Received April 27, 2021, accepted April 30, 2021, date of publication May 11, 2021, date of current version June 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079131

A New Modulation Recognition Method Based on Flying Fish Swarm Algorithm

JINGPENG GAO[1,2], XU WANG[2], RUOWU WU[1], AND XIONG XU[1]

[1]State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang 471003, China
[2]College of Information and Communication, Harbin Engineering University, Harbin 150001, China

Corresponding author: Jingpeng Gao (gaojingpeng@hrbeu.edu.cn)

This work was supported by the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE) through Researchers under Grant CEMEE2021K0103B.

ABSTRACT The modulation recognition method based on deep learning plays a significant role in the intelligent communication system. To further improve the recognition rate, especially in the case of small samples with a low signal-to-noise ratio, this paper proposes a new modulation recognition method based on flying fish swarm algorithm. First, Short-Time Fourier Transform, Choi-Williams Distribution, and Cyclic Spectrum are combined to complete multi-channel signal processing. Second, AlexNet, VGGNet, GoogLeNet, and ResNet are transferred to realize feature extraction. Third, the support vector machine classifies the modulations after dimension reduction and feature fusion. Finally, the flying fish swarm is proposed to optimize the signal processing methods, the types of networks, the layers of networks, the dimensions of features, and the parameters of the support vector machine. The method can accurately recognize BPSK, QPSK, OQPSK, 8PSK, 4ASK, QAM16, QAM32, and QAM64. The simulation results show that the average recognition rate of modulation is 94.5% at SNR of 0 dB and 84.7% at SNR of −4 dB. Besides, the proposed modulation recognition method possesses good robustness under low SNR conditions.

INDEX TERMS Modulation recognition, deep learning, flying fish swarm algorithm, multi-channel signal processing, feature extraction.

I. INTRODUCTION

Modulation recognition of communication signals has played an important role in new intelligent communication systems, which can help to demodulate and decode signals. Generally speaking, modulation recognition can be divided into interpulse modulation recognition and intrapulse modulation recognition. Interpulse modulation recognition is the pulse repetition interval modulation recognition [1], which is used for signal sorting [2]. Intrapulse modulation recognition can be further divided into intrapulse unintentional modulation recognition and intrapulse intentional modulation recognition [3]. The former is the classification of individual radiation sources [4].The latter is the classification of signal modulation mode [5]. This paper focuses on intrapulse intentional modulation recognition, which is referred to as modulation classification in the following contents.

Many scholars have devoted themselves to studying various methods of modulation recognition of communication signals [6]. In general, the recognition process can be divided into three parts: signal processing, feature extraction, and classification [7]. In the field of non-cooperative communication, the working environment is often under a low signal-to-noise ratio (SNR). Communication signals have the characteristics of diversity and fewer samples, which leads to the low recognition rates of existing modulation recognition methods.

The purpose of signal processing is to transform a signal into an image, which can be extracted features by the convolutional neural network (CNN) [8]. With the increasing complexity of the electromagnetic environment, spectrum aliasing will occur when the signal is analyzed from the traditional Fourier transform [9]. Short-Time Fourier Transform (STFT) is used to combine the time information of the signal with the frequency information, which has low computational complexity and no cross-terms[10]. Choi-Williams Distribution (CWD) has a higher resolution than STFT. CWD is still affected by the cross terms, although it can suppress the cross-terms to some extent [11]. Cyclic Spectrum(CS) has a better performance for periodically stationary modulated

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

signals [12]. CS carries out a three-dimensional spectrum analysis of signals.

Feature extraction based on deep learning is the key part of the modulation recognition method. Many neural networks have been proposed to achieve better performance [13]. In 2012, Hinton uses AlexNet to perform gesture recognition in the ILSVRC competition and won the championship. This network developed the idea of the LeNet network and applied the CNN network to a very deep and wide network [14]. The VGGNet is developed by Oxford University Computer Vision Combination and Google DeepMind researchers. It explores the relationship between the depth and performance of CNN [15]. GoogLeNet is proposed to average the accuracy of bilingual manual translation on some test sets [16]. ResNet is proposed by four scholars from Microsoft Research, which solves the problem of gradient disappearance or gradient explosion caused by the deepening of network depth [1 7]. O'Shea presents the concept of radio transformer networks as a means to incorporate expert domain knowledge in the machine learning model, which achieves competitive accuracy with respect to traditional schemes relying on expert features [18]. However, it is very difficult for the non-cooperative communication field to build a complete database like other databases. Transfer learning exploits existing knowledge to solve different and related domain problems [19]. Hashemian has solved the problem of weak generalization ability due to too little customer data during training by using transfer learning [20]. Due to the small samples of communication signals, it is sensible to apply transfer learning to feature extraction. AlexNet is used to recognize the modulation of eight communication signals. This study achieves a recognition rate of 78.1% at 0dB [21]. Peng uses GoogLeNet to identify the modulation mode of the communication signal based on constellation transformation, which achieves 77.4% at 0dB in the case of 1000 samples [22]. However, the above two kinds of studies related to transfer learning only choose a single signal processing method, which cannot fully express the essential information of the signals. What's more, the above studies do not consider the layers of neural networks and the parameters of the classifiers.

The single-channel feature extraction method can be composed of the signal processing and transferred network model. The feature information obtained by the single-channel method is limited, which is not conducive to subsequent classification. Hence, this paper uses a multi-channel signal processing method. More abundant feature information can be obtained by the fusion of multi-channel features. Although fusing features of different scales is an important means to improve the classification performance, the direct fusion of a large number of extracted features will lead to information redundancy. Feature dimension reduction can effectively resolve this conflict [23]. Yang uses Principal Component Analysis (PCA) to transform the original input linearity into new uncorrelated features. The results show that the PCA dimension reduction effect is better than that of feature classification without dimension reduction [24]. However, the selection of the number of features before fusion will affect the recognition accuracy. To achieve a better recognition rate, not only the dimension of features but also the selection relationship between the signal processing methods, neural network, and corresponding network layer should be further considered.

Classification is the other key step of the modulation recognition method based on deep learning [25]. O'Shea conducts an in-depth study on the performance of deep learning based radio signal classification for radio communications signals [26]. As for communication signals with small samples, it not able to train the network adequately. Support Vector Machine (SVM) is a classification technology based on a risk minimization structure, which has a high generalization ability [27]. SVM can effectively solve the problem of constructing high dimensional data model under the condition of limited samples. At the same time, it can converge to the global optimum [28]. However, the selection of SVM parameters has a significant impact on classification results. Jain uses an artificial bee colony algorithm to optimize the parameters of SVM, which can find the optimal solution easily. However, the initial search speed of this algorithm is very slow [29]. Liang proposes a differential evolution algorithm to optimize the parameters of SVM, but the algorithm converged to the extreme value prematurely [30]. To optimize the classifier parameters quickly and effectively to further improve the recognition, an effective optimization method needs to be proposed.

Given the above methods, the problems can be drawn as follows, the first one is that the appropriate signal processing methods, neural network, and corresponding network layer need to be selected for feature extraction. The second one is that the appropriate dimensions of features need to be selected for classification before feature fusion. The last one is that the key parameters of SVM need to be optimized to further improve the recognition rate.

This study regards the above problem as a feasible solution space optimization process. The intelligent algorithm is a popular parameter optimization method in recent years [31]. It has good optimization ability for huge solution space. Artificial fish swarm algorithm is a swarm intelligence algorithm with better global optimization performance [32]. But the accuracy is low when the solution space is large [33]. The cuckoo algorithm combines the principle of Levy flight. Levy flight has the advantage of both long and short paths, which can increase the diversity of solutions [34]. However, the convergence speed of this algorithm is very slow. Because of the above two algorithms, we propose a new Flying Fish Swarm (FFS) algorithm iteratively optimize the objective function and realize the optimal adaptation of parameters, to effectively recognize the communication modulation mode.

In this paper, a new modulation recognition method based on the Flying Fish Swarm algorithm is proposed. The modulations that the proposed algorithm can recognize include

Binary Phase-Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), Offset Quadrature Phase-Shift Keying (OQPSK), 8 Phase-Shift Keying (8PSK), 4 Amplitude-Shift Keying (4ASK), Quadrature Amplitude Modulation 16 (QAM16), Quadrature Amplitude Modulation 32 (QAM32) and Quadrature Amplitude Modulation 64 (QAM64). We use STFT, CS, and CWD multi-channel signal processing methods to analyze signals. We also build a Convolutional Neural Network Set (CNNS) based on transfer learning, including ResNet-50, VGGNet-19, GoogLeNet, and AlexNet. The selective relationship among the signal processing methods, the transferred network models in CNNS, and the appropriate number of network layers constitutes part of the feasible solution space. Moreover, the dimension of features after dimension reduction and the parameters of SVM will have a direct impact on the recognition rate, which constitutes the rest of the feasible solution space. Then, this paper uses the FFS algorithm to select the above parameters quickly and reasonably. Compared with the previous work, the average recognition rate of the proposed method is 94.5% at SNR of 0 dB and 84.7% at SNR of −4 dB. Monte Carlo simulation results show that the proposed method is suitable for different communication signal modulation modes under a wide range of SNR.

The main contributions of this paper are as follows:

1) A new modulation recognition method based on the FFS algorithm is proposed. This method can solve the problem of low recognition rate under low SNR and small sample size.
2) CNNS is built with the idea of transfer learning, which includes four kinds of neural networks and different network layers. Besides, the selection of features after PCA is designed to retain more useful information of the signal.
3) FFS algorithm is proposed. Compared with the traditional algorithm, the algorithm can find the optimal solution of the combination of CNNS, the dimensions of features through PCA, and the parameters of SVM more accurately.

The structure of this paper is organized as follows, section II establishes the model of communication signal and the system structure. Section III introduces modulation recognition based on deep learning. Section IV discusses FFS algorithm in detail. The simulation results are shown in section V. Finally, conclusions are drawn in section VI.

II. SIGNAL MODEL AND SYSTEM STRUCTURE

A. SIGNAL MODEL

Through RF Front End Modules and acquisition & tracking in Baseband Digital Signal Processing Module in the receiver, we get the digital processed communication signals, on this basis, we carry out modulation classification.

The algorithm is proposed to realize the recognition of communication signals. The modulated signal is jammed by additive white Gaussian noise, which is expressed as follows,

$$s(t) = x(t) + n(t) = A\exp(j\varphi(t)) + n(t), \quad (1)$$

where $x(t)$ represents the communication signal and $n(t)$ represents the white Gaussian noise, A represents the amplitude of the signal, $\varphi(t)$ represents the instantaneous phase of the communication signal. Eight kinds of common modulated signals in communication are selected in this paper, including BPSK, QPSK, 8PSK, OQPSK, 4ASK, QAM16, QAM32, and QAM64 [35].

TABLE 1. The instantaneous phase of communication signals with different modulations.

