

Article

An Efficient and Lightweight Model for Automatic Modulation Classification: A Hybrid Feature Extraction Network Combined with Attention Mechanism

Zhao Ma *, Shengliang Fang *, Youchen Fan, Gaoxing Li and Haojie Hu

School of Aerospace Information, Space Engineering University, Beijing 101400, China

* Correspondence: nuaamazhao@163.com (Z.M.); eefsllyl@163.com (S.F.)

Abstract: This paper proposes a hybrid feature extraction convolutional neural network combined with a channel attention mechanism (HFECNET-CA) for automatic modulation recognition (AMR). Firstly, we designed a hybrid feature extraction backbone network. Three different forms of convolution kernels were used to extract features from the original I/Q sequence on three branches, respectively, learn the spatiotemporal features of the original signal from different “perspectives” through the convolution kernels with different shapes, and perform channel fusion on the output feature maps of the three branches to obtain a multi-domain mixed feature map. Then, the deep features of the signal are extracted by connecting multiple convolution layers in the time domain. Secondly, a plug-and-play channel attention module is constructed, which can be embedded into any feature extraction layer to give higher weight to the more valuable channels in the output feature map to achieve the purpose of feature correction for the output feature map. The experimental results on the RadiomL2016.10A dataset show that the designed HFECNET-CA has higher recognition accuracy and fewer trainable parameters compared to other networks. Under 20 SNRs, the average recognition accuracy reached 63.92%, and the highest recognition accuracy reached 93.64%.



Citation: Ma, Z.; Fang, S.; Fan, Y.; Li, G.; Hu, H. An Efficient and Lightweight Model for Automatic Modulation Classification: A Hybrid Feature Extraction Network Combined with Attention Mechanism. *Electronics* **2023**, *12*, 3661. <https://doi.org/10.3390/electronics12173661>

Academic Editors: Lin Li, Chuanjin Dai and Rui Guo

Received: 12 July 2023

Revised: 24 August 2023

Accepted: 29 August 2023

Published: 30 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: automatic modulation recognition; attention mechanism; deep learning; lightweight

1. Introduction

Automatic modulation recognition of signals is the foundation of signal demodulation in non-cooperative communication. As a key technology in the field of cognitive communication, it is an important prerequisite for achieving efficient spectrum sensing, understanding, and utilization in non-cooperative communication scenarios. In the military and civilian fields, signal automatic modulation recognition has a wide range of applications, such as spectrum monitoring, radio fault detection, signal interception, and interference [1,2].

Traditional signal modulation recognition methods can be divided into methods based on maximum likelihood theory [3,4] and expert features [5]. The method based on likelihood theory is to construct decision criteria and establish a maximum likelihood classifier based on obtaining the statistical characteristics of each modulation signal. However, due to its high computational complexity and narrow applicability, it is rarely applied in practical scenarios [6]. The method based on expert features is to transform the received signal into a certain feature space through specific analysis and processing and then design a classifier for classification. Typical expert features include cyclic spectrum feature, high-order Cumulant feature, and wavelet transform feature, while common classifiers include support vector machine [7], decision tree [8], artificial neural network [9], etc. The recognition accuracy of this type of method depends on the extracted statistical features and is limited by the weak learning ability of traditional classifiers. As a result, its overall recognition accuracy is generally low.

The rise of deep learning technology [10] in recent years has made significant progress in fields such as image classification and speech recognition, providing new research ideas for automatic modulation recognition. The powerful feature extraction and self-learning capabilities of deep learning technology are expected to address the limitations of traditional methods.

In 2016, O 'SHEA et al. first proposed a modulation recognition method using a convolutional neural network (CNN) to directly process the original in-phase and quadrature (I/Q) signals and opened the data set (RML2016.10a), attracting a large number of researchers to participate and promoting the development of this field [11]. Herma Wan et al. adjusted the structure based on CNN to further improve the identification accuracy of the model [12]. Rajendran et al. preprocessed the I/Q signal into amplitude and phase (A/P) form and then used LSTM for temporal feature extraction, achieving high recognition accuracy [13]. Y. Wang et al. converted the original I/Q signal into a constellation map and used CNN to carry out the research on modulation recognition from the perspective of image recognition [14]. In addition to CNN, Dehua Hong et al. noticed the advantages of RNN in time series feature extraction, introduced it into the research of automatic modulation recognition, and achieved good results [15]. Nathan E. West et al. combined the advantages of CNN and LSTM in spatial features and temporal features and proposed a CLDNN network with higher recognition accuracy [16]. Jia lang Xu et al. proposed an MCLDNN framework by extracting features from both single and combined I/Q sequences of modulated signals, further demonstrating the feasibility of hybrid networks [17]. Zhang et al. proposed a data preprocessing method that significantly improves the accuracy of CNN-based AMC recognition, proving that it is necessary to process I/Q signals into a data format suitable for CNN input [18].

To further improve the accuracy of modulation recognition, an intuitive idea is to fully utilize the different data forms of modulation signals, extract more feature information, and perform feature fusion or decision fusion. Tuo Wang et al. proposed a Multi-clue Fusion (MCF) framework that combines I/Q signals and A/P signals. The framework transformed the original I/Q data into three data forms for feature extraction and decision fusion, enriching the feature forms and greatly improving recognition accuracy, surpassing existing research results [19]. Seung Hwan Kim et al. designed a mixed deep learning model based on signal and image, which fused the two feature maps for comprehensive decision-making and achieved good results [20].

