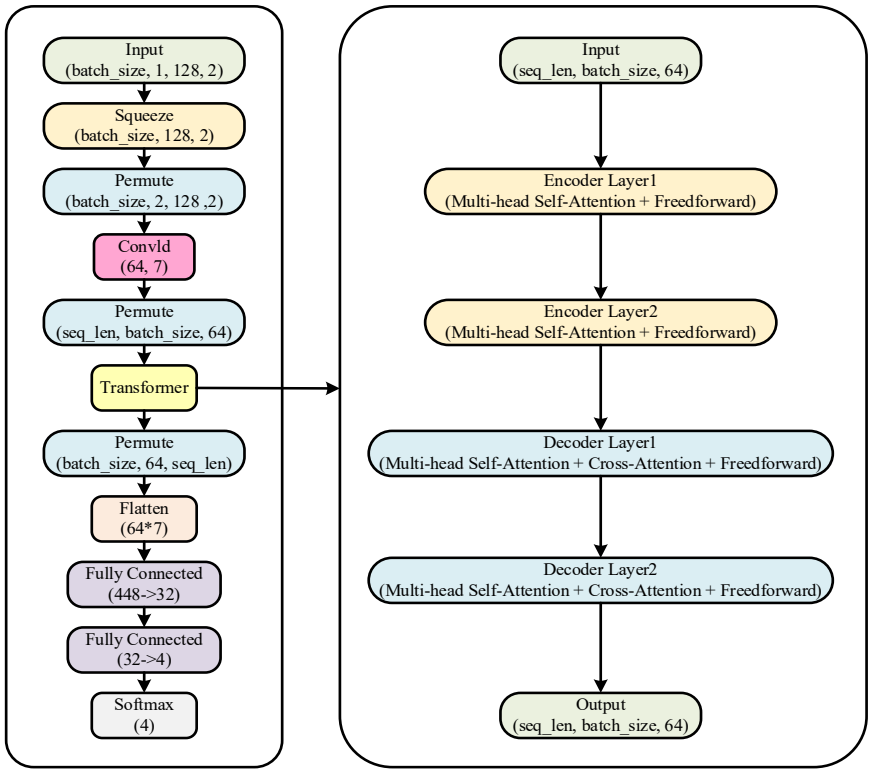


Figure 1: CNN-Transformer Signal Recognition Model Architecture

During the model training process, the cross-entropy loss function and the Adam optimizer were employed, with an initial learning rate set to 0.0005. Weight decay was also incorporated to prevent overfitting. A learning rate scheduler was utilized to dynamically adjust the learning rate based on the validation set loss, reducing it to a minimum value of 1e-6 as needed. To prevent the issue of gradient explosion, gradient clipping was applied during each backpropagation step. Additionally, an early stopping strategy was implemented, monitoring the validation set loss and terminating the training process when the loss ceased to decrease significantly.

4. Experimental Results

The entire training process was set to 200 epochs. During each training epoch, the losses for both the training set and the validation set were computed. The training loss and validation loss curves were plotted, as shown in Figure 2.

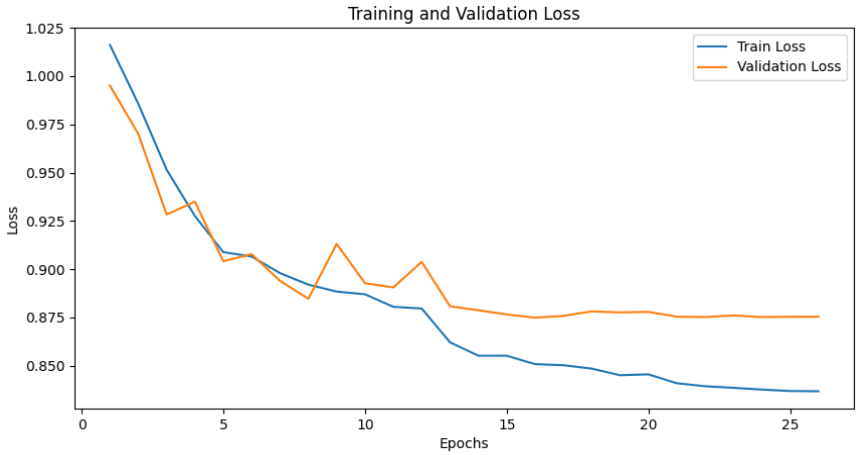


Figure 2: Training and Validation Loss

From the training loss and validation loss curves, it can be observed that in the initial stage, both the training loss and validation loss rapidly decreased to around 0.90. Between epochs 5 and 15, the rate of

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

decrease slowed, and after 15 epochs, the curves stabilized, with the training loss and validation loss remaining in the range of 0.85 to 0.88.