Modulation	Phase/Amplitude
BPSK	$\varphi(t) = 2\pi f_c t + j\pi, j = 0, 1, A{=}1$
QPSK	$\varphi(t) = 2\pi f_c t + j\pi, j = 1/4, 3/4, 5/4, 7/4$
OQPSK	$\varphi(t) = 2\pi f_c t + j\pi - \pi/4,$ $j = 1/4, 3/4, 5/4, 7/4, A{=}1$
8PSK	$\varphi(t) = 2\pi f_c t + (2j-1)\pi,$ $j = 1, 2, \cdots, 8, A{=}1$
4ASK	$\varphi(t) = 2\pi f_c t, A{=}0, 1, 2, 3$
QAM16	$\varphi(t) = 2\pi f_c t, A{=}A_I + jA_Q,$ $A_I = \pm1, \pm3, A_Q{=}\pm1, \pm3$
QAM32	$\varphi(t) = 2\pi f_c t, A{=}A_I + jA_Q,$ $A_I = \pm1, \pm3, \pm5, A_Q{=}\pm1, \pm3, \pm5$
QAM64	$\varphi(t) = 2\pi f_c t, A{=}A_I + jA_Q, A_I = \pm1,$ $\pm3, \pm5, \pm7, A_Q{=}\pm1, \pm3, \pm5, \pm7$

The instantaneous phases of different signal modulation modes are given in Table 1, where f_c represents the carrier frequency of the signal, A_I represents the in-phase component, A_Q represents the quadrature component.

B. SYSTEM STRUCTURE

The signal recognition system includes five parts, which are signal processing, feature extraction, feature fusion and dimension reduction, classification, and parameter optimization based on the FFS algorithm. First, CWD, STFT, and CS are used to process the modulated signals in communication. Second, ResNet, VGGNet, GoogLeNet, and AlexNet are used to construct CNNS to extract the features of the signals based on the idea of transfer learning. Moreover, we use PCA to reduce the dimension of features and remove unnecessary information based on the overall characterization of signal features. Finally, we fuse the features and send them to SVM to classify. We propose a new FFS algorithm to optimize the types of networks, the layers of networks, the dimensions of features, the parameters of the support vector machine. Fig.1 shows the communication signal modulation recognition system.

III. MODULATION RECOGNITION BASED ON DEEP LEARNING

A. SIGNAL PROCESSING

Signal processing plays an important role in the modulation recognition of signals. Multi-channel signal processing methods can provide strong support for subsequent feature extraction. It can more comprehensively characterize the characteristics of signals. CS, STFT, and CWD process signals from different aspects according to their different characteristics, which are commonly used in recent years. Among them, CS analysis can better characterize the

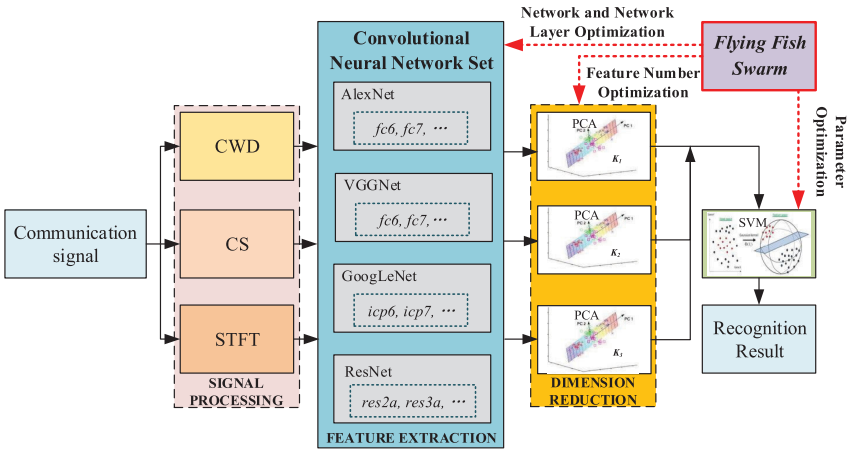


FIGURE 1. The figure shows the communication signal modulation recognition system.

cyclostationarity characteristics of signals, and STFT can reflect the time and frequency characteristics of signals. CWD transform is an improved type of Wigner distribution, which can better reflect the energy spectral density distribution characteristics of signals.

1) CYCLIC SPECTRUM

Modulation recognition of communication signals requires a series of signal processing such as sampling, quantization, and coding. The above processing has periodic characteristics. CS theory is based on cyclostationary characteristics. It can get good results by using a cyclic spectrum to process communication signals. When the modulated signal $s(t)$ is received, the signal processing of cyclic spectrum analysis is carried out,

$$S_s^{\alpha}(f) = \int_{-\infty}^{\infty} R_s^{\alpha}(\tau) \exp(-j2\pi f \tau) d\tau, \tag{2}$$

where α is the cyclic frequency. $R_s^{\alpha}(\tau)$ is,

$$R_s^{\alpha}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t+\frac{\tau}{2}) \times s^*(t-\frac{\tau}{2}) \exp(-j2\pi\alpha t) dt, \tag{3}$$

where $R_s^{\alpha}(\tau)$ is the cyclic autocorrelation of signal $s(t)$, T is the length of time series. Since the signal is sampled to obtain discrete values within the time interval Δt, the formula (2) is discretized. Cyclic spectral density is an estimate, and its value can be obtained by the time-smoothing period gram,

$$S_s^{\alpha}(f) \approx S_{sT_w}^{\alpha}(t,f)_{\Delta t} = \frac{1}{T_w \Delta t} \int_{t-(\Delta t/2)}^{t+(\Delta t/2)} \times X_{T_w}(u, f+\frac{\alpha}{2}) X_{T_w}^*(u, f-\frac{\alpha}{2}) du, \tag{4}$$

where T_w is the length of the short FFT window, and $X_{T_w}(t,f) = \int_{t-(T_w/2)}^{t+(T_w/2)} s(u)\exp(-j2\pi f u)du$. Besides, some estimates of (f_0, α_0) represent a very small region in the dual-frequency plane, which requires a sufficient number of estimates to fully represent the cyclic spectrum. Due to the large amount of computation to obtain the estimates, efficient algorithms are needed. Time smoothing FFT Accumulation (FAM) divides the dual-frequency plane into smaller regions, called channel-to-channel regions [36]. The fast Fourier transform is used to calculate the estimated value of a block, which can effectively reduce the amount of calculation needed to estimate the cyclic spectrum. In this paper, the FAM algorithm is used to analyze the cyclic spectrum of signals. The cyclic spectral density function for estimating time under discrete condition is,

$$S_{X_{N'}}^{\gamma}(\gamma, k) = \frac{1}{N} \sum_{n=0}^{N-1} [\frac{1}{N'} X_{N'}(n, k+\frac{\gamma}{2}) X_{N'}^*(n, k-\frac{\gamma}{2})], \tag{5}$$

where $X_{N'}(n,k)$ is the discrete Fourier transform of $s(n)$, k, γ are the discrete values of f and α, respectively. Parameter N is the total number of discrete samples in observation time, N' is the number of points in FFT.

After calculating the communication signal by CS analysis, eight different signal cycle spectra are obtained. Fig.2 shows the cyclic spectra of the signals. The coordinate axis respectively represents frequency, cyclic frequency, and amplitude.

The three-dimensional spectrum of CS shows that the non-cyclostationary signal can be distinguished well based on the CS theory. To use advanced neural network technology to extract features, it is necessary to send two-dimensional images to the network. So the three-dimensional images are projected as two-dimensional images.

2) SHORT TIME FOURIER TRANSFORM

The process of signal analysis mainly reflects the characters of signal changing with time. The time-domain description of signals mainly includes signal energy, the average value of time function, waveform center, duration, time width, and

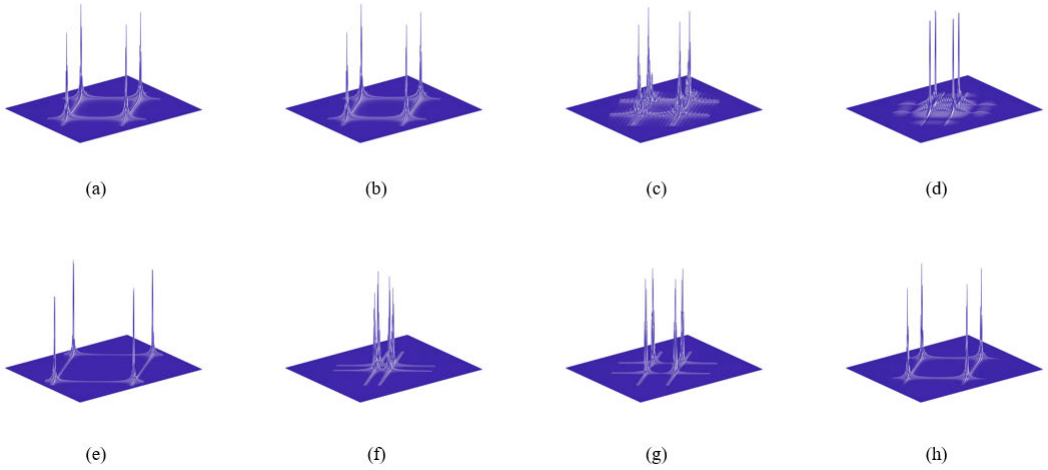


FIGURE 2. The CS transform images of signals are shown in this figure, (a) BPSK, (b) QPSK, (c) OQPSK, (d) 8PSK, (e) 4ASK, (f) QAM16, (g) QAM32, (h) QAM64.

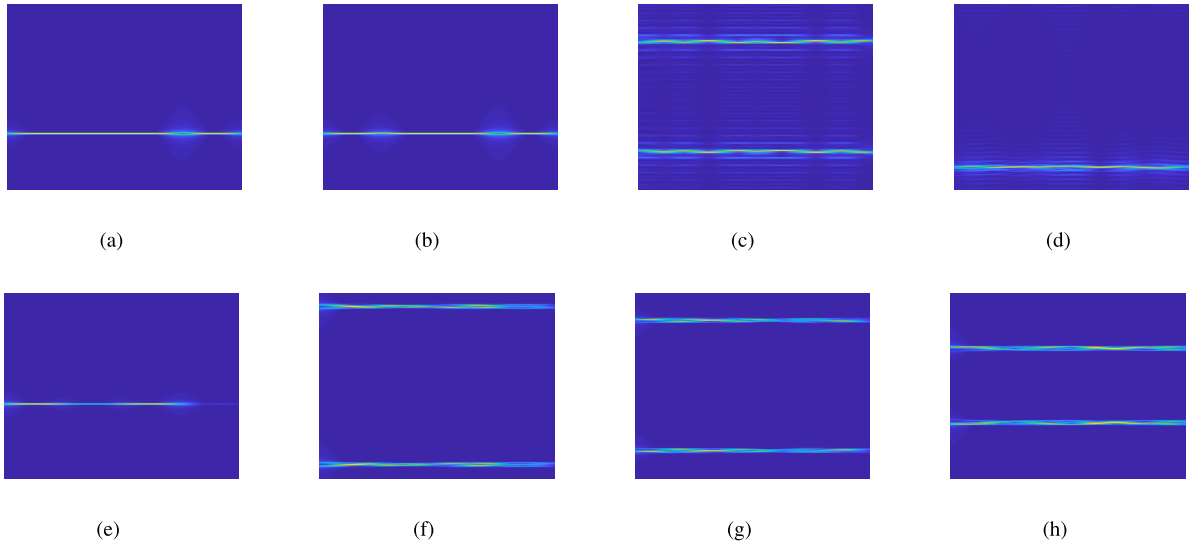


FIGURE 3. The STFT images of signals are shown in this figure, (a) BPSK, (b) QPSK, (c) OQPSK, (d) 8PSK, (e) 4ASK, (f) QAM16, (g) QAM32, (h) QAM64.

other parameters. If the time domain parameters of the two signals are identical, it cannot be said that the signals are identical. In practical application, not only do the parameters change with time but also some parameters change with frequency and phase. So when analyzing the signal, we need to use the joint time-frequency analysis method.

