Signal Modulation Recognition based on Convolutional Autoencoder and Time-Frequency Analysis

Xiangyu Wu
Harbin Engineering University
College of Information and Communication Engineering
Harbin, Heilongjiang, China
wu_xiang_yu@126.com

Jing Zhang
Harbin Engineering University
College of Information and Communication Engineering
Harbin, Heilongjiang, China
zhangjing03@hrbeu.edu.cn

Changbo Hou*
Key Laboratory of Advanced Marine Communication and Information Technology, Ministry of Industry and Information Technology
Harbin Engineering University
Harbin, Heilongjiang, China
houchangbo@hrbeu.edu.cn

Guowei Liu
Harbin Engineering University
College of Information and Communication Engineering
Harbin, Heilongjiang, China
1275589137@hrbeu.edu.cn

Jie Zhang
Harbin Engineering University
College of Information and Communication Engineering
Harbin, Heilongjiang, China
s320080159@hrbeu.edu.cn

Jin Liu
Harbin Engineering University
College of Information and Communication Engineering
Harbin, Heilongjiang, China
lj1727391387@126.com

***Abstract*—With the development of communication technology, patterns of communication signals have become more complex and diverse. Therefore, the technology of identifying the modulation mode of the communication signal in transmission, especially the modulation recognition technology based on artificial intelligence, has become an extremely important technology in the communication field. For a variety of modulation methods, the traditional method is extremely complicated to implement, and cannot meet the requirements of accurate identification in a short time. In order to increase the speed and reduce the redundancy, this paper proposes a method based on the convolutional autoencoder and the residual network which can realize the denoising, identification and classification of different modulated signals. This method generates ten different modulation types of signals under each signal-to-noise ratio. After the model is trained, the data set is input to the convolutional autoencoder to denoise, and then the data set denoised by the autoencoder is input to the residual network to obtain the classification and recognition accuracy of each modulation type. And an average recognition rate of 92.86% was achieved at -2dB.**

Keywords: Time-frequency diagram; Convolutional autoencoder; Residual network; Modulation recognition

I. Introduction

With the rapid development of communication transmission, signal processing, pattern recognition and other fields, automatic modulation classification (AMC) [1]-[3] has become one of the most important technologies in the field of communication transmission. With the development of communication equipment, the transmission speed of signals has become more and more rapid and the signals have become more and more difficult to identify. This is a very headache for technicians engaged in traditional modulation recognition. Therefore, how to achieve fast and accurate recognizing the modulation type of communication signals has been a hot topic in recent years. In the civil field, in order to ensure the legal and orderly conduct of communications, modulation recognition technology is widely used in non-professional electromagnetic spectrum monitoring [4] and management; In the military field, if the detected electromagnetic signal can accurately and quickly identify the modulation type used by the opponent's communication, not only can the function and complexity of the opponent's communication equipment be judged, but also the information transmitted by the opponent can be estimated, then take the lead in the electronic warfare [5].

II. Literature Review

There are many traditional modulation recognition methods, and they can be roughly divided into maximum likelihood hypothesis [6] and pattern recognition [7]. The principle of traditional modulation recognition can be divided into three parts: signal preprocessing [8], feature extraction [9] and classification and recognition [10]. Signal preprocessing includes carrier synchronization, frequency down conversion, noise suppression, and estimation of parameters such as signal-to-noise ratio (SNR), symbol period, and carrier frequency; Feature extraction is the most important part of modulation recognition. The pre-defined characterization signal is extracted from the data, and the characteristics of the modulation type are selectively extracted from the characterization signal. This process is to extract the feature parameters, and the feature extraction includes analysis methods in time domain[11], frequency domain[12], and transform domains [13]; Classification and recognition is to select and determine the appropriate decision rules and recognition classifiers on the basis of extracting the feature parameters.

However, the traditional modulation recognition process requires cumbersome feature extraction, and the generated manual signal [14] has serious uncertainties, making the traditional method not suitable for complex communication environments. As a result, the generalization performance and

2767-6684/21/$31.00 ©2021 IEEE
DOI 10.1109/DSA52907.2021.00096

anti-noise performance of traditional modulation recognition technology are poor. Therefore, the method based on deep learning has become a more common method to identify the modulation type of communication signals. There are many ways to apply deep learning to the field of modulation recognition, for example, based on the Inception network can improve the convergence speed and expand the receiving domain [15]. Under low SNRs, adopting the method based on capsule network [16] can improve the accuracy of modulation recognition; In addition, the image-based signal recognition method is also a commonly used method in deep learning. The principle is to convert the signal into a time-frequency diagram [17] or constellation diagram [18], [19], and then input to the convolutional network. For example, the use of graph convolutional network (GCN) [20] method for signal modulation recognition can obtain higher recognition accuracy than other convolutional neural networks (CNN). Similar to the method of GCN, this paper proposes a method to convert the modulated signal under each SNR into a time-frequency diagram through time-frequency transformation. The time-frequency diagram is used as the identification medium, and the convolutional autoencoder (CAE) is used to denoise.