By loading the pre-trained model, signals under different Signal-to-Noise Ratio conditions were classified, and the model's accuracy was calculated. The trend of classification accuracy with varying SNR values was obtained. Figure 3 shows the classification accuracy under different SNR values for a model including only the convolutional neural network and for the model with the incorporated Transformer. At SNR values below -5dB, the accuracy was low, and the classification effect was not significant. As the SNR increased, the accuracy gradually improved, indicating that the model performed better in low-noise environments. At SNR values of approximately 20dB and above, the accuracy stabilized. The improved model's classification accuracy reached over 90%, whereas the classification accuracy of the model before improvement was only around 75%. This demonstrates that the incorporation of the Transformer effectively enhances the classification performance.

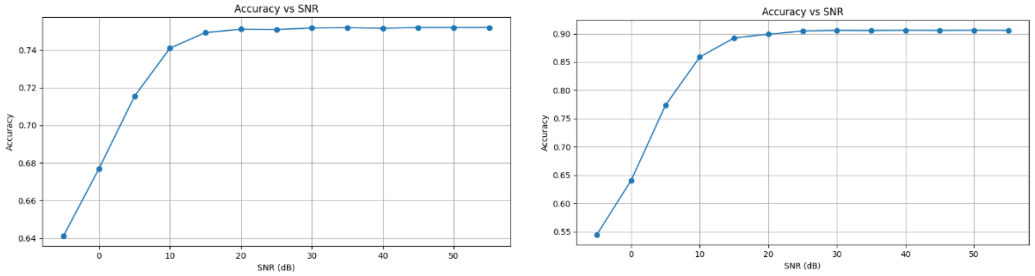
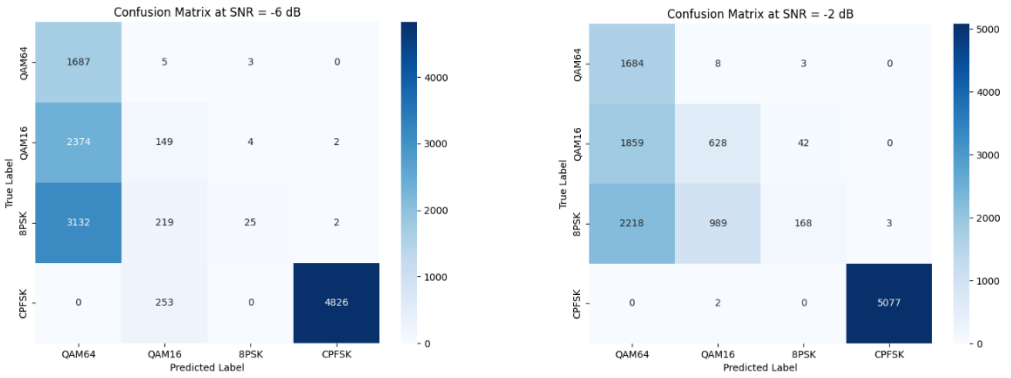


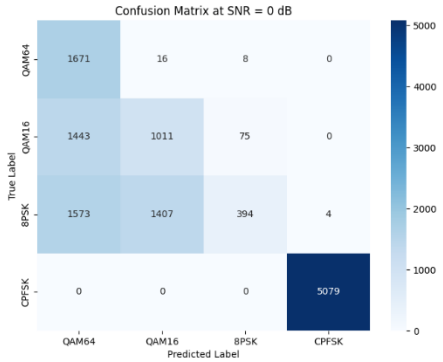
Figure 3: Classification Accuracy at Different SNR Values Before and After Model Improvement