With the deepening of research, the recognition accuracy of AMC models based on deep learning is gradually improving, but at the cost of larger model sizes and higher computational complexity. However, in today's Edge devices with limited computing and storage capacity, such as IoT devices and UAV devices, it is impossible to deploy too large network models, and in actual confrontation scenarios, there is a high requirement for the response speed of models [21]. Therefore, the lightweight research of AMC models has gradually become a research hotspot and an urgent problem to be solved. Thien Huynh and Fuxin Zhang et al. proposed lightweight AMC architecture based on CNN [22] and CNN + LSTM [23], respectively. Shi Fengyun and Fan Jia introduced an attention mechanism based on designing lightweight CNN architecture, which contributed to the lightweight of the AMC model [24,25]. Wang et al. introduced deep separable convolutions, which is a lightweight convolutional structure, into the design of AMC models, significantly reducing the model's parameter count at the expense of certain recognition accuracy [26]. Chenghong Xiao et al. proposed a novel complex-valued deep separable convolutional neural network structure, which combines the rich representation ability of complex-valued neural networks and the lightweight advantages of separable convolutions [27].

Although the existing AMC model has a relatively high recognition performance, and the research on model lightening has made some progress, there is still much room for improvement in the balance between model size and model accuracy. Therefore, in this paper, we propose a lightweight and excellent performance AMC framework, which effectively solves the contradiction between the model size and the model precision in the

current AMC field. The framework can achieve high accuracy by only using the original I/Q sequence and a simple convolutional neural network.

The main contributions of this paper are as follows:

- (1) A hybrid feature extraction module based on CNN is designed to extract the signal features more effectively, three different convolution kernels are designed to extract the feature maps from the original I/Q sequence from different perspectives, and the mixed feature map is obtained by channel fusion. Furthermore, we design a time-domain convolution module to extract the deep features of the mixed feature image. In all the feature extraction modules, we use smaller convolution kernels and down-sampling to reduce the computational complexity, which can make the backbone network extremely lightweight.
- (2) Channel attention module is introduced. A lightweight channel attention module is added to the feature extraction module to further enhance the ability of model feature expression. This module compresses the feature maps of each channel into a channel descriptor through self-learning, obtaining the importance of each channel feature map and playing a role in feature correction. It greatly improves the performance of the model while not causing a parameter burden.

2. Signal Model and the Proposed Model

2.1. Signal Model

This article considers a single input, single output system, assuming that the observed signal received by the receiver is $X(t)$, which can be represented as [28]

$$X(t) = h(t) * s(t) + n(t), \quad (1)$$

where $h(t)$ represents the channel impulse response, including multiplicative noise such as channel fading, $s(t)$ represents the modulation signal generated by the transmitter, and $n(t)$ represents additive white Gaussian noise (AWGN). The problem of signal modulation recognition can be described as determining the modulation method of $s(t)$ by observing the signal $X(t)$.

After sampling by the receiver, in-phase/quadrature signal $X(n)$ is obtained:

$$X(N) = \left[(x_1^I, x_1^Q), (x_2^I, x_2^Q), \dots, (x_N^I, x_N^Q) \right], \quad (2)$$

where N is the sampling length. Therefore, when I/Q signals are used as network inputs, their data format can be written as

$$X_{\text{input}} = \begin{bmatrix} x_1^I & x_2^I & \dots & x_N^I \\ x_1^Q & x_2^Q & \dots & x_N^Q \end{bmatrix}. \quad (3)$$

2.2. Proposed Model

Figure 1 proposed a hybrid feature extraction model (HFECNET-CA) combined with an attention mechanism. HFECNET-CA improves the information utilization rate of the original signal by increasing the network width and introduces the channel attention mechanism to make the network more “focused” on more key feature channels to realize higher recognition accuracy with less calculation cost. It consists of a backbone network (HFECNET) and an additional channel attention module (SE block), whereas the backbone network consists of a hybrid feature extraction module, a deep temporal feature extraction module, and a classification module.

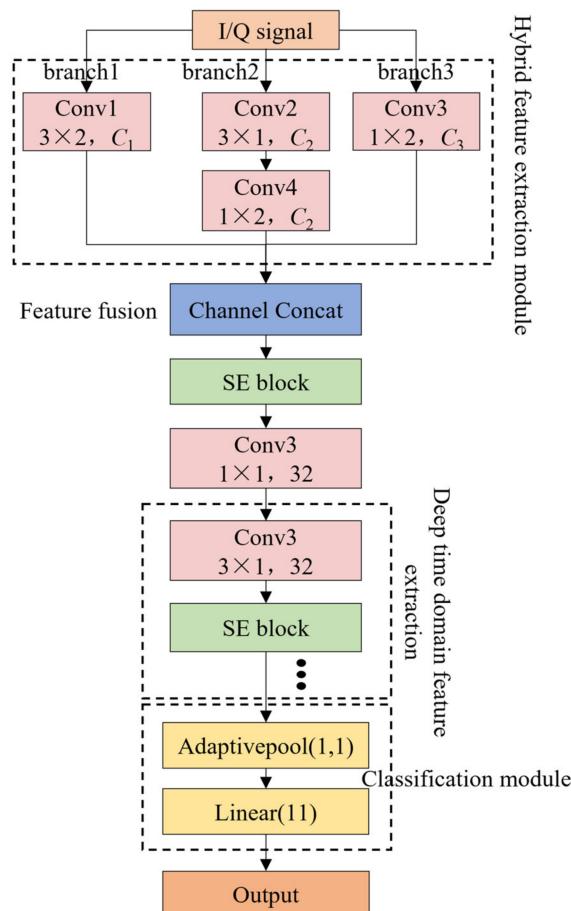


Figure 1. The structure of the proposed HFECNE-CA.