III. PRINCIPLES OF METHOD

This section presents the model of the CAE and the model of the residual network.

A. *Time-frequency Diagram*

First do Fourier transform of the simulated signal, and then use time as the horizontal axis, frequency as the vertical axis, and use color to indicate the amplitude to draw a spectrogram. A picture contains the frequency and amplitude of the signal changes over time, so it is also called "time-frequency diagram". Time-frequency diagram is one of the very common ways to analyze signals. After the signal is converted into a two-dimensional color image, the relationship between frequency and time can be observed in a more intuitive way.

The conventional Fourier transform cannot express the frequency components at any time. In order to make a comprehensive analysis, a time-frequency analysis method is usually used to transform a one-dimensional time-domain signal into a two-dimensional time-frequency plane. There are many methods of time-frequency analysis, such as: spectrogram, Short-time Fourier Transform (STFT), Wigner-Ville Distribution (WVD), Pseudo Wigner-Ville Distribution (PWVD), Smoothed Pseudo Wigner-Ville Distribution (SPWVD), etc.

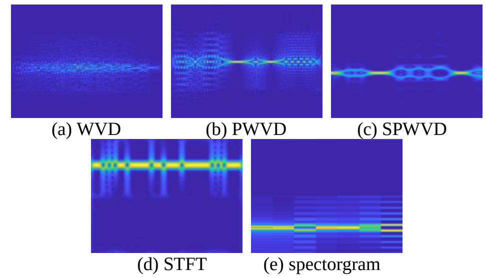

Figure 1. Time-frequency diagrams generated by various time-frequency transformations

WVD is a typical quadratic transform, which is defined as the Fourier transform of the signal instantaneous correlation function, reflecting the instantaneous time-frequency relationship of the signal. Because there is no window function involved in WVD, there is no restriction between the time domain and the frequency domain in WVD, so the resolution is very high. On the other hand, since there is no window function in WVD, it is susceptible to interference from cross-terms. Fig.1 shows the time-frequency diagrams generated by several commonly used time-frequency transformations under the same conditions. It can be clearly seen from the figure that compared to WVD and PWVD, SPWVD has a better suppression effect on cross terms.

The design purpose of PWVD is to suppress cross-term interference, introduce parameter optimization in WVD, and construct a matching transformation kernel according to the selection of parameters. If the transformation kernel and the signal model are relatively consistent, better time-frequency concentration can be achieved. The values of WVD and PWVD do not satisfy the energy positiveness in the conventional sense. In order to achieve the positive energy distribution characteristics, the SPWVD can be obtained by convolving WVD with a smoothing function. Its distribution definition is:

$$SPWVD_{g,h}(x;t,f) = \iint g(u)h(\tau)x(t-u+\tau/2)x^{*}(t-u-\tau/2)e^{2j\pi f\tau}du\,d\tau \tag{1}$$

Because there is a window function g (u) in SPWVD for smoothing in both time and frequency, it has a better effect of eliminating cross terms. Its time-frequency characteristics and aggregation performance are also maintained well, so in all possible Cohen-like time-frequency distributions, SPWVD is one of the most general distributions. Moreover, in the case of a narrow frequency band, choosing SPWVD can obtain higher time-frequency resolution, and the advantages are very prominent.

B. *Convolutional Autoencoder*

In deep learning, autoencoder is a very commonly used unsupervised learning model. The autoencoder is mainly composed of two parts: an encoder and a decoder. The encoder encodes the original representation into a hidden layer representation, and the decoder decodes the hidden layer representation into the original representation. It can also be said that the former compresses the input into a latent space representation, and the latter reconstructs the input from the hidden space. The autoencoding process is shown in Fig.2. When the time-frequency diagram is input to the encoder, the encoder will generate a code, which represents a representation of the input time-frequency diagram in the latent space. Then input the code which output by the encoder into the decoder. The decoder will generate a message. In theory, the information generated by the decoder should be exactly the same as the input time-frequency diagram. In practice, there must be a gap between the two parts. The gap is called the reconstruction error, and the reconstruction error can be reduced by adjusting the parameters of the encoder and decoder.