The traditional Fourier transform compares the signal with the sinusoidal basis function, but the sinusoidal basis function is extended in the whole time domain, which is not conducive to accurately reflect the exact location of a frequency. STFT solves the contradiction between localization in the time domain and frequency domain. It uses window function $g(\tau)$ to intercept signal $s(\tau)$. It gets the Fourier transform of the intercepted local signal at time t. On the one hand, $g(\tau)$ is a window function, so it can be limited support in the time domain. On the other hand, the spectrum of $\exp(j\Omega t)$ is line spectrum, so the basis function $g(\tau - t)\exp(j\Omega t)$ of STFT should be limited support in the time domain and frequency domain. In this way, STFT can locate the signal $s(t)$. When the modulation signal $s(t)$ is received, STFT is performed,

$$STFT_s(t, \Omega) = \int_{-\infty}^{+\infty} s(\tau)g^*(\tau - t)\exp(-j\Omega t)d\tau. \quad (6)$$

The window function should be a symmetrical real function. In this paper, the window function should be constant. With the increase of the number of window function points, the main lobe of the amplitude-frequency characteristics will be narrowed, and the leakage of the spectrum will be reduced. Fig.3 shows the time-frequency image of STFT. In the figure, the x-axis represents time and the y-axis represents frequency.

3) CHOI-WILLIAMS DISTRIBUTION

STFT belongs to linear time-frequency transform, which only reflects the distribution characteristics of the signal in the

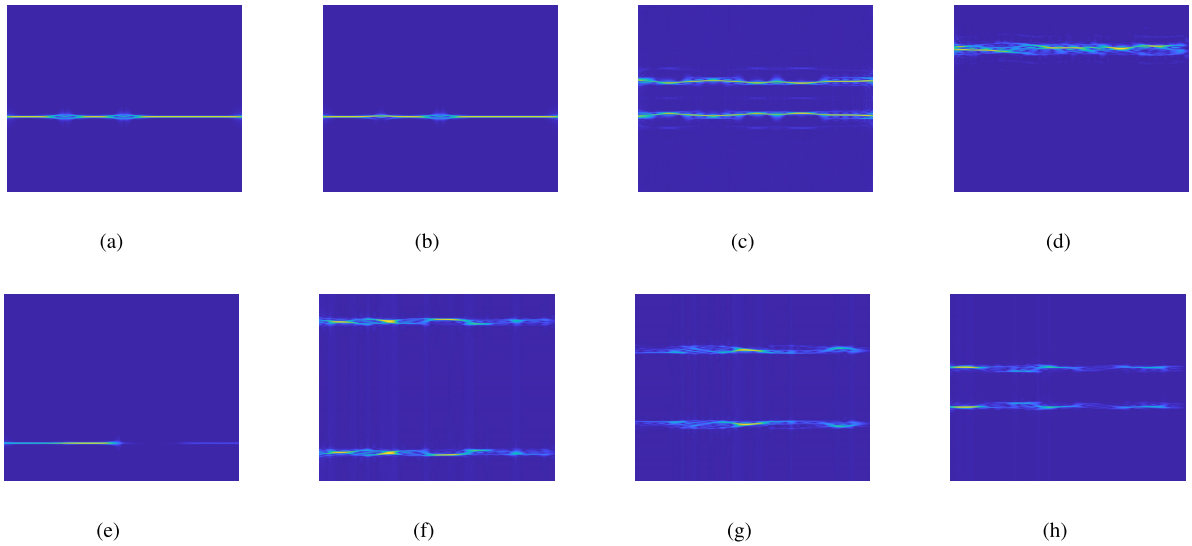


FIGURE 4. The CWD transform images of signals are shown in this figure, (a) BPSK, (b) QPSK, (c) OQPSK, (d) 8PSK, (e) 4ASK, (f) QAM16, (g) QAM32, (h) QAM64.

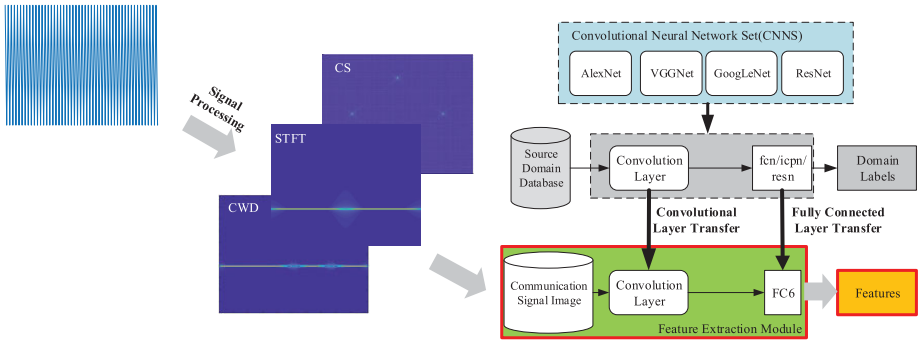


FIGURE 5. The figure shows the feature extraction based on transfer learning.

time and frequency domain. However, in many practical applications, the distribution of energy spectral density is needed. Quadratic time-frequency distribution representation can realize the distribution of energy spectral density of signals. Therefore, this paper uses CWD to process the signal. Assuming that the modulation signal is $s(t)$, CWD time-frequency conversion is performed on the signal,

$$C_s(t,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(j2\pi\zeta(u-t)) \times \exp[(\pi\zeta\tau)^2/2\sigma]s(u+\tau/2) \times s^*(u-\tau/2)\exp(-j\omega\tau)d\zeta du d\tau, \quad (7)$$

where ζ is frequency shift, τ is time shift, u is integral variable. Fig.4 shows CWD transform image of the modulated signal, where the x-axis represents time and the y-axis represents frequency.

B. FEATURE EXTRACTION

The signal in the field of non-cooperative communication is not easy to acquire, and the sample size is small. Therefore, these samples cannot be trained by themselves using a neural network. Transfer learning is a method to recognize and apply knowledge and skills learned in previous domains to novel tasks. In the task of communication signal recognition, we use STFT, CWD, and CS to get two-dimensional signal images. The feature extraction method of these images is similar to those by ResNet-50, VGGNet-19, GoogLeNet, and AlexNet in the ImageNet competition. Therefore, CNNS is proposed with these four pre-trained neural networks to extract the features of modulated signals in communication. CNNS is a solution space containing neural networks and corresponding network layers. It exists in this paper relying on intelligent algorithms. As shown in formula (8), it is a row vector expression according to the corresponding coding rules.

In the process of network transfer, each network has a different generalization ability for communication signals. Moreover, different neural network layers in a network have different representations of images when extracting features. Therefore, when transferring the neural network, we need to optimize the combination of network types and layers. Fig.5 shows CNNS feature extraction based on

transfer learning. Then the network types and corresponding neural network layers are regarded as feasible solutions for feature extraction, so the feasible solution space **NET** is,

$$\mathbf{NET} = [net_1, net_2, net_3], \tag{8}$$

where $net_i (i = 1, 2, 3)$ represent the different feature extraction methods after three time-frequency transforms CS, STFT and CWD, which i represent i-th time-frequency transform.The above neural network layers are encoded and performed decimal encoding in the order of AlexNet, VGGNet, GoogLeNet, and ResNet. Each net_i has a ranging from 1 to the sum of the network layer.

The signal $s(t)$ after the i-th time-frequency transform is,

$$\mathbf{s}_i(t) = \left\{s_{i1}(t), s_{i2}(t), \ldots, s_{ij}(t)\right\} \quad (j = 1, \ldots, J), \tag{9}$$

where J is sample size. Initialize the feasible solution space **NET**. Then select the corresponding neural network layer of CNNS according to the encoding in **NET**. And we send the signal $\mathbf{s}_i(t)$ which is represents the signal of the i-th time-frequency transform to the neural network for feature extraction to get the corresponding extracted features $\mathbf{A}_{J\times M_i}$, (10), as shown at the bottom of the page, where $fea_{jm}(\mathbf{NET}[net_i])$ represents the m-th feature of the j-th sample obtained by feature extraction from the i-th ($i = 1, 2, 3$) time-frequency transform. M_i represents the number of extracted features.

C. DATA DIMENSION REDUCTION AND FEATURE FUSION

From the above research, it can be seen that communication signals can acquire the features of signals from different aspects after multi-transform. After feature extraction, to preserve the signal features of different aspects before classification, the features are fused. However, direct fusion can lead to redundancy of features and a waste of computing resources. If the extracted features are directly sent to SVM for classification, good recognition results cannot be obtained. Because the features are very complex [37]. Dimension reduction is an application of unsupervised learning. It cannot only remove noise and reduce over-fitting but also make SVM easier to obtain valuable information [38]. To solve the problem of dimension disaster caused by too many features, we reduce the dimension of extracted features.

PCA is a statistical analysis method that transforms the original multiple variables into a few new comprehensive variables through linear transform [39]. These new variables (also known as principal components) are not correlated with each other [40]. It can effectively express the information of the original variables without losing or minimizing the information of the original variables, so it can eliminate the interaction between the original data components.

The process of PCA is as follows. De-mean each of the columns in the sample matrix $\mathbf{Y}'$, then calculate covariance matrix,

$$\mathbf{C_y} = cov\left(\mathbf{Y}'^T\right) = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1r} \\ C_{21} & C_{22} & \cdots & C_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r1} & C_{r2} & \cdots & C_{rr} \end{pmatrix}, \tag{11}$$

Calculate all eigenvalues λ_u and corresponding eigenvectors $\mathbf{q}_u$ in matrix $\mathbf{C_y}$, where $u = 1, 2, \ldots, r$. Rank the eigenvalues λ_u from largest to smallest and calculate the completeness of the first g eigenvalues to the overall information, which also becomes the cumulative contribution rate $\eta(g)$,

$$\eta(g) = \sum_{k=1}^{g} \lambda_k / \sum_{k=1}^{r} \lambda_k. \tag{12}$$

When $\eta(g)$ is greater than the preset rate value, the eigenvectors corresponding to the first g eigenvalues are combined to form the transformation matrix $\mathbf{P}'$,

$$\mathbf{P}' = \left[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_g\right] \quad g < r. \tag{13}$$

On this basis, the dimensionality reduction matrix $\mathbf{Z}$ is obtained,

$$\mathbf{Z} = \left(\mathbf{P}'\right)^T \mathbf{Y}. \tag{14}$$

As an unsupervised dimensionality reduction algorithm, PCA does not require data prior information. In this paper, PCA is used to reduce the dimensionality of the extracted features.

After feature extraction, the calculation of the obtained sample feature matrix $\mathbf{A}_{J\times M_i}$ is too complicated, so it needs to reduce the feature dimension. This paper uses PCA to solve this problem. After the sample feature matrix $\mathbf{A}_{J\times M_i}$ is processed by PCA, the projection matrix $\mathbf{W} = \left\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_{K_i}\right\}$ is obtained. Thus, fast computation from high dimension to low dimension can be realized.