2.2.1. The Backbone Network—HFECNET

Existing studies have proved that a deeper network in the field of modulation recognition cannot improve the network performance but will increase the network parameters and calculation amount, which can result in network overfitting [29]. Therefore, we consider increasing the width of the network to improve the ability of network feature extraction and the information utilization of the original data. For the original I/Q signal ($N \times 2$), we redesign the feature extraction part of the network, i.e., the aforementioned hybrid feature extraction module and deep temporal feature extraction module.

When humans recognize something, they usually observe and summarize its characteristics from different perspectives in order to distinguish the similarities and differences between things. For convolutional neural networks, different forms of convolutional kernels mean different “perspectives” for feature extraction. The data format of the modulated signal is $N \times 2$. Therefore, when using CNN for feature extraction, the convolutional kernel can take three forms: $n \times 2$, $n \times 1$, 1×2 ($n > 1$).

We believe that each form of convolutional kernel represents observing and extracting signal features from different perspectives. The convolutional kernel in the form of $n \times 2$ can perform convolution operations on both IQ signals in the time domain, so it can be considered as the extraction of general features of the signal. The $n \times 1$ form convolution kernel can only perform convolution operations on I-path or Q-path signals at the same time, so it can be considered as the extraction of shallow time-domain features of single-path signals. The convolutional kernel in the form of 1×2 performs convolution operations on I-path and Q-path signals at the same sampling position, which can be considered as the extraction of I/Q-related features. The size of n represents the size of the convolutional nucleus receptive field. To ensure sufficient receptive field and perception of local features, n can be taken as 3.

Therefore, when designing the hybrid feature extraction module, we used three different forms of convolutional branches to extract signal features from different perspectives. The first branch uses a 3×2 convolution kernel to extract general features; the second branch firstly adopts a 3×1 convolution kernel to respectively extract the time-domain characteristics of the I and Q paths and then connects a 1×2 convolution kernel for extracting the deep I/Q correlation characteristics. The third branch directly uses a 1×2 convolution kernel to extract I/Q-related features. Through the above three branches, the amplitude and phase information between complex signals I and Q can be fully utilized to extract features from different levels and perspectives. The horizontal dimension of the output feature map of each branch is reduced from 2 to 1 by the corresponding convolution operation, and the subsequent calculation amount is reduced by 50%. In addition, at the end of each branch, a max-pooling layer with a stride length of (2, 1) is added to reduce the dimension of the characteristic image in the time-domain direction so that the size of the output characteristic image is further reduced, and the subsequent calculation cost is reduced. At last, we obtain the mixed feature map of the original signal by channel fusion of the feature maps of the three branches. Assuming that the number of channels of the output characteristic graph of each branch is C_1 , C_2 , and C_3 , the size of the final output characteristic graph of the module is $(C, N/2, 1)$, where $C = C_1 + C_2 + C_3$.

In the time-domain feature extraction module, we use the experience of the VGG network for reference and use multiple 3×1 small convolution kernels to extract deep time-domain features to reduce the number of convolution kernel parameters under the condition of ensuring the receptive field. After each convolution kernel, the max-pooling layer is used to reduce the data dimension, which further reduces the time-domain direction calculation.

In the classification stage, the Flatten operation is replaced by Adaptive Average Pooling, which compresses the feature map of each channel into a new feature value to improve the generalization performance of the network. Finally, only one fully connected layer is used in the classification, which avoids a large number of training parameters and calculation amounts caused by using multiple fully connected layers.

At the same time, between the convolution layer and the activation function, a Batch Normalization (BN) operation is added to increase the robustness and training speed of the model and prevent the network from overfitting; the network uses ReLU as the activation function and Sigmoid as the classification function in the classification layer.

2.2.2. Channel Attention Module

Mnih V et al. first proposed the attention mechanism in the field of image recognition [30]. The so-called attention mechanism is to focus on the information that is more critical to the current task to obtain more detailed information about the target that needs to be focused on while suppressing other useless information, thus improving the efficiency and accuracy of task processing.

Our work introduces the channel attention mechanism (SE Block) proposed in [31] to further improve information utilization and network performance. SE Block obtains the importance of each channel of the feature map by self-learning and then uses this importance to assign a weight value to each feature so that the model focuses on the key feature channels. And we made some small changes based on the original structure to avoid the side effects caused by feature dimension reduction. The implementation process of SE Block is shown in Figure 2, which mainly includes three parts: Squeeze, Excitation, and Scale.