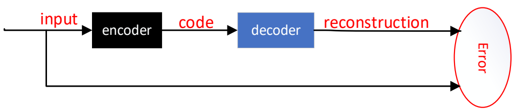

Figure 2. Operation process of autoencoder

CAE is based on unsupervised learning of the autoencoder, and includes convolution and pooling of CNN. The structure diagram is shown in Fig.3, and its training goal is to minimize the reconstruction error, and realize the sample reconstruction by using the mapping relationship between the input layer and the output layer to extract features.

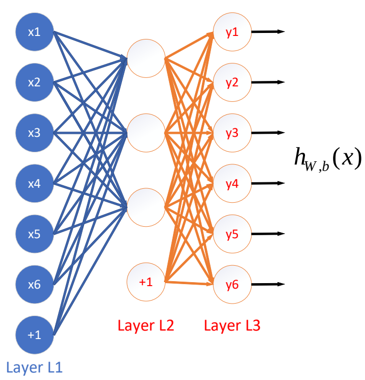

Figure 3. Autoencoder structure

Suppose there are k convolution kernels：

$$h^k = \sigma(x * \omega^k + b^k) \quad (2)$$

Comparing the input sample with the result obtained by using feature reconstruction finally, and optimizing through BP algorithm, a complete CAE can be obtained：

$$E = \frac{1}{2n}\sum (x_i - y_i)^2 \quad (3)$$

C. Residual Network

In CNN, as the number of convolutional layers increases, the loss value of the training set will decrease accordingly, and then tends to be saturated. However, this rule is limited. When the number of layers increases to a certain value, the training loss value will not decrease, but will increase sharply. This phenomenon is called degradation, at this time, the features learned by the deep network are almost the same as those learned by the shallow network, which also makes deepening the layer meaningless. In addition, if you continue to increase the number of layers after degradation has occurred, the gradient will disappear. This phenomenon is called gradient explosion.

Compared with the traditional CNN, the residual network can reach a higher depth, which means that the extracted feature values are more reliable and more representative. The data volume of a color RGB image is the total multiplied by its length, width, and color gamut, and each data set will have hundreds or thousands of images. For such a huge amount of data, the traditional CNN has poorer features and recognition speed. Therefore, we adopt a residual network that can learn a deeper level.

The residual network is composed of many columns of residual blocks. A residual block in the same layer can be expressed as:

$$x_{l+1} = x_l + \mathcal{F}(x_l, W_l) \quad (4)$$

For a deeper level, the relationship between L and l can be expressed as:

$$x_L = x_l + \sum_{i=l}^{L-1} \mathcal{F}(x_i, W_i) \quad (5)$$

According to the chain rule of derivatives in BP, the gradient of the loss function with respect to can be expressed as:

$$\begin{aligned}\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x_l} &= \frac{\partial \varepsilon}{\partial x_L}\left(1 + \frac{\partial}{\partial x_l}\sum_{i=l}^{L-1}\mathcal{F}(x_i, W_i)\right) \\ &= \frac{\partial \varepsilon}{\partial x_L} + \frac{\partial \varepsilon}{\partial x_L}\frac{\partial}{\partial x_l}\sum_{i=l}^{L-1}\mathcal{F}(x_i, W_i)\end{aligned} \quad (6)$$

As shown in (5), during the whole training process, $\frac{\partial}{\partial x_l}\sum_{i=l}^{L-1}\mathcal{F}(x_i, W_i)$ cannot always be -1, that is to say, the problem of gradient disappearance will not occur in the residual network.

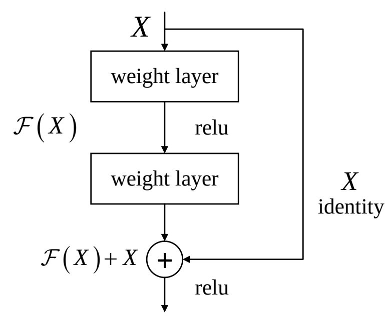

Figure 4. Residual unit implemented in the form of layer jump connection

In the residual network, the residual unit is realized in the form of layer jump connection, that is, the input of the unit and the output of the unit are directly added together, and then activated, the connection form is shown in Fig.4. From a macro perspective, when propagating between layers, the input signal can propagate directly from any lower layer to the upper layer. Since it contains a natural identity mapping, the problem of network degradation can be solved.

IV. RESULTS AND ANALYSIS

The data set of this paper has 10 signal modulation types, which are 2ASK, 4ASK, 2FSK, 4FSK, 8FSK, BPSK, QPSK, 16QAM, 32QAM, and 64QAM. The simulation data experiment parameters are shown in TABLE 1.