To ensure the reliability of modulation recognition and further improve the effectiveness of the system, PCA is used to reduce the dimension to remove the redundant information from the huge number of features. Then the features are fused. The i-th value of dimension reduction parameter K_i is the

$$\mathbf{A}_{J\times M_i} = \begin{bmatrix} fea_{11}(\mathbf{NET}[net_i]) & fea_{12}(\mathbf{NET}[net_i]) & \cdots & fea_{1m}(\mathbf{NET}[net_i]) & \cdots & fea_{1M_i}(\mathbf{NET}[net_i]) \\ fea_{21}(\mathbf{NET}[net_i]) & fea_{22}(\mathbf{NET}[net_i]) & \cdots & fea_{2m}(\mathbf{NET}[net_i]) & \cdots & fea_{2M_i}(\mathbf{NET}[net_i]) \\ \vdots & \vdots & & \vdots & & \vdots \\ fea_{j1}(\mathbf{NET}[net_i]) & fea_{j2}(\mathbf{NET}[net_i]) & \cdots & fea_{jm}(\mathbf{NET}[net_i]) & \cdots & fea_{jM_i}(\mathbf{NET}[net_i]) \\ \vdots & \vdots & & \vdots & & \vdots \\ fea_{J1}(\mathbf{NET}[net_i]) & fea_{J2}(\mathbf{NET}[net_i]) & \cdots & fea_{Jm}(\mathbf{NET}[net_i]) & \cdots & fea_{JM_i}(\mathbf{NET}[net_i]) \end{bmatrix}, \tag{10}$$

number of eigenvectors contained in the dimension reduction matrix. For the selection of K_i, $\mathbf{W}$ is defined as a matrix consisting of column vectors of all feature mapping vectors, which can preserve the information in the data. An optimal function can be obtained by algebraic linear transform from the matrix as follows,

$$\max_{\vec{w}_{K_i}} tr(\mathbf{W}^T \mathbf{A}_{J\times M_i} \mathbf{A}_{J\times M_i}{}^T \mathbf{W}), \quad s.t. \mathbf{W}^T \mathbf{W} = \mathbf{I}, \tag{15}$$

where $tr(\cdot)$ is the trace function of matrix.Because of the large amount of calculation in formula (11), the following formula replaces it.

$$g_1(\vec{w}_{K_1}) = \min_{K_1} \left\| \frac{\sum_{m=1}^{K} (\mathbf{A}_{J\times M_i} \cdot \mathbf{A}_{J\times M_i}{}^T)}{\sum_{m=1}^{M_i} (\mathbf{A}_{J\times M_i} \cdot \mathbf{A}_{J\times M_i}{}^T)} - CON \right\|_2, \tag{16}$$

where CON represents contribution degree. Then fusion of these features as follows,

$$g(\vec{w}_K) = [g_1(\vec{w}_{K_1}), g_2(\vec{w}_{K_2}), g_3(\vec{w}_{K_3})]. \tag{17}$$

Similar to network coding, the number of features is also encoded after PCA dimension reduction. The final new dimension reduction coding matrix is as follows,

$$\mathbf{EC} = [EC_1, EC_2, EC_3], \tag{18}$$

where $EC_i (i = 1, 2, 3)$ represents the signal features that can be retained after feature dimension reduction, and the value range of EC_i is,

$$EC_i \geq K_i. \tag{19}$$

According to the range of EC_i, the number of features after PCA dimension reduction is optimized and selected, which can be well distinguished by the classifier. After dimension reduction of a large number of features, the features are fused. These features are sent to SVM for classification.

D. CLASSIFICATION

After the above signal processing and feature extraction of CNNS, we get a large number of image features for dimension reduction. Then we send these features into the classifier to achieve the goal of signal modulation recognition. SVM transforms the eigenvector into a high-dimensional feature space by using non-linear mapping and constructs the optimal classification hyperplane in a high-dimensional feature space. Then the purpose of classifying in high-dimensional feature space can be achieved. According to the principle of functional, a kernel function $K(\vec{w}, \vec{w})$ is introduced as,

$$K(\vec{w}, \vec{w}) = \phi(\vec{w}) \cdot \phi(\vec{w}), \tag{20}$$

where $\vec{w}$ is any one of the feature vectors after feature fusion. $\phi(\vec{w})$ mapping function of feature space.The selection of penalty factor C determines the performance of the classifier. In order to achieve good recognition and classification results, we need to optimize the parameters of SVM kernel function and penalty factor. Therefore, the objective function can be obtained,

$$\begin{cases} \phi(\vec{w}) = \frac{1}{2}\|\vec{w}\|^2 + C\sum_{l=1}^{n} \xi_l, \\ \text{subject to} : y_l[(\eta \cdot \vec{w}) - b] + \xi_l - 1 > 0, \quad \xi_l > 0, \end{cases} \tag{21}$$

where y_l is the class mark of $\vec{w}$, which realizes the transition from non-linear classification to linear classification, η is the normal vector of the hyperplane, C is the penalty factor for misclassification samples, ξ_l is the relaxation variable for relaxing constraints, b is the threshold. To find the minimum of the objective function is to solve the quadratic convex optimization problem. Because of the complexity of calculation, this paper transforms the above functions into dual problems,

$$Q(\beta) = \sum_{i=1}^{n} \beta_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{n} \beta_i \beta_l y_i y_l K(\vec{w}, \vec{w}), \tag{22}$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ is the Lagrange multiplier. By further simplification, the final decision function is as follows,

$$f(\vec{w}) = sgn[(\eta \cdot \phi(\vec{w})) + b]. \tag{23}$$

Generally, the sigmoid kernel function is widely used. It increases flexibility by adjusting parameters,

$$K(\vec{w}, \vec{w}) = \tanh(\sigma \cdot (\vec{w}^T \vec{w}) + \theta), \tag{24}$$

where σ and θ are the parameters of the kernel function, which decide the classification results. Combined with formula (17), we encode the classifier's parameters to be optimized to obtain the solution space,

$$\mathbf{SV} = [C, \sigma, \theta], \tag{25}$$

where the constraints of C, σ, θ are,

$$\begin{cases} C > 0, \\ \sigma > 0, \\ \theta \leq 0. \end{cases} \tag{26}$$

Combining the formulas (8), (18) and (25), parameter vector to be optimized is as follows,

$$\begin{aligned} \mathbf{V}' &= [\mathbf{NET}, \mathbf{EC}, \mathbf{SV}] \\ &= [net_1, net_2, net_3, EC_1, EC_2, EC_3, C, \sigma, \theta]. \end{aligned} \tag{27}$$

These values of parameters affect the performance of the classification system, so it is necessary to further optimize the parameters through an intelligent algorithm.

IV. PARAMETER OPTIMIZATION BASED ON FLYING FISH SWARM ALGORITHM

Swarm intelligence algorithm has good performance in parameter optimization. Different swarm intelligence algorithms have different optimization characteristics, and their accuracy and processing speed are also different.

Artificial fish swarm algorithm is a kind of swarm intelligence algorithm using the behavior law of biota [41].

Based on the characteristics of fish foraging, the algorithm simulates four behaviors: foraging, team, track, and random. In practical application, the algorithm will optimize the objective function through these four behavior modes.

1) Foraging behavior.
The problem solved by the algorithm can be regarded as the extremum problem, and the maximum and minimum problems can be converted into each other. Take the minimum, for example, if the current position of individual fish i in the shoal is $\mathbf{X}_i$, randomly select a position $\mathbf{X}_j$ within its visual range D,

$$\mathbf{X}_j = \mathbf{X}_i + rand(0, 1) \cdot D. \tag{28}$$

where rand (0,1) represents the random number between 0 and 1.
When $Y_i > Y_j$ is satisfied, the other individual fish moves one step in this direction, and the moving distance is the moving step length of the individual fish, Y_i is the food source concentration corresponding to individual fish i,

$$\mathbf{X}_{i|NEXT} = \mathbf{X}_i + rand(0, 1) \cdot G \cdot \frac{\mathbf{X}_j - \mathbf{X}_i}{\left\|\mathbf{X}_j - \mathbf{X}_i\right\|}. \tag{29}$$

If not, a new position is randomly selected, then proceed to determine whether the forward condition is met. After n attempts, the shoal will behave randomly.

2) Random behavior.
Random behavior is the act of randomly selecting a position within the visual range of an individual fish and then moving one step in that direction,

$$\mathbf{X}_{i|NEXT} = \mathbf{X}_i + rand(0, 1) \cdot G \tag{30}$$

3) Team behavior.
The individual fish in position $\mathbf{X}_i$ can find the number of its partner fish in the range of vision is N_f, if the conditions are met,

$$d_{i,j} = \left\|\mathbf{X}_i - \mathbf{X}_j\right\| < D, \tag{31}$$

$\mathbf{X}_j$ is regarded as its partner fish. The position of each partner fish is used to calculate the corresponding central position $\mathbf{X}_c$ and the food source concentration Y_c. If the conditions are met,

$$\frac{Y_c}{N_f} > \delta Y_i, \tag{32}$$

where δ represents the crowding factor of fish swarm. This formula shows that the food concentration in the partner center is large and the crowding degree is low, so one step can be moved toward position $\mathbf{X}_c$, and the mode of movement is,

$$\mathbf{X}_{i|NEXT} = \mathbf{X}_i + rand(0, 1) \cdot G \cdot \frac{\mathbf{X}_c - \mathbf{X}_i}{\|\mathbf{X}_c - \mathbf{X}_i\|} \tag{33}$$

If not, perform foraging behavior.

4) Track behavior.
After finding the partner fish and the corresponding food source concentration according to the team behavior, find the partner fish $\mathbf{X}_{\max}$ with the largest food concentration $Y_{\max}$, if the conditions are met,

$$\frac{Y_{\max}}{N_f} > \delta Y_i, \tag{34}$$

it indicates that the position of partner fish $\mathbf{X}_{\max}$ has higher food concentration, and the crowding near the fish is low, so move towards the position $\mathbf{X}_{\max}$. If not, perform foraging behavior.

The artificial fish swarm algorithm is not directly applicable to the situation with large solution space, but the cuckoo algorithm can increase the diversity of solutions by using the Levy flight mode and has excellent global performance. Therefore, a new FFS algorithm is proposed in this paper, which utilizes the flight mode of Levy's flight between long and short distances to optimize the movement mode of individual fish in the artificial fish swarm algorithm, so as to ensure the diversity of solutions in the optimization process of the proposed algorithm and improve the adaptability of the proposed algorithm to a large solution space.