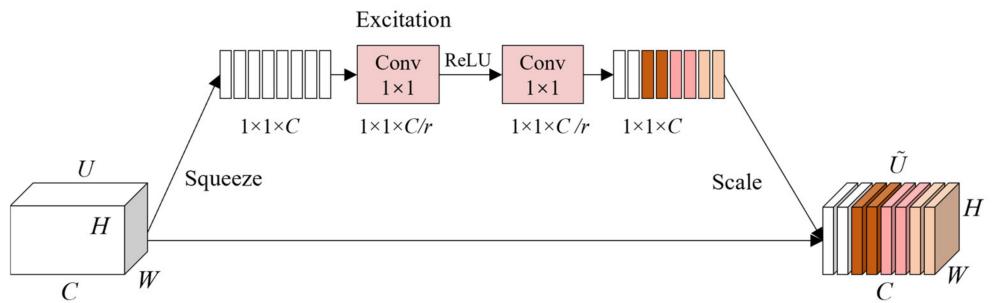


Figure 2. The structure of the channel attention module—SE block.

(1) **Squeeze:** compress the two-dimensional feature map U ($H \times W$) of each channel into a channel descriptor z through Global Average Pooling (GAP), and the calculation formula of the c th element of z is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (4)$$

where $F_{sq}(u_c)$ represents the compression operation performed on the feature map u_c of the c th channel, and H and C represent the height and width of the feature map, respectively.

The Squeeze operation compresses the size of the feature graph from $[H, W, C]$ to $[1, 1, C]$ so that each element in z has a global receptive field for the channel of the feature map U .

(2) **Excitation:** This operation is intended to fully capture channel dependencies. The feature weight vector obtained by the Squeeze operation is activated by two FC layers, and the channel weight is controlled between 0 and 1 by the Sigmoid activation function.

In this part, we've made some improvements to the Excitation section. First, we replace the two fully connected layers with two 1×1 convolution layers to avoid the complicated tensor dimension transformation in the calculation process when using fully connected layers. Secondly, we change the channel dimension reduction between two full connections into channel dimension increase and map the obtained weight vector to a high dimension to further improve the expression ability of feature weight.

(3) **Scale:** Weight the normalized channel weight obtained above to the feature of each channel; that is, multiply the feature weight by the original feature map to obtain the calibrated feature map U .

The structure parameters of the proposed HFECNET-CA can be configured flexibly. The HFETNET-CA is composed of a hybrid feature extraction layer and H temporal feature extraction layers, where the hybrid feature extraction layer contains C feature channels in total. To represent the HFECNET-CA network with different structural parameters, the specific structural parameters are represented by HFECNET-CA(C_1, C_2, C_3, H). C_1, C_2 , and C_3 , respectively, represent the number of channels of the three branches of the hybrid feature extraction section, and H represents the number of layers of the time-domain feature extraction layer. Meanwhile, HFECNET-CA(C_1, C_2, C_3, H) degenerates to HFECNET(C_1, C_2, C_3, H) when no additional channel attention module is used.

3. Datasets and Implementation Details

3.1. Dataset

This paper uses the dataset RML2016.10a generated by Timothy J et al. using the GNU Radio software platform for experimental verification. The RML2016.10a dataset contains 20 signal-to-noise ratios ($-20:2:18$ dB) for a total of 22,000 samples for 11 modulation schemes: WBFM, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, 4PAM, 16QAM, 64QAM, QPSK, and 8PSK. The number of samples of each modulation mode under each SNR condition is 1000, and the data format of each sample is 128×2 , wherein 128 represents the signal length.

3.2. Implementation Details

The experiment randomly divided the data set into a training set, verification set, and test set according to the ratio of 8:1:1. The cross-entropy loss function and ADAM optimizer are used. The initial learning rate is set to 0.001. If the verification loss does not decrease within 10 epochs, it is reduced to 0.1 times the original to improve the training efficiency. The batch size is set to 400. When the verification loss does not decrease within 20 epochs, the training is stopped, and the model with the minimum verification loss is used to predict the signal modulation mode. The hardware environment is Intel(R) Core (TM) i710700k CPU@3.8 GHz, 32GB RAM, NVIDIA GeForce GTX3090 GPU; Software environment: python3.8, PyTorch deep learning framework.

4. Experimental Results and Discussion

4.1. Model Performance Comparison

In the first experiment, to verify the superiority of HFECNET-CA model performance, HFECNET-CA (32, 32, 32, 4) was compared with six SOTA AMR models: IC-AMCNET [12], GRU2 [15], CLDNN [16], MCLDNN [17], MCNet [22], and PET-CGDNN [23]. The performance indexes compared in the experiment include the number of parameters, the test response time of a single sample, the highest recognition accuracy under all SNRs, and the average recognition accuracy under 20 SNRs.

The experimental results are shown in Table 1. The number of parameters of the HFECNET-CA (32, 32, 32, 4) model is 47,979, which is far lower than that of other benchmark models. Its highest recognition accuracy and average recognition accuracy at 20 SNRS are 93.64% and 63.92%, respectively, both of which are 1% higher than the highest performance in the benchmark model. From the perspective of testing time, HFECNET-CA (32, 32, 32, 4) is much lower than other benchmark models but slightly higher than MCNET. Considering its excellent recognition performance, this time cost can be ignored. Overall, HFECNET-CA (32, 32, 32, 4) has the smallest number of parameters but achieves the best recognition performance compared to other baseline models.