To further analyze the recognition accuracy of various signal modulation schemes, the experiment employed a confusion matrix method to evaluate the classification accuracy of four types of signals under different Signal-to-Noise Ratio conditions. This approach allows for a more intuitive presentation of the model's classification performance for different types of signals in various noise environments. As shown in Figure 4, the horizontal axis of the confusion matrix represents the categories that the model predicted the input signals to belong to, while the vertical axis represents the true categories of the input signals. Each cell (i, j) contains a number indicating the quantity of samples with true label i that were predicted by the model as category j. Under low SNR conditions at -6 dB, CPFSK classification performed very well, while some QAM64 samples were misclassified as QAM16 and 8PSK. Most QAM16 and 8PSK samples were misclassified, particularly with QAM16 being frequently misclassified as 8PSK. As the SNR increased, the performance of the classification model gradually improved, and the instances of misclassification decreased. At an SNR of 18 dB, the classification accuracy for CPFSK reached 100%, while the classification accuracies for QAM64 and 8PSK were approximately 90%. However, the classification accuracy for QAM16 remained relatively lower at around 70%.

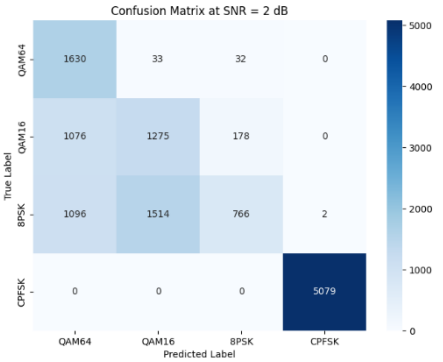


(a) Confusion Matrix at SNR = -6 dB *(b) Confusion Matrix at SNR = -2 dB*

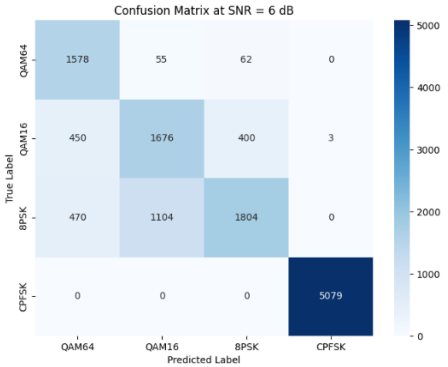
ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426



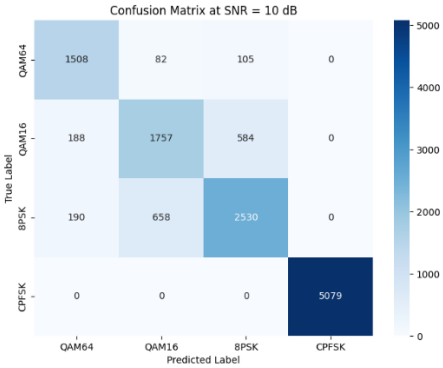
(c) Confusion Matrix at SNR = 0 dB



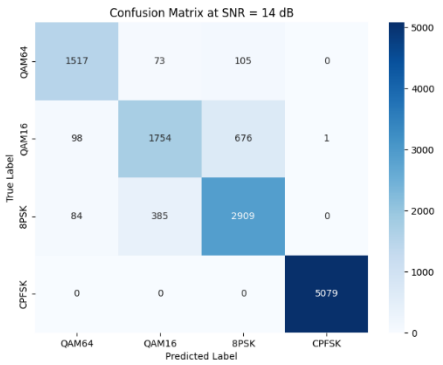
(d) Confusion Matrix at SNR = -2 dB



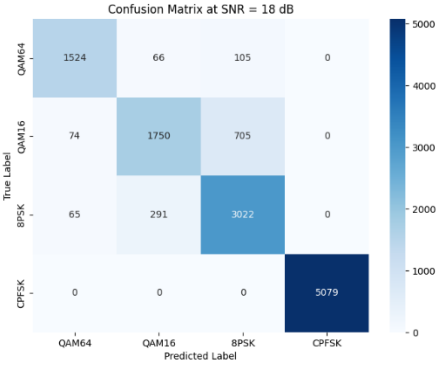
(e) Confusion Matrix at SNR = 6 dB



(f) Confusion Matrix at SNR = 10 dB



(g) Confusion Matrix at SNR = 14 dB



(h) Confusion Matrix at SNR = 18 dB

Figure 4: Confusion Matrix for the Recognition of Four Modulation Schemes under Different SNR Conditions

5. Conclusions

This paper proposes a method that combines Convolutional Neural Networks and Transformer models to process and classify electromagnetic signals. The experiments generated signal data with different modulation schemes by constructing a signal transmission model, and designed and trained a deep learning network for feature extraction and classification. Compared to traditional electromagnetic signal recognition methods, this approach does not require manual feature design and can automatically extract deep features. Additionally, by leveraging the local feature extraction capabilities of CNNs and the global dependency capturing abilities of Transformers, this method exhibits good generalization performance when handling signals with various modulation schemes.