TABLE 1. SIMULATION DATA EXPERIMENTAL PARAMETERS

Signal parameters	***Accurate value***
Sampling rate	40kHz
Carrier frequency	4-16kHz
Sampling points	500
The number of signals for each modulation for each SNR	200

The SNR is -6dB to 4dB, and the step size is 2dB. The signal of each modulation method generates 200 time-frequency diagrams as a data set under a single SNR. The maximum number of iterations is 1240, a total of 20 rounds, 62 iterations per round. Fig.5 is the modulation recognition rate of the two modulation signals before and after denoising. Firstly, Fig.5 shows that as the SNR increases, the recognition rate is also increasing. Secondly, it can be clearly seen that the recognition rate after denoising is higher than before denoising, especially for the most obvious improvement of the recognition rate when SNR=-6dB, which can prove the effectiveness of the denoising method.

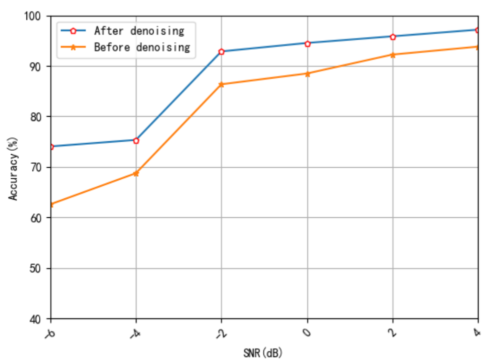

Figure 5. Accuracy under different SNRs

Confusion matrix in Fig.6 shows that QPSK and 32QAM have a certain degree of confusion. How to effectively identify signals with similar modulation types is also the next step to be studied. In addition, other modulation methods have very good classification effects. According to the degree of convergence of the confusion matrix and the loss curve, the effectiveness of the residual network in modulation recognition can be proved.

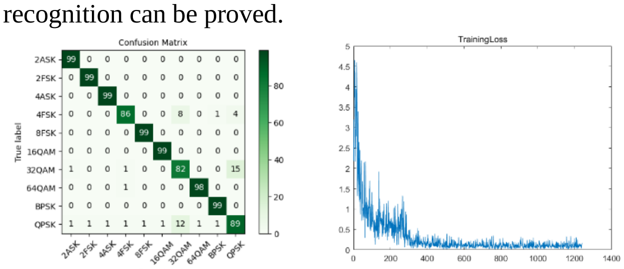

Figure 6. Confusion matrix and loss curve under 4dB

The confusion of recognition mainly occurs in the modulation types of BPSK, QPSK and 16QAM, because the time-frequency diagrams of the three modulation signals are affected by background noise, some subtle features become blurred or even lost, resulting in insufficient characterization of the extracted features and a decrease in the classification accuracy.

V. CONCLUSION

Based on deep learning, this paper uses CAE for denoising and residual network to recognize the modulation of the simulated signal. First, the signal is subjected to SPWVD time-frequency transformation to obtain time-frequency diagrams with different SNRs, and then denoised by the CAE, and finally the denoised images are input to the residual network to identify and classify the modulated signal. Experimental results show that the recognition rate under -6dB has increased by 12%, and the recognition rate under other SNRs has also increased by at least 6%, it can be seen that the denoising effect is good. The recognition rate after 4dB denoising reached 97.16%, which shows that the recognition and classification effect is good.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (62001137); the National Key Research and Development Program of China under Grant 2018AAA0102702; the Natural Science Foundation of Heilongjiang Province (JJ2019LH2398); the Fundamental Research Funds for the Central Universities (3072021CF0805).

REFERENCES

[1] P. Ghasemzadeh, S. Banerjee, M. Hempel and H. Sharif, "Accuracy Analysis of Feature-based Automatic Modulation Classification with Blind Modulation Detection," *2019 International Conference on Computing, Networking and Communications (ICNC)*, 2019, pp. 1000-1004, doi: 10.1109/ICCNC.2019.8685638.

[2] C. B. Hou, X. Zhang, X. Chen, “Electromagnetic Signal Feature Fusion and Recognition based on Multi-Modal Deep Learning” [J]. Int J Performability Eng, vol.16, no.6, pp.941-949,2020.

[3] C. B. Hou, G. W. Liu, L. J. Hua, Y. Lin, "Radar Signal Separation and Recognition based on Semantic Segmentation," *2020 7th International Conference on Dependable Systems and Their Applications (DSA), 2020*, pp. 385-390, doi: 10.1109/DSA51864.2020.00068.