$L(\lambda)$ is Levy random search path and the random step size is Levy distribution [42],

$$L(\lambda) = \frac{u}{|v|^{\frac{1}{\lambda}}}, \tag{35}$$

where λ is Levy distribution index, $\lambda = 1.5$. u obeys normal distribution,

$$u \sim N(0, \sigma_n^2), \tag{36}$$

where the variance σ_n^2 is,

$$\sigma_n^2 = \left(\frac{\Gamma(1+\lambda) \times \sin\left(\pi \times \frac{\lambda}{2}\right)}{\Gamma\left\{\left(\frac{1+\lambda}{2}\right) \times \lambda \times 2^{\frac{\lambda-1}{2}}\right\}}\right)^{\frac{2}{\lambda}}, \qquad \Gamma(\lambda) = \int_0^\infty e^{-u} u^{\lambda-1} du. \tag{37}$$

v obeys standard normal distribution,

$$v \sim N(0, 1). \tag{38}$$

Change the step size of the artificial fish swarm algorithm from a fixed value to Levy flight update:

$$\mathbf{X}_{i|NEXT} = \mathbf{X}_i + \alpha \oplus L(\lambda) \tag{39}$$

where $\oplus$ is point-to-point multiplication, α is step size control quantity, which is used to control the search range of step size.

The moving step size G of individual fish in the FFS algorithm can be changed to *step*,

$$step = \left\{\frac{\Gamma(1+\lambda) \times \sin\left(\pi \times \frac{\lambda}{2}\right)}{\Gamma\left\{\left[\frac{1+\lambda}{2}\right] \times \lambda \times 2^{\frac{\lambda-1}{2}}\right\} \cdot N(0,1)}\right\}^{\frac{1}{\lambda}} \cdot N(0, 1) \cdot G, \tag{40}$$

where, $N(0, 1)$ represents the random number between 0 and 1 that conforms to the standard normal distribution, G is the maximum moving step of flying fish.

The random behavior of the artificial fish swarm algorithm is changed into the cuckoo search process, and the specific update method is combined with the Levy flight optimization idea, so the modified artificial fish swarm behavior is as follows,

$$\mathbf{X}_{i|NEXT} = \mathbf{X}_i + rand(0,1) \cdot \left\{ \frac{\Gamma(1+\lambda) \times \sin\left(\pi \times \frac{\lambda}{2}\right)}{\Gamma\left\{\left[\frac{1+\lambda}{2}\right] \times \lambda \times 2^{\frac{\lambda-1}{2}}\right\} \cdot N(0,1)} \right\}^{\frac{1}{\lambda}} \cdot N(0,1) \cdot G \cdot \frac{\mathbf{X}_c - \mathbf{X}_i}{\|\mathbf{X}_c - \mathbf{X}_i\|}. \quad (41)$$

Corresponding track behavior is expressed as:

$$\mathbf{X}_{i|NEXT} = \mathbf{X}_i + rand(0,1) \cdot \left\{ \frac{\Gamma(1+\lambda) \times \sin\left(\pi \times \frac{\lambda}{2}\right)}{\Gamma\left\{\left[\frac{1+\lambda}{2}\right] \times \lambda \times 2^{\frac{\lambda-1}{2}}\right\} \cdot N(0,1)} \right\}^{\frac{1}{\lambda}} \cdot N(0,1) \cdot G \cdot \frac{\mathbf{X}_{\max} - \mathbf{X}_i}{\|\mathbf{X}_{\max} - \mathbf{X}_i\|}. \quad (42)$$

In order to further improve the convergence accuracy of the algorithm, this paper proposes a further optimization design for the random range of step size,

$$Rand\,(\cdot) = rand(0, e^{1/N} - 1) \quad (43)$$

where, $rand(0, e^{1/N} - 1)$ represents the random number between 0 and $e^{1/N} - 1$. With the increase of the number of iterations N, the random range will gradually decrease, and a fine search near the optimal solution will be conducted with a smaller step length, finally obtaining a solution with better precision.

In this paper, the initial fish population is NP, the maximum number of iterations is ITA, the maximum number of attempts is NUM, the flight steps control amount is α, the maximum moving step of flying fish is G, the perceptual range of flying fish is PER, and the congestion factor is Q. When the dimension of the target search space is D, the individual state of flying fish is set as vector $\mathbf{V} = [v_1, v_2, \ldots, v_D]$. The behavior patterns of FFS populations are flying foraging behavior, team behavior, and track behavior.

1) Solution space.
The selection of the network layer and the number of features in the solution space is discretized, while the range of parameters in the classifier is continuous. When using the FFS algorithm to optimize the solution space, it is very difficult to balance the two aspects. In other words, when the algorithm is updated, there will be a continuous number of updates to the feasible solution. However, the network layer and the number of features need to be represented by integers. To find the corresponding network layer and the number of features in the solution space, we round and optimize some parameters of the solution space. Therefore, the paper partially consolidates the solution space,

$$\mathbf{V} = \left[\left\lceil net_1' \right\rceil, \left\lceil net_2' \right\rceil, \left\lceil net_3' \right\rceil, \times \left\lceil EC_1' \right\rceil, \left\lceil EC_2' \right\rceil, \left\lceil EC_3' \right\rceil, C, \sigma, \theta\right], \quad (44)$$

where $\lceil \cdot \rceil$ means round up, $net_i'(i = 1, 2, 3)$ is the continuous corresponding representation at the network layer which value can be found corresponding to the network layer. Since it is meaningless for the network layer net_i' is zero, this point is deleted. This condition also applies to the number of features $\lceil EC_i' \rceil$ of the i-th dimension reduction. The determination of each parameter to be optimized in the solution space will affect the features $\vec{w}$ after dimension reduction and the recognition result of the classifier.

2) Objective function.
According to the decision function of the classifier, it can be known that the smaller the deviation between the result of recognizing and the signal label, the better the recognition rate. Therefore, the objective function is established as follows,

$$Y(\vec{w}) = \|sgn[(\eta \cdot \phi(\vec{w})) + b] - y_l\|_2^2 + \|K(\vec{w}, \vec{w}) - \phi(\vec{w}) \cdot \phi(\vec{w})\|_2. \quad (45)$$

If this objective function is smaller, the recognition rate is excellent. This paper uses the FFS algorithm to find the optimal value to find the minimum value of the objective function.

3) Optimizing behavior.
In the ita-th ($ita = 1, 2, \ldots, ITA$) generation flying fish, the current state of flying fish is $\mathbf{V}_b^{(ita)}(b = 1, 2, \ldots, NP)$. Random selection of a state $\mathbf{VI}_p^{(ita)}$ within its perception range satisfies $d_{b,p} = \left\|\mathbf{VI}_p^{(ita)} - \mathbf{V}_b^{(ita)}\right\| < PER$. If the food concentration is $Y_b^{(ita)}(\vec{w}) > Y_{Ip}^{(ita)}(\vec{w})$, flying fish will move in the direction,

$$\mathbf{V}_{bnext}^{(ita)} = \mathbf{V}_b^{(ita)} + Rand(\cdot) \cdot step \cdot \frac{\mathbf{VI}_p^{(ita)} - \mathbf{V}_b^{(ita)}}{\left\|\mathbf{VI}_p^{(ita)} - \mathbf{V}_b^{(ita)}\right\|}, \quad (46)$$

Otherwise, the state $\mathbf{VI}_p^{(ita)}$ is re-selected, and after the NUM-th attempts, the condition is still not satisfied, then the flying foraging behavior is carried out.

Flying fish then performs team behavior. In the ita-th ($ita = 1, 2, \ldots, ITA$) generation of flying fish, the current state of flying fish is $\mathbf{V}_b^{(ita)}$. It is known that there are NF flying fishes in its perception range. The central position of these flying fish is $\mathbf{VC}^{(ita)}$,

$$\mathbf{VC}^{(ita)} = \frac{1}{NF} \sum_{p=1}^{NF} \mathbf{VI}_p^{(ita)}. \quad (47)$$

If $\mathbf{VC}^{(ita)} \cdot NF < Q \cdot \mathbf{V}_b^{(ita)}$ is satisfied, indicating that the center is located at a high food concentration and not too

crowded, then the flying fish takes a step forward to the center location,

$$\mathbf{V}_{bnext}^{(ita)} = \mathbf{V}_{b}^{(ita)} + Rand(\cdot) \cdot step \cdot \frac{\mathbf{VC}^{(ita)} - \mathbf{V}_{b}^{(ita)}}{\left\| \mathbf{VC}^{(ita)} - \mathbf{V}_{b}^{(ita)} \right\|}. \quad (48)$$

If $\mathbf{VC}^{(ita)} \cdot NF \geq Q \cdot \mathbf{V}_{b}^{(ita)}$, the flying foraging behavior will be carried out. Then perform the track behavior. It is known that the optimal state of flying fish in the perceptual range is $\mathbf{V}_{\max}^{(ita)} = \max\left\{\mathbf{V}_{1}^{(ita)}, \mathbf{V}_{2}^{(ita)}, \ldots, \mathbf{V}_{NF}^{(ita)}\right\}$. If $\mathbf{V}_{\max}^{(ita)} \cdot NF < Q \cdot \mathbf{V}_{b}^{(ita)}$ is satisfied, the results show that the higher food concentration with less crowded, the flying fish moves forward one step in the direction,

$$\mathbf{V}_{bnext}^{(ita)} = \mathbf{V}_{b}^{(ita)} + Rand(\cdot) \cdot step \cdot \frac{\mathbf{V}_{\max}^{(ita)} - \mathbf{V}_{b}^{(ita)}}{\left\| \mathbf{V}_{\max}^{(ita)} - \mathbf{V}_{b}^{(ita)} \right\|}. \quad (49)$$

Otherwise, flying foraging behavior will be carried out. The parameter optimization process of the FFS algorithm is given by Algorithm1 and Algorithm 2. This paper uses the FFS algorithm to optimize the objective function. Through continuous iteration, the solution space will eventually converge to a fixed value $\mathbf{V}_{best}$. First, the network layer in the solution space is used to extract the features of the signals. Second, the number of features can be fixed. Finally, the classification is combined with the penalty factor and kernel function parameters in the SVM to obtain the recognition result. Besides, the criteria for selecting this algorithm to solve the problem in this paper are as follows: On the one hand, there is a certain mapping relationship between the problem to be solved and the solution space. The solution space is a collection of finite states. On the other hand, the extreme point must exist in the solution space.

TABLE 2. Signal parameters.

Content	Detailed description
Carrier frequency	10MHz
Sampling frequency	1.12MHz
Symbol rate	280kHz
Number of symbol	1024
Sample value	400 samples for each modulation signal (280 training samples and 120 test samples)
SNR	-6dB, -4dB, -2dB, 0dB, 2dB, 4dB, 6dB

V. SIMULATION AND DISCUSSION

This paper assumes that the received modulation includes BPSK, QPSK, 8PSK, OQPSK, 4ASK, QAM16, QAM32, and QAM64. Each signal is disturbed by additive white Gaussian noise. SNR is defined as $SNR = 10\log_{10}(\sigma_s^2)/(\sigma_\varepsilon^2)$, where σ_s^2 is the variance of signal and σ_ε^2 is the variance of white Gaussian noise. Signal parameters are shown in Table 2. The experimental equipment of the simulation operating system is Windows 10(Microsoft, Redmond, WA, USA), the processor is Intel R CoreTM i7-8700K (Intel, Santa Clara, CA, USA), the memory is 64GB. To make the simulation results universal, the simulation results were obtained through 800 Monte Carlo experiments. The parameters are used to complete the signal processing. Based on this, the FFS algorithm is used to optimize the parameters. Table 3 shows the initial parameters of the FFS algorithm.

Algorithm 1 Flying Fish Swarm Algorithm Part 1

Input: Population size NP, maximum number of iterations ITA, maximum number of attempts NUM, flight step control amount α, maximum moving step of flying fish G, perceptual range of flying fish PER, congestion factor Q.