Table 1. Model comparison on RML2016.10a datasets.

Model	Parameters	Test Times (ms/Sample)	Highest Accuracy	Average Accuracy
IC-AMCNET	1,264,011	0.24	85.59%	56.83%
GRU2	151,179	0.57	87.86%	58.9%
C LDNN	248,817	0.86	82.90%	55.46%
MCLDNN	406,119	1.18	92.36%	62.24%
MCNet	120,267	0.10	84.54%	56.41%
PET-CGDNN	71,871	0.28	92.54%	61.55%
HFECNET-CA (32, 32, 32, 4)	47,979	0.14	93.64%	63.92%

Figure 3 is the comparison curve of the identification accuracy of all models under 20 SNRs. It can be seen from the figure that HFECNET-CA (32, 32, 32, 4) has the best identification performance under all SNRs. Figure 4 shows the confusion matrix for each model at a signal-to-noise ratio of 10dB. The horizontal axis of the confusion matrix represents the actual modulation class, and the vertical axis represents the modulation class predicted by the model. It can be seen from the figure that the confusion between 16-QAM and 64-QAM, as well as between WBFM and AM-DSB, is the main factor affecting the accuracy of model identification. The confusion between 16-QAM and 64-QAM is because they have overlapping constellation points in the digital domain. Compared with other benchmark models, HFETNET-CA (32, 32, 32, 4) can distinguish 16-QAM and 64-QAM very well. Although the confusion between WBFM and AM-DSB is still not well resolved by HFECNET-CA (32, 32, 32, 4), it is a significant improvement over other SOTA models.

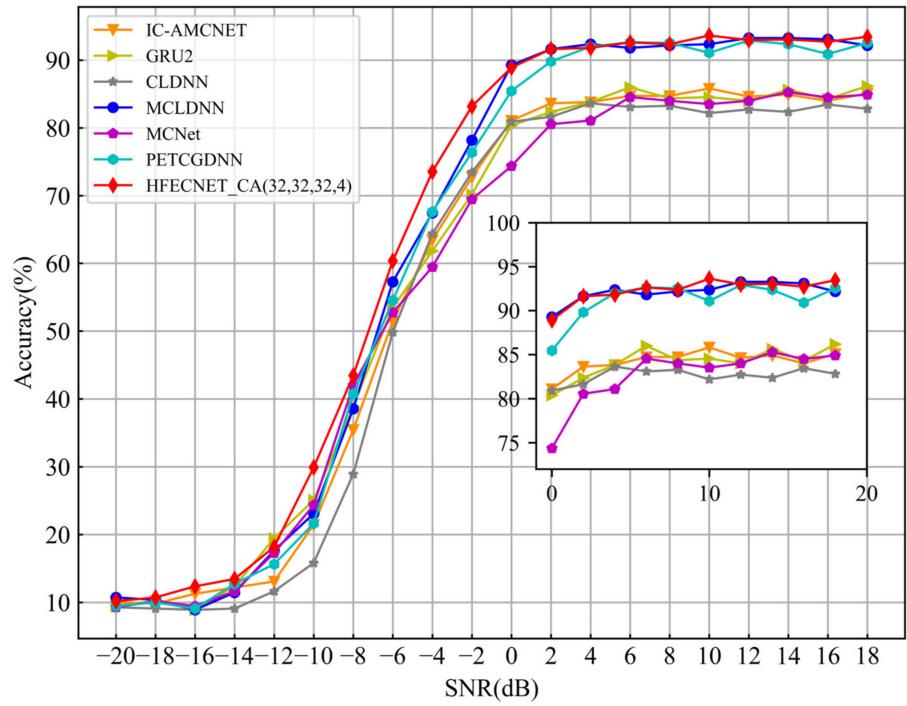


Figure 3. Recognition accuracy comparison on RadioML2016.10a dataset between HFECNET-CA (32, 32, 32, 4) and the other frameworks.