[4] J. Villain, V. Deniau, A. Fleury, E. P. Simon, C. Gransart and R. Kousri, "EM Monitoring and Classification of IEMI and Protocol-Based Attacks on IEEE 802.11n Communication Networks," in *IEEE Transactions on Electromagnetic Compatibility*, vol. 61, no. 6, pp. 1771-1781, 2019, doi: 10.1109/TEMC.2019.2900262.

[5] M. E. Yıldırım, Y. B. Salman and J. Park, "Clustering for Electronic Warfare Information," *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, 2018, pp. 1195-1197.

[6] S. Yang and D. Qin, "Improved Maximum Likelihood Syncmark Detection for Magnetic Recording Channels," in *IEEE Transactions on Magnetics*, vol. 49, no. 7, pp. 3691-3694, 2013, doi: 10.1109/TMAG.2013.2248054.

[7] N. Yu, Q. Xu and H. Wang, "Wafer Defect Pattern Recognition and Analysis Based on Convolutional Neural Network," in *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 4, pp. 566-573, 2019, doi: 10.1109/TSM.2019.2937793.

[8] Q. Qiu, "Design of multi channel gain-control signal preprocessing equipment," *2013 IEEE International Conference of IEEE Region 10*

(TENCON 2013), 2013, pp. 1-4, doi: 10.1109/TENCON.2013.6719055.

[9] Z. Han, J. Wang, S. Lun and X. Wang, "Robust feature extracting of speech signal based on wavelet packet transform," *Proceedings of the 29th Chinese Control Conference*, 2010, pp. 2832-2837.

[10] C. B. Hou, G. W. Liu, L. J. Hua, Y. Lin, "Multimodal Feature Fusion Recognition of Modulated Signals Based on Image and Waveform Domain," *2020 7th International Conference on Dependable Systems and Their Applications (DSA)*, 2020, pp. 337-342, doi: 10.1109/DSA51864.2020.00061.

[11] S. S. A. Obayya and R. Letizia, "Multiresolution Time Domain analysis of optical guided-wave devices," *2007 Workshop on Computational Electromagnetics in Time-Domain*, 2007, pp. 1-4, doi: 10.1109/CEMTD.2007.4373530.

[12] D. Namdev and A. Bansal, "Frequency Domain Analysis for Audio Data Forgery Detection," *2015 Fifth International Conference on Communication Systems and Network Technologies*, 2015, pp. 702-705, doi: 10.1109/CSNT.2015.168.

[13] A. L. Wang, C. Y. Qi, J. W. Dong, S. L. Meng and D. M. Li, "Multimodal medical image fusion in nonsubsampled contourlet transform domain," *Proceedings of 2013 2nd International Conference on Measurement, Information and Control*, 2013, pp. 169-172, doi: 10.1109/MIC.2013.6757939.

[14] N. Ding, Q. He, C. Wu and J. Fetzer, "Modeling Traffic Control Agency Decision Behavior for Multimodal Manual Signal Control Under Event Occurrences," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2467-2478, Oct. 2015, doi: 10.1109/TITS.2015.2409174.

[15] Y. Tang, W. Gong, X. Chen and W. Li, "Deep Inception-Residual Laplacian Pyramid Networks for Accurate Single-Image Super-Resolution," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1514-1528, 2020, doi: 10.1109/TNNLS.2019.2920852.

[16] L. Li, J. Huang, Q. Cheng, H. Meng and Z. Han, "Automatic Modulation Recognition: A Few-Shot Learning Method Based on the Capsule Network," in *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 474-477, 2021, doi: 10.1109/LWC.2020.3034913.

[17] C. Hou, Y. Li, X. Chen, J. Zhang, “Automatic modulation classification using KELM with joint features of CNN and LBP”, *Physical Communication*, vol. 45, no. 3, pp. 101259, April 2021.

[18] S. Wu, X. Tao and Y. Chen, "A New Constellation Diagram-Based Signal Detection for GSM-MIMO Systems," in *IEEE Communications Letters*, vol. 22, no. 3, pp. 502-505, 2018, doi: 10.1109/LCOMM.2017.2776948.

[19] T. Huynh-The, C. -H. Hua, V. -S. Doan, Q. -V. Pham, T. -V. Nguyen and D. -S. Kim, "Deep Learning for Constellation-based Modulation Classification under Multipath Fading Channels," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 300-304, doi: 10.1109/ICTC49870.2020.9289413.

[20] Y. Liu, Y. Liu, C. Yang, "Modulation Recognition With Graph Convolutional Network," in *IEEE Wireless Communications Letters,* vol. 9, no. 5, pp. 624-627, 2020, doi: 10.1109/LWC.2019.2963828.