Ouput: Optimal solution $\mathbf{V}_{best}$

Initialize flying fish parameters, NP feasible solutions $\mathbf{V}_{b}^{(ita)}(b = 1, 2, \ldots, NP)$
for each $ita = 1 : ITA$ **do**
  for each $b = 1 : NP$ **do**
    find NF, randomly select $\mathbf{VI}_{p}^{(ita)}$ according to $\left\| \mathbf{VI}_{p}^{(ita)} - \mathbf{V}_{b}^{(ita)} \right\| < PER$
    $\mathbf{VC}^{(ita)} \leftarrow \left\{\mathbf{VI}_{p}^{(ita)}, NF\right\}$ according to formula(48)
    if $\mathbf{VC}^{(ita)} \cdot NF < Q \cdot \mathbf{V}_{b}^{(ita)}$
      $\mathbf{V}_{bnext}^{(ita)} \leftarrow \left\{\mathbf{V}_{b}^{(ita)}, \mathbf{VC}^{(ita)}, G\right\}$ according to formula(49)
      $Y_{bnext_1}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{bnext}^{(ita)}$ according to formula(46)
    else
      for each $r = 0 : NUM$ **do**
        $Y_{b}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{b}^{(ita)}$, $Y_{p}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{VI}_{p}^{(ita)}$ according to formula(46)
        if $Y_{b}^{(ita)}(\vec{w}) > Y_{Ip}^{(ita)}(\vec{w})$
          $\mathbf{V}_{bnext}^{(ita)} \leftarrow \left\{\mathbf{V}_{b}^{(ita)}, \mathbf{VI}_{p}^{(ita)}, G\right\}$
          $Y_{bnext_1}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{bnext}^{(ita)}$ according to formula(46)
          break
        elseif $r = NUM$
          $\mathbf{V}_{bnext}^{(ita)} \leftarrow \left\{\mathbf{V}_{b}^{(ita)}, \alpha, L(\lambda)\right\}$ according to formula(35)
          $Y_{bnext_1}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{bnext}^{(ita)}$ according to formula(46)
        else
          randomly select $\mathbf{VI}_{p}^{(ita)}$ according to $\left\| \mathbf{VI}_{p}^{(ita)} - \mathbf{V}_{b}^{(ita)} \right\| < PER$
      end for

TABLE 3. The initial parameters of FFS algorithm.

Parameter	Symbol	Value
Population size	NP	200
Maximum number of attempts	NUM	50
Flight step control amount	α	1.5
Maximum moving step of flying fish	G	0.1
Perceptual range of flying fish	PER	1
Congestion factor	Q	0.6

Algorithm 2 Flying Fish Swarm Algorithm Part 2

$\mathbf{V}_{\max}^{(ita)} \leftarrow \left\{\mathbf{VI}_p^{(ita)}, NF\right\}$
if $\mathbf{V}_{\max}^{(ita)} \cdot NF < Q \cdot \mathbf{V}_b^{(ita)}$
$\mathbf{V}_{bnext}^{(ita)} \leftarrow \left\{\mathbf{V}_b^{(ita)}, \mathbf{V}_{\max}, G\right\}$ according to formula(49)
$Y_{bnext_2}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{bnext}^{(ita)}$ according to formula(46)
else
for each $r = 0 : NUM$ **do**
$Y_b^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_b^{(ita)}$, $Y_{Ip}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{VI}_p^{(ita)}$ according to formula(46)
if $Y_b^{(ita)}(\vec{w}) > Y_{Ip}^{(ita)}(\vec{w})$
$\mathbf{V}_{bnext}^{(ita)} \leftarrow \left\{\mathbf{V}_b^{(ita)}, \mathbf{VI}_p^{(ita)}, G\right\}$
$Y_{bnext_2}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{bnext}^{(ita)}$ according to formula(46)
break
elseif $r = NUM$
$\mathbf{V}_{bnext}^{(ita)} \leftarrow \left\{\mathbf{V}_b^{(ita)}, \alpha, L(\lambda)\right\}$ according to formula(35)
$Y_{bnext_2}^{(ita)}(\vec{w}) \leftarrow \vec{w} \leftarrow \mathbf{V}_{bnext}^{(ita)}$ according to formula(46)
else
randomly select $\mathbf{VI}_p^{(ita)}$ according to $\left\|\mathbf{VI}_p^{(ita)} - \mathbf{V}_b^{(ita)}\right\| < PER$
end for
$\mathbf{V}_{best} \leftarrow Y_{best}^{(ita)}(\vec{w}) \leftarrow \min\left\{Y_{bnext_1}^{(ita)}(\vec{w}), Y_{bnext_2}^{(ita)}(\vec{w})\right\}$
end for
end for

A. COMBINATION MODE

Table 4 shows the recognition rate of a single signal processing method through a single neural network when the SNR is 0dB. The sample size is 400 and the number of features after dimension reduction is 150. In the table, we use CS, CWD, and STFT to process eight kinds of communication signals. Four kinds of neural networks are used to extract the features of signals. The features of the fc6 layer in AlexNet and VGGNet, the features of the icp4 layer in GoogLeNet, and the features of the res3a layer in ResNet are extracted and sent to PCA for dimension reduction respectively. Then, implement classification through SVM.

As can be seen from Table 4, the recognition rate using CS and ResNet is the best, reaching 90.2 % at SNR of 0 dB. Using STFT and then using AlexNet for feature extraction has the worst classification recognition rate, with a recognition rate of 81.7%. For the same network, different signal processing methods will get different recognition results. On the one hand, for AlexNet, VGGNet, and ResNet, when the network type and network layer are fixed, the signal processing method of CS is better than that of STFT and CWD. However, for GoogLeNet, the recognition rate obtained by using CWD is better than CS and STFT. On the other hand, for CS, the recognition rate obtained by applying ResNet for feature extraction is better than the other three networks. For STFT, the recognition rate obtained by applying ResNet for feature extraction is also better than the other three networks. However, for CWD, the recognition rate obtained by GoogLeNet is better than other networks in the table. In brief, different signal processing methods have different effects on the recognition rate. Not only that, but different neural network structures also have different effects on the recognition rate. Table 5 shows the recognition rate of different network layers for the same neural network. For the three signal processing methods of CS, CWD, and STFT, we use icp1 to icp9 layers of GoogLeNet to extract the features of communication signals. Then we send these features to PCA for dimension reduction, and finally, send them to SVM for classification.

It can be seen from Table 5 that when the SNR is 0dB, different network layers of the same neural network have different recognition rates on communication signals. For the same network layer of the same neural network, different recognition rates can still be obtained by using different signal processing methods. For CS and CWD, using the features of the icp4 layer in GoogLeNet compared to other network layers can get the best results. However, for STFT, using the features of the icp5 layer in GoogLeNet for classification can get better recognition results than other network layers.

Based on the above analysis, different signal processing methods, different neural networks, and different neural network layers will have different effects on the recognition results of communication signals. However, it is impossible to enumerate these combinations one by one. To further improve the modulation recognition rate of communication signals, an optimization algorithm is needed to find the optimal combination.

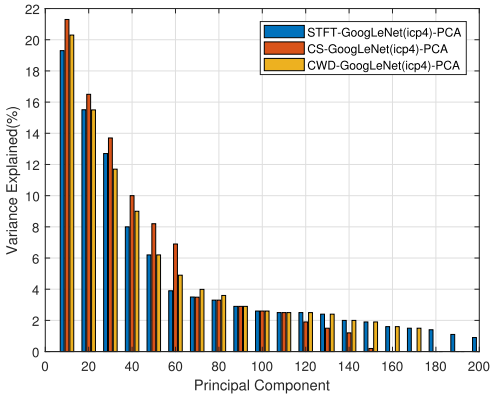


FIGURE 6. The figure shows the analysis of PCA dimension reduction components, blue bar, orange bar and yellow bar are STFT-GoogLeNet(icp4)-PCA, CS-GoogLeNet(icp4)-PCA and CWD-GoogLeNet(icp4)-PCA respectively.

B. DIMENSION REDUCTION

In this part, we use PCA to reduce the dimension of the extracted image features. Fig.6 shows the contribution of

TABLE 4. The recognition rate of a single signal processing method through a single neural network.

Signal Processing	Layers of Network	Recognition Rate	Signal Processing	Layers of Network	Recognition Rate
CS	AlexNet(fc6)	85.6%	CS	GoogLeNet(icp4)	89.2%
CWD	AlexNet(fc6)	85.1%	CWD	GoogLeNet(icp4)	89.9%
STFT	AlexNet(fc6)	81.7%	STFT	GoogLeNet(icp4)	86.8%
CS	VGGNet(fc6)	86.9%	CS	ResNet(res3a)	90.2%
CWD	VGGNet(fc6)	86.5%	CWD	ResNet(res3a)	88.5%
STFT	VGGNet(fc6)	84.8%	STFT	ResNet(res3a)	87.3%

TABLE 5. The recognition rate of different network layers for the same neural network.

Layers of Network	Signal Processing	Recognition Rate	Signal Processing	Recognition Rate	Signal Processing	Recognition Rate
GoogLeNet(icp1)	CS	77.6%	CWD	80.8%	STFT	69.5%
GoogLeNet(icp2)	CS	84.3%	CWD	83.6%	STFT	76.8%
GoogLeNet(icp3)	CS	88.7%	CWD	86.3%	STFT	82.4%
GoogLeNet(icp4)	CS	89.2%	CWD	89.9%	STFT	86.8%
GoogLeNet(icp5)	CS	89.1%	CWD	88.7%	STFT	87.7%
GoogLeNet(icp6)	CS	86.5%	CWD	85.2%	STFT	86.1%
GoogLeNet(icp7)	CS	84.1%	CWD	84.2%	STFT	85.7%
GoogLeNet(icp8)	CS	83.4%	CWD	80.4%	STFT	80.7%
GoogLeNet(icp9)	CS	75.7%	CWD	70.7%	STFT	76.5%

TABLE 6. The part of cumulative contribution.

CS			CWD			STFT		
Com-ponent	Individual Contribution	Cumulative Contribution	Com-ponent	Individual Contribution	Cumulative Contribution	Com-ponent	Individual Contribution	Cumulative Contribution
121	0.24%	93.79%	141	0.48%	94.77%	141	0.38%	93.99%
122	0.23%	94.02%	**142**	**0.35%**	**95.12%**	142	0.31%	94.30%
123	0.21%	94.23%	143	0.29%	95.41%	143	0.27%	94.57%
124	0.20%	94.43%	144	0.21%	95.62%	144	0.25%	94.82%
125	0.20%	94.63%	145	0.18%	95.80%	**145**	**0.23%**	**95.05%**
126	0.19%	94.82%	146	0.15%	95.95%	146	0.21%	95.26%
127	0.17%	94.99%	147	0.12%	96.07%	147	0.17%	95.43%
128	**0.17%**	**95.16%**	148	0.11%	96.18%	148	0.17%	95.60%
129	0.06%	95.22%	149	0.11%	96.29%	149	0.14%	95.74%
130	0.02%	95.24%	150	0.08%	96.37%	150	0.08%	95.82%

PCA feature dimension reduction for multi-transform. The features are sorted in descending order of contribution. The first 200 features after PCA were divided into 20 groups at intervals of 10. The value of PCA main component parameters determines the effect of dimension reduction for high-dimensional feature data. The three groups of histograms are CWD-GoogLeNet(icp4)-PCA, CS-GoogLeNet(icp4)-PCA and STFT-GoogLeNet(icp4)-PCA. After dimension reduction through PCA, some features that can reflect the principle contribution are retained. By grouping features, these main components can meet the contribution requirements. The part of contribution is shown in Table 6. As can be seen from Table 6, when the number of the features is 128, the cumulative contribution reaches 95.16% in the dimension reduction method based on CS-PCA analysis. When the component is 142, the cumulative contribution reaches 95.12% in the dimension reduction method based on CWD-PCA analysis. When the component is 145, cumulative contribution reaches 95.05% in the dimension reduction method based on STFT-PCA analysis. Generally, when the contribution degree $CON = 0.95$, the corresponding principal component $K_1 = 128$, $K_2 = 145$ and $K_3 = 142$.