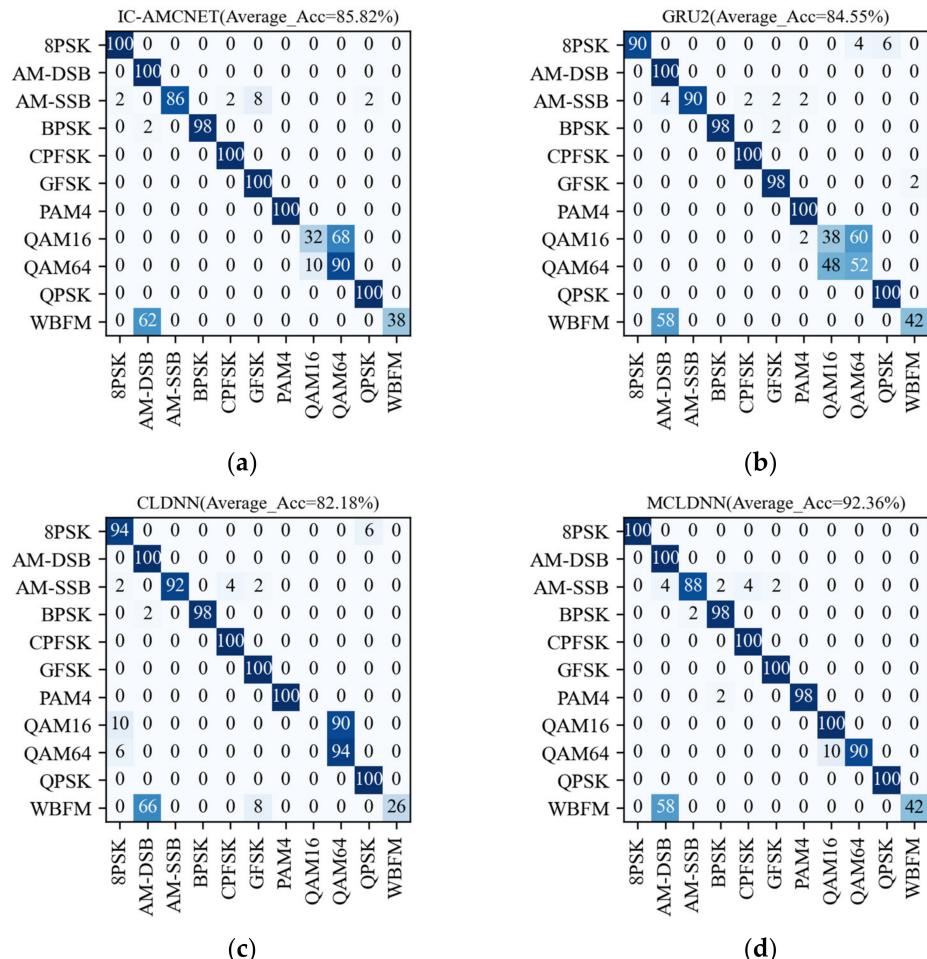


Figure 4. Cont.

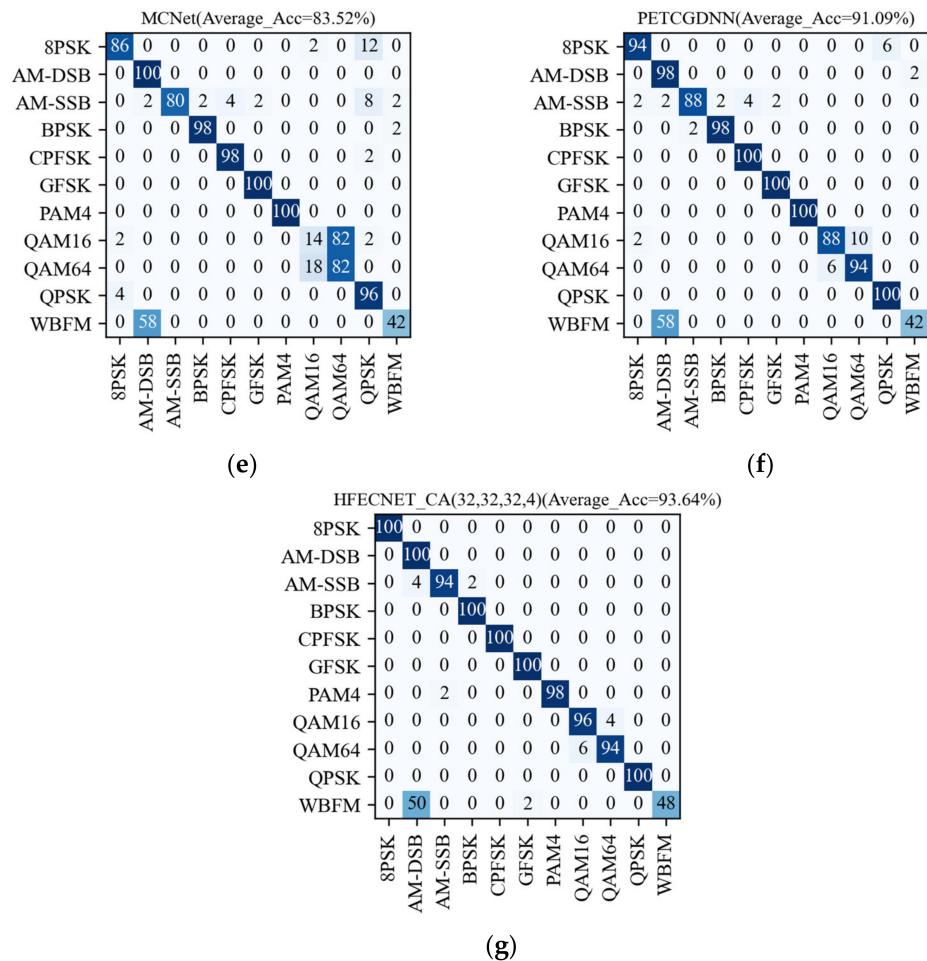


Figure 4. Confusion matrix of (a) IC-AMCNET, (b) GRU2, (c) CLDNN, (d) MCLDNN, (e) MCNet, (f) PET-CGDNN, and (g) HFECNET-CA (32, 32, 32, 4) on RadioML2016.10a dataset at 10 dB SNR.

4.2. HFECNET-CA Structure Effectiveness Analysis

In the second set of experiments, we analyzed the effects of the three convolution branches and the attention module in the hybrid feature extraction module on the performance of the proposed framework. For this purpose, four control frame groups are provided: HFECNET (32, 32, 32, 4) (removes the attention block), HFECNET-CA (0, 32, 32, 4) (removes the first convolution branch), HFECNET-CA (32, 0, 32, 4) (removes the second convolution branch), and HFECNET-CA (32, 32, 0, 4) (removes the third convolution branch).