With the advancement of computational capabilities, models combining CNNs and Transformers can efficiently recognize and classify signals in complex electromagnetic environments, making them applicable in fields such as wireless communication and radar signal processing. Through further

ISSN 2616-5767 Vol.7, Issue 4: 187-195, DOI: 10.25236/AJETS.2024.070426

research, this model can also address open set recognition problems, identifying signals of unknown categories. Moreover, this method can be extended to multimodal signal processing, integrating data from different sensors and signal sources to meet the demands of intelligent and automated applications.

References

[1] Cui M, Wu Z, Lu Y, et al. Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions [J]. IEEE Communications Magazine, 2022, 61(1): 40-46.
[2] Li Q, Yang S, Chen H. Performance Evaluation System Based on Multi-Indicators for Signal Recognition [J]. IEEE Access, 2022, 11: 2820-2830.
[3] Li F Z, Liu Y, Wu P X, et al. A survey on recent advances in meta-learning[J]. Chinese Journal of Computers, 2021, 44(2): 422-446.
[4] Huang J, Wu B, Li P, et al. Few-shot learning for radar emitter signal recognition based on improved prototypical network[J]. Remote Sensing, 2022, 14(7): 1681.
[5] Shao P, Chen Z. Radio frequency fingerprint feature extraction based on I/Q data distribution features[C]//2022 Photonics & Electromagnetics Research Symposium (PIERS). IEEE, 2022: 161-165.
[6] Wang H, Li J, Guo L, et al. Fractal complexity-based feature extraction algorithm of communication signals [J]. Fractals, 2017, 25(04): 1740008.
[7] Xi H C, Li B, Mai W H, et al. Feature extraction for evaluating the complexity of electromagnetic environment based on adaptive multiscale morphological gradient and nonnegative matrix factorization [J]. Journal of Electrical and Computer Engineering, 2022, 2022.
[8] Yang B, Hong T, Ma L, et al. A Recognition Algorithm for Complex Spatial Electromagnetic Signal Perception Based on KNN[C]//2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT). IEEE, 2022: 1-3.
[9] Wu J, Wu B, Niu H, et al. A Novel Support Vector Machine Based Radar Individual Recognition Algorithm Under Inconsistent Noise Condition[C]//IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2020: 2847-2850.
[10] Chirov D S, Kandaurova E O. Synthesis of informative features for recognition of the type of pulse repetition interval modulation of signals from radars[C]//2019 Systems of Signals Generating and Processing in the Field of on Board Communications. IEEE, 2019: 1-6.
[11] Malekian A, Chitsaz N. Concepts, procedures, and applications of artificial neural network models in streamflow forecasting[M]//Advances in streamflow forecasting. Elsevier, 2021: 115-147.
[12] Taherdoost H. Enhancing social media platforms with machine learning algorithms and neural networks [J]. Algorithms, 2023, 16(6): 271.
[13] Chen Z, Shi Y, Wang Y, et al. Unlocking signal processing with image detection: a frequency hopping detection scheme for complex EMI environments using STFT and CenterNet[J]. IEEE Access, 2023.
[14] Tu Y, Lin Y, Hou C, et al. Complex-valued networks for automatic modulation classification[J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 10085-10089.
[15] Hou C, Fu D, Hua L, et al. The recognition of multi-components signals based on semantic segmentation [J]. Wireless Networks, 2023, 29(1): 147-160.
[16] Liu Z, He T, Wu T, et al. Open-set recognition of LPI radar signals based on a slightly convolutional neural network and support vector data description[J]. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, 2024, 37(2): e3213.
[17] Tao Guanhong, Liao Kaisheng, Zhou Lin. A New Method for Electromagnetic Signal Modeling and Modulation Recognition Based on Deep Learning[J]. Electronic Information Countermeasure Technology, 2019, 34(5): 10-15.
[18] Ashish V. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: I.