The data analysis shows that the number of features obtained by PCA and STFT is more than the other two transforms. STFT cannot take good care of the local characteristics of the signal when analyzing the communication signal, because of the low resolution. Therefore, after dimension reduction, more features are needed to represent each signal. In the PCA based on CWD, it can achieve higher resolution than STFT, which makes CWD require fewer features than STFT. In PCA based on CS, the principal component is the least among the three signal processing methods. Because CS can effectively suppress the stationary colored noise. Moreover, CS analyzes signals from frequency and cyclic frequency. Communication signals mostly have cyclic stationary characteristics. CS can better characterize each signal. After dimension reduction, modulation of the signal can be characterized by relatively few features. Therefore, the number of different features will affect the corresponding component values, which will further affect the modulation recognition rate of the signal.

Under the condition that the contribution degree *CON* is satisfied, Fig.7 shows the recognition rate of the different number of features. When the GoogLeNet(icp4) is used

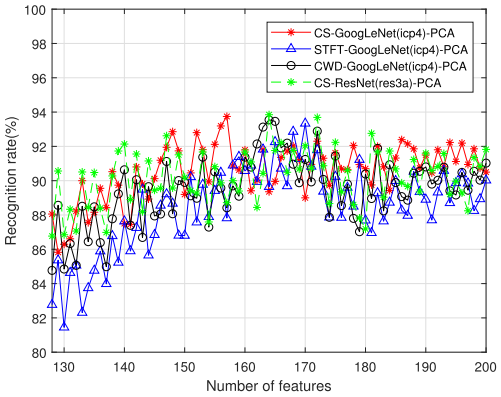


FIGURE 7. The figure shows recognition rate of different number of features, red line, blue line, black line and green line are CS-GoogLeNet(icp4)-PCA, STFT-GoogLeNet(icp4)-PCA, CWD-GoogLeNet(icp4)-PCA, and CS-ResNet(icp4)-PCA.

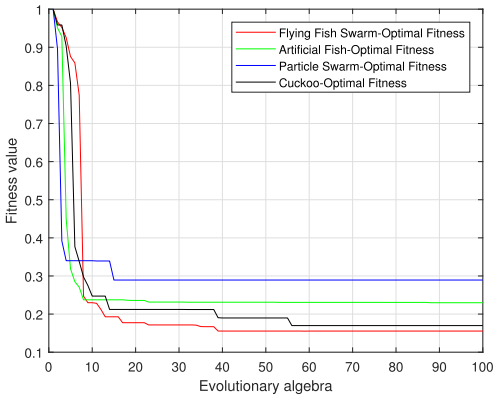


FIGURE 8. The figure shows the convergence performance of different algorithms, red line, green line, blue line and black line are flying fish swarm-optimal fitness, artificial fish-optimal fitness, particle swarm-optimal fitness and cuckoo-optimal fitness respectively.

to extract features, the recognition effects of PCA feature dimension corresponding to different signal processing methods are different. Taking CS as an example, features extracted from GoogLeNet(icp4) and ResNet(res3a) have different PCA dimension reduction effects. Different signal processing methods and networks extract different features, and then the PCA dimension reduction effect is different.

Besides, as the number of features increases, the accuracy of modulation recognition increases. As the number of features continues to increase, the recognition rate will slowly decrease. The reason for the increase in the recognition rate in the figure is that as the features increase, the classifier can classify them more clearly. However, when the features increase to a certain extent, the number of redundant information increases due to the overlapping of the features obtained by PCA dimension reduction. Besides, when these features are fused, the recognition rate will be affected by the amount of feature redundancy. However, the influence degree from the three-way feature cannot be known. To further improve the recognition rate after feature fusion, this paper uses the proposed intelligent algorithm to optimize parameters. It should be added that the fluctuation of the curve in the figure is caused by the randomness of a single recognition method. The three curves in the figure have different recognition rates. Because the number of features required is different after different signal processing.

C. PARAMETER OPTIMIZATION BASED ON INTELLIGENT ALGORITHM

Fig.8 shows the convergence performance of different algorithms. The fitness function value is normalized. The number of iterations is 100 times. The fitness value obtained by the FFS algorithm proposed in this paper is the smallest. The algorithm finally converges to 0.156. There is an improvement of 0.075 over the performance of the artificial fish algorithm and an improvement of 0.014 over the Cuckoo algorithm. Besides, the number of convergence iterations of the proposed algorithm has 16 more iterations than the artificial fish algorithm, and 17 fewer iterations than the Cuckoo algorithm. As for particle swarm, the algorithm finally converges to 0.289. The number of convergence iterations is 16.

In the early stage of iteration, the particle swarm optimization algorithm has the fastest convergence performance, and the FFS algorithm proposed in this paper has the slowest convergence performance. This is because the updated principle of particle swarm optimization is relatively easy. The FFS algorithm is based on the artificial fish swarm algorithm combined with the Levy flight optimization mechanism in the Cuckoo algorithm. The updating method of multiple behavior optimization leads to slow convergence in the early stage. During the later stage of iteration, due to the poor global search ability of particle swarm optimization, it converges to a higher fitness value than the other three algorithms. On the contrary, the artificial fish algorithm has better global searchability, so a lower fitness value than particle swarm optimization can be achieved. The cuckoo algorithm can increase the variety of solutions, so the fitness value is lower than the artificial fish algorithm, but the convergence speed is slower. FFS algorithm combines the multiple behavior selection of artificial fish swarm algorithm and the advantages of long and short paths in the Cuckoo algorithm. Simulation results show that FFS has better global optimization ability. On the one hand, the FFS algorithm increases the diversity of feasible solutions. On the other hand, it improves accuracy through multiple behavior selection. While ensuring local optimization performance, the global search capability is increased.

To evaluate the performance of FFS, the time complexity is analyzed. The time complexity of particle swarm is $O(ITA*N*D)$. As for artificial fish and Levy optimization of the Cuckoo algorithm, the time complexity is $O(ITA*N*D)$. The time complexity of the foraging behavior mentioned in this paper is $O(ITA*N*D)$. The time complexity of the team behavior is $O((N+1)*ITA)$. The time complexity of track behavior is $O((N+1)*ITA)$. We usually take the highest

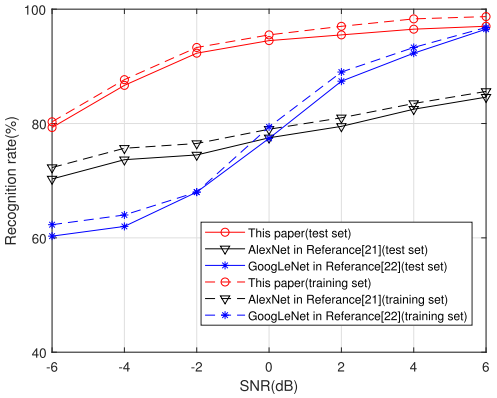


FIGURE 9. The figure shows the recognition rate under different SNR.

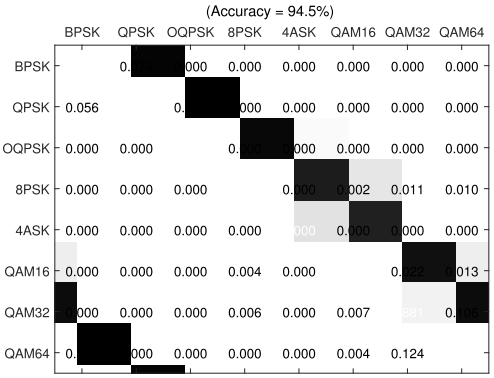


FIGURE 10. This figure shows the confusion matrix of recognition results under 0 dB, which the accuracy is 94.5%.

power item as time complexity, thus the time complexity of the algorithm is $O(ITA * N * D)$, which is the same as the artificial fish and Cuckoo algorithm.

D. EXPERIMENT WITH SNR

Fig.9 shows the recognition rate under different SNR compared with reference [16] and [22]. Reference [16] uses the AlexNet network to perform modulation recognition on gray-scale constellations. Reference [22] uses GoogLeNet to perform modulation recognition on three-channel constellations. The SNR range set in this paper is from −6dB to 6dB. From Fig.9, it can be seen that the recognition rate by the method proposed in this paper is better than that of the modulation recognition in reference [21] and reference [22]. When the SNR is less than 0dB, the recognition rate of reference [22] is lower than that of reference [21]. However, when the SNR is above 0dB, the recognition rate of reference [22] is better than that of reference [21]. Therefore, the modulation recognition method of reference [21] has better performance under low SNR conditions. The modulation recognition method proposed in this paper has a better recognition rate under the low SNR. When SNR is between 2dB and 6dB, it is difficult to improve the overall recognition rate significantly. Because some of the signals are too similar, which is affected by the QAM32 and QAM64 recognition rates. Under the condition of −4dB, the recognition rate of the method proposed in this paper can reach 84.7%, which is 11.1% higher than that of reference [21]. Under the condition of 0dB, the recognition rate of the method proposed in this paper can reach 94.5%, which is 16.4% higher than the recognition rate of reference [21].

Fig.10 shows the confusion matrix of the recognition rate under the condition of SNR of 0dB. As can be seen from the table, the recognition rate of 4ASK and OQPSK is the highest among the eight signals, which is 100%. This result is caused by the low similarity of its other signals. QAM64 has the lowest recognition rate at 87.2%. Because after orthogonal modulation, the spectrum of the CS transform is similar to that of QAM64 and QAM32. The modulation of the two signals is the same, the difference is only in the constellation coordinates. This paper mainly solves the problem of various types of communication signals. To solve such a problem, a secondary classifier needs to be designed, and further classification of QAM is necessary. Besides, for BPSK and QPSK, due to the indistinguishability of the two after CWD transformation, the results of mutual misclassification occurred.

Fig.11 shows the recognition rates of eight modulation recognition signals with different SNR. The recognition rate obtained by the method proposed in this paper is superior to that of reference [21] and reference [22]. For 4ASK, because the modulation method of this signal is significantly different from other signals, the recognition rate of the signal can reach 100% at 0dB. Besides, for OQPSK, because this signal is phase-delayed, and the feature extraction in this paper is comprehensive, its recognition rate can reach 100% at 0dB. However, due to the similarity between QAM32 and QAM64, the recognition rates are slightly lower. It can be seen from the table that under the condition of −4dB, the calculation time of reference [21] is 3.6s, and the calculation time of reference [22] is 6.2s. The calculation time obtained by the method mentioned in this paper is 4.6s. Due to the limited conditions of the experimental simulation equipment used in this paper, the calculation time obtained in this paper is acceptable in the real communication system.