Table 2 shows the overall performance of each control model on the data, and Table 3 further shows the recognition accuracy of each modulation mode when the signal-to-noise ratio is 10dB. Comparing the specific indicators of HFECNET (32, 32, 32, 4) and HFECNET-CA (32, 32, 32, 4), it can be found that the introduction of the channel attention mechanism is helpful to solve the confusion between QAM16 and QAM64 and can improve the recognition accuracy of WBFM to a certain extent, so the overall recognition performance of the model is improved. At the same time, because the network architecture designed in this paper is small, it seems that the introduction of the channel attention module will have a great influence on the volume of the model, but in general, the volume of the model is small, which does not affect the conclusion that the attention module SENET introduced in this paper is lightweight.

Table 2. Performance comparison between HFECNET-CA (32, 32, 32, 4) and its varieties.

Model	Parameters	Highest Accuracy	Average Accuracy
HFECNET (32, 32, 32, 4)	18,571	92.18%	61.81%
HFECNET-CA (32, 32, 32, 4)	47,979	93.64%	63.92%
HFECNET-CA (0, 32, 32, 4)	42,507	92.5%	61.2%
HFECNET-CA (32, 0, 32, 4)	40,491	92.5%	61.7%
HFECNET-CA (32, 32, 0, 4)	42,635	92.72%	62.44%

Table 3. Recognition accuracy on RadioML2016.10a at 10 dB SNR.

	HFECNET (32, 32, 32, 4)	HFECNET-CA (32, 32, 32, 4)	HFECNET-CA (0, 32, 32, 4)	HFECNET-CA (32, 0, 32, 4)	HFECNET-CA (32, 32, 0, 4)
8PSK	0.94	1	0.98	0.94	1
AM-DSB	1	1	0.82	1	1
AM-SSB	0.9	0.94	0.88	0.88	0.88
BPSK	0.98	1	0.98	0.98	0.98
CPFSK	1	1	1	1	1
GFSK	1	1	1	1	1
PAM4	0.98	0.98	0.98	1	1
QAM16	0.9	0.96	0.98	0.88	0.88
QAM64	0.8	0.94	0.92	0.84	0.86
QPSK	0.98	1	1	1	1
WBFM	0.44	0.48	0.56	0.42	0.42

Comparing HFECNET-CA (0, 32, 32, 4), HFECNET-CA (32, 0, 32, 4), HFECNET-CA (32, 32, 0, 4), and HFECNET-CA (0, 32, 32, 4), the results show that the introduction of each convolution branch improves the overall recognition performance of the model (about 1~2%). The first branch (extraction of general features) is helpful in improving the recognition rate of AM-DSB and AM-SSB, while the second and third convolution branches (extraction of IQ-related features and deep IQ-related features) can solve the confusion problem between QAM16 and QAM64. This also fully demonstrates that the multi-view hybrid feature extraction method proposed in this paper is reasonable and effective.

5. Conclusions

In this paper, we proposed a lightweight and efficient ARM model based on CNN, named HFECNET-CA. To improve the information utilization ratio of the original IQ signal, this model adopts three forms of convolutional branches to extract features from different perspectives and performs feature fusion and deep time-domain feature extraction, fully exploring the distinguishable features in the signal. Furthermore, a channel attention mechanism was introduced for feature correction, which enables the model to have better feature expression capabilities with less parameter overhead. The experimental results on RML2016.10a indicate that HFECNET-CA achieves higher recognition accuracy with a smaller network architecture and can better distinguish between QAM16 and QAM64 compared to other benchmark models.

Author Contributions: Conceptualization, Z.M. and Y.F.; methodology, S.F.; software, Z.M. and G.L.; validation, Z.M. and H.H.; formal analysis, H.H. and G.L.; resources, S.F.; data curation, Z.M. and G.L.; writing—original draft preparation, Z.M.; writing—review and editing, Z.M. and G.L.; supervision, H.H. and Y.F.; funding acquisition, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Key Basic Research Projects of the Basic Strengthening Program, grant number 2020-JCJQ-ZD-071.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to express their appreciation to the editors for their rigorous and efficient work, and the reviewers for their helpful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. O’Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179. [[CrossRef](#)]
2. Meng, F.; Chen, P.; Wu, L.; Wang, X. Automatic Modulation Classification: A Deep Learning Enabled Approach. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10760–10772. [[CrossRef](#)]
3. Dulek, B. Online Hybrid Likelihood Based Modulation Classification Using Multiple Sensors. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 4984–5000. [[CrossRef](#)]
4. Chang, D.C.; Shih, P.K. Cumulants-based modulation classification technique in multipath fading channels. *IET Commun.* **2015**, *9*, 828–835. [[CrossRef](#)]
5. Huang, S.; Yao, Y.; Wei, Z.; Feng, Z.; Zhang, P. Automatic Modulation Classification of Overlapped Sources Using Multiple Cumulants. *IEEE Trans. Veh. Technol.* **2017**, *66*, 6089–6101. [[CrossRef](#)]
6. Kim, B.; Kim, J.; Chae, H.; Yoon, D.; Choi, J.W. Deep neural network-based automatic modulation classification technique. In Proceedings of the 2016 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 19–21 October 2016; pp. 579–582.
7. Shuli, D.; Zhipeng, L.; Linfeng, Z. A modulation recognition algorithm based on cyclic spectrum and SVM classification. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 2123–2127.
8. Furtado, R.S.; Torres, Y.P.; Silva, M.O.; Colares, G.S.; Pereira, A.M.C.; Amoedo, D.A.; Valadao, M.D.M.; Carvalho, C.B.; da Costa, A.L.A.; Junior, W.S.S. Automatic Modulation Classification in Real Tx/Rx Environment using Machine Learning and SDR. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 10–12 January 2021; pp. 1–4.
9. Ya, T.; Lin, Y.; Wang, H. Modulation Recognition of Digital Signal Based on Deep Auto-Ancoder Network. In Proceedings of the 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Prague, Czech Republic, 25–29 July 2017; pp. 256–260.
10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
11. O’Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. In Proceedings of the Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, 2–5 September 2016; pp. 213–226.
12. Hermawan, A.P.; Ginanjar, R.R.; Kim, D.-S.; Lee, J.-M. CNN-Based Automatic Modulation Classification for Beyond 5G Communications. *IEEE Commun. Lett.* **2020**, *24*, 1038–1041. [[CrossRef](#)]
13. Rajendran, S.; Meert, W.; Giustiniano, D.; Lenders, V.; Pollin, S. Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors. *IEEE Trans. Cogn. Commun. Netw.* **2018**, *4*, 433–445. [[CrossRef](#)]
14. Wang, Y.; Liu, M.; Yang, J.; Gui, G. Data-Driven Deep Learning for Automatic Modulation Recognition in Cognitive Radios. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4074–4077. [[CrossRef](#)]
15. Hong, D.; Zhang, Z.; Xu, X. Automatic modulation classification using recurrent neural networks. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 695–700.
16. West, N.E.; O’Shea, T.J. Deep Architectures for Modulation Recognition. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; pp. 1–6.
17. Xu, J.; Luo, C.; Parr, G.; Luo, Y. A Spatiotemporal Multi-Channel Learning Framework for Automatic Modulation Recognition. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1629–1632. [[CrossRef](#)]
18. Zhang, H.; Huang, M.; Yang, J.; Sun, W. A Data Preprocessing Method for Automatic Modulation Classification Based on CNN. *IEEE Commun. Lett.* **2021**, *25*, 1206–1210. [[CrossRef](#)]
19. Wang, T.; Hou, Y.; Zhang, H.; Guo, Z. Deep Learning Based Modulation Recognition With Multi-Cue Fusion. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1757–1760. [[CrossRef](#)]
20. Kim, S.-H.; Moon, C.-B.; Kim, J.-W.; Kim, D.-S. A Hybrid Deep Learning Model for Automatic Modulation Classification. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 313–317. [[CrossRef](#)]
21. Wang, Y.; Guo, L.; Zhao, Y.; Yang, J.; Adebisi, B.; Gacanin, H.; Gui, G. Distributed learning for automatic modulation classification in edge devices. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 2177–2181. [[CrossRef](#)]
22. Huynh-The, T.; Hua, C.-H.; Pham, Q.-V.; Kim, D.-S. MCNet: An Efficient CNN Architecture for Robust Automatic Modulation Classification. *IEEE Commun. Lett.* **2020**, *24*, 811–815. [[CrossRef](#)]
23. Zhang, F.; Luo, C.; Xu, J.; Luo, Y. An Efficient Deep Learning Model for Automatic Modulation Recognition Based on Parameter Estimation and Transformation. *IEEE Commun. Lett.* **2021**, *25*, 3287–3290. [[CrossRef](#)]
24. Shi, F.; Yue, C.; Han, C. A lightweight and efficient neural network for modulation recognition. *Digit. Signal Process.* **2022**, *123*, 103444. [[CrossRef](#)]

25. Jia, F.; Yang, Y.; Zhang, J.; Yang, Y. A hybrid attention mechanism for blind automatic modulation classification. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4503. [[CrossRef](#)]
26. Wang, Z.; Sun, D.; Gong, K.; Wang, W.; Sun, P. A Lightweight CNN Architecture for Automatic Modulation Classification. *Electronics* **2021**, *10*, 2679. [[CrossRef](#)]
27. Xiao, C.; Yang, S.; Feng, Z. Complex-Valued Depthwise Separable Convolutional Neural Network for Automatic Modulation Classification. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2522310. [[CrossRef](#)]
28. O’Shea, T.J.; West, N. Radio machine learning dataset generation with gnu radio. In Proceedings of the GNU Radio Conference, 2016; Available online: <https://pubs.gnuradio.org/index.php/grcon/article/view/11> (accessed on 11 July 2023).
29. Zhang, F.; Luo, C.; Xu, J.; Luo, Y.; Zheng, F.-C. Deep learning based automatic modulation recognition: Models, datasets, and challenges. *Digit. Signal Process.* **2022**, *129*, 103650. [[CrossRef](#)]
30. Mnih, V.; Heess, N.; Graves, A. Recurrent models of visual attention. In Proceedings of the Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014.
31. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.