E. EXPERIMENT WITH ROBUSTNESS

Aiming at the problem of training with fewer samples proposed in this paper, this part proves that modulation recognition still has robustness under the condition of fewer samples by comparing the recognition rates under different samples. Fig.12 shows the robustness experience result. Firstly, we test the robustness of the proposed method in a sample range of 200 to 3000. The recognition rates of different sample sizes under −4dB, 0dB, and 4dB conditions are given. When the sample size is 200, the recognition rate of each SNR is lower than that of 400 samples. When the sample size is more than or equal to 400, the recognition rate does not

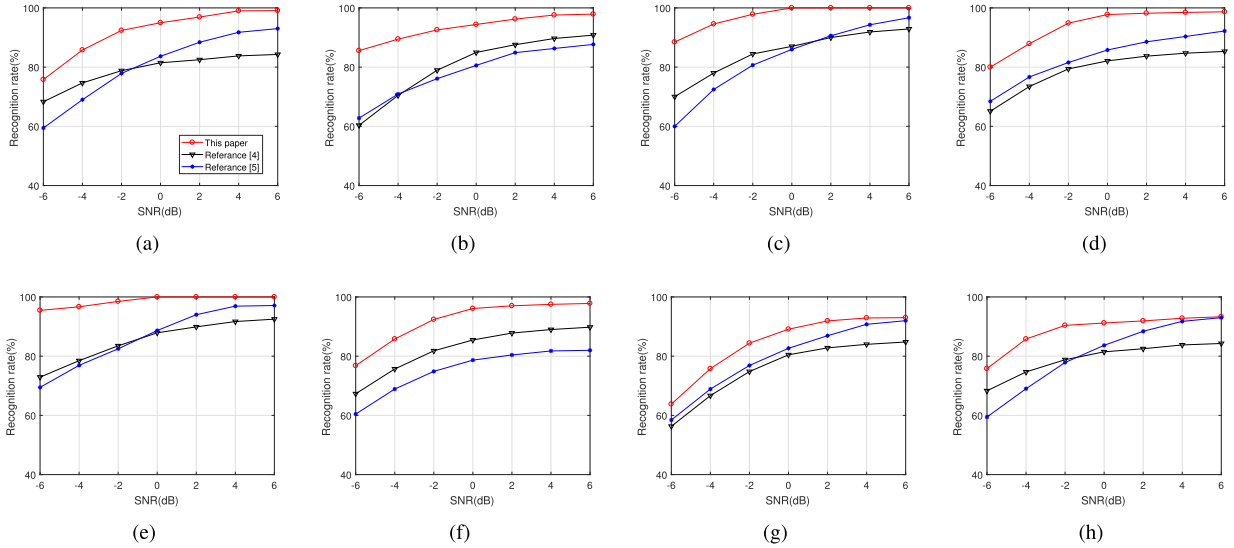


FIGURE 11. The recognition rate of modulation signals are shown in this figure, (a) BPSK, (b) QPSK, (c) OQPSK, (d) 8PSK, (e) 4ASK, (f) QAM16, (g) QAM32, (h) QAM64.

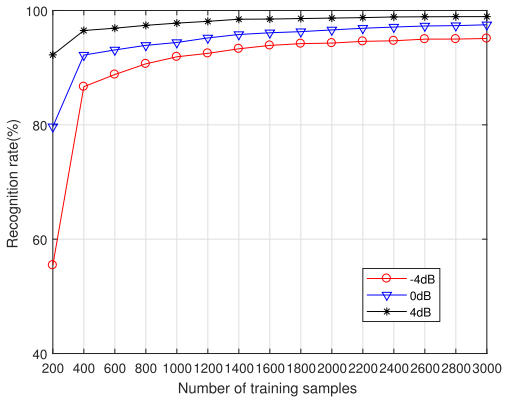


FIGURE 12. The figure shows the robustness experiment result, red line, blue line and black line are −4dB, 0dB and 4dB respectively.

change much. There is little difference between the recognition rate of the sample size 400 and the sample size 3000. The reason is that the modulation recognition method proposed in this paper can more fully characterize the signal sample. Moreover, using CNNS to extract its features can increase the adaptability of the extraction performance. When the SNR is −4dB, because of the influence of noise and other factors, the recognition rate of the sample size 200 is quite different from that of the sample size 400. With the increase of SNR, the recognition rate decreases. When the SNR is 4 dB, the recognition method proposed in this paper has better recognition stability. Therefore, when the sample size is 400, the recognition rate has certain robustness.

VI. CONCLUSION

This paper proposes a new modulation recognition method based on the FFS algorithm, which can recognize eight different modulated signals. The proposed method solves the problem of low accuracy in modulation recognition, which is under the small samples with low SNR conditions. The proposed FFS algorithm can be combined with CNNS to find the optimal solution. It can balance the signal processing methods, the types of networks, the specific layers of the network, the dimensions of features, and the parameters of SVM. The overall recognition rate can reach 94.5% at 0dB. The overall recognition rate can reach 84.7% at −4dB. Moreover, the method has certain theoretical significance and practical application value in other recognition fields under small samples.

Although the algorithm proposed in this study can have better performance, there is still plenty of room for improvement. For example, some parameters of the FFS algorithm can be further optimized. Deep Reinforcement Learning (DRL) has excellent capabilities of exploration and exploitation capabilities. In the future, we plan to use DRL to solve the parameter optimization problem with modulation recognition.

REFERENCES

[1] K. Gençol, A. Kara, and N. At, "Improvements on deinterleaving of radar pulses in dynamically varying signal environments," *Digit. Signal Process.*, vol. 69, pp. 86–93, Oct. 2017.

[2] K. Gençol, N. At, and A. Kara, "A wavelet-based feature set for recognizing pulse repetition interval modulation patterns," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 24, pp. 3078–3090, Apr. 2016.

[3] A. M. Ali, E. Uzundurukan, and A. Kara, "Assessment of features and classifiers for Bluetooth RF fingerprinting," *IEEE Access*, vol. 7, pp. 50524–50535, 2019.

[4] M. K. M. Fadul, D. R. Reising, and M. Sartipi, "Classification of OFDM-based radios under Rayleigh fading using RF-DNA and deep learning," *IEEE Access*, vol. 9, pp. 17100–17113, Jan. 2021.

[5] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2016, pp. 213–226.

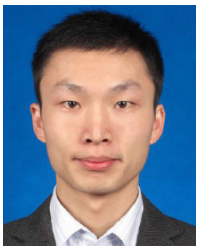
[6] G. Wang, C. Gong, and Z. Xu, "Signal characterization for multiple access non-line of sight scattering communication," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4138–4154, Sep. 2018.

[7] M. Zebarjadi and M. Teimouri, "Non-cooperative burst detection and synchronisation in downlink TDMA-based wireless communication networks," *IET Commun.*, vol. 13, no. 7, pp. 863–872, Apr. 2019.
[8] M. Zhang, M. Diao, and L. Guo, "Convolutional neural networks for automatic cognitive radio waveform recognition," *IEEE Access*, vol. 5, pp. 11074–11082, Jun. 2017.
[9] M. C. M. M. Souza, A. Grieco, N. C. Frateschi, and Y. Fainman, "Fourier transform spectrometer on silicon with thermo-optic non-linearity and dispersion correction," *Nature Commun.*, vol. 9, no. 1, p. 665, Feb. 2018.
[10] C. Mateo and J. A. Talavera, "Short-time Fourier transform with the window size fixed in the frequency domain," *Digit. Signal Process.*, vol. 77, pp. 13–21, Jun. 2018.
[11] Z.-C. Zhang, "Choi–Williams distribution in linear canonical domains and its application in noisy LFM signals detection," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 82, Mar. 2020, Art. no. 105025.
[12] S. Cao and Z. Weiyan, "Carrier frequency and symbol rate estimation based on cyclic spectrum," *J. Syst. Eng. Electron.*, vol. 31, no. 1, pp. 37–44, Jan. 2020.
[13] T. Zhang, C. Shuai, and Y. Zhou, "Deep learning for robust automatic modulation recognition method for IoT applications," *IEEE Access*, vol. 8, pp. 117689–117697, 2020.
[14] Y. Zhang, D. Liu, J. Liu, Y. Xian, and X. Wang, "Improved deep neural network for OFDM signal recognition using hybrid grey wolf optimization," *IEEE Access*, vol. 8, pp. 133622–133632, 2020.
[15] I. Hammad and K. El-Sankary, "Impact of approximate multipliers on VGG deep learning network," *IEEE Access*, vol. 6, pp. 60438–60444, 2018.
[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.
[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[18] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.
[19] A. Van Opbroek, H. C. Achterberg, M. W. Vernooij, and M. De Bruijne, "Transfer learning for image segmentation by combining image weighting and kernel learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 213–224, Jan. 2019.
[20] M. Hashemian, F. Razzazi, H. Zarrabi, and M. S. Moin, "A privacy-preserving distributed transfer learning in activity recognition," *Telecommun. Syst.*, vol. 72, no. 1, pp. 69–79, Sep. 2019.
[21] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y.-D. Yao, "Modulation classification using convolutional neural network based deep learning model," in *Proc. 26th Wireless Opt. Commun. Conf. (WOCC)*, Newark, NJ, USA, Apr. 2017, pp. 1–5.
[22] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 718–727, Mar. 2019.
[23] T. Luo, C. Hou, F. Nie, and D. Yi, "Dimension reduction for non-Gaussian data by adaptive discriminative analysis," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 933–946, Mar. 2019.
[24] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
[25] C. Yang, Z. He, Y. Peng, Y. Wang, and J. Yang, "Deep learning aided method for automatic modulation recognition," *IEEE Access*, vol. 7, pp. 109063–109068, Aug. 2019.
[26] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
[27] S. Routray, A. K. Ray, C. Mishra, and G. Palai, "Efficient hybrid image denoising scheme based on SVM classification," *Optik*, vol. 157, pp. 503–511, Mar. 2018.
[28] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, Apr. 2012.
[29] S. Jain, V. Bajaj, and A. Kumar, "Efficient algorithm for classification of electrocardiogram beats based on artificial bee colony-based least-squares support vector machines classifier," *Electron. Lett.*, vol. 52, no. 14, pp. 1198–1200, Jul. 2016.
[30] W. Liang, L. Zhang, and M. Wang, "The chaos differential evolution optimization algorithm and its application to support vector regression machine," *J. Softw.*, vol. 6, no. 7, pp. 1297–1304, Jul. 2011.
[31] W. Li and S. Gao, "Prospective on energy related carbon emissions peak integrating optimized intelligent algorithm with dry process technique application for China's cement industry," *Energy*, vol. 165, pp. 33–54, Dec. 2018.
[32] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, "Artificial fish swarm algorithm: A survey of the state-of-the-art, hybridization, combinatorial and indicative applications," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 965–997, Dec. 2014.
[33] Y. Feng, S. Zhao, and H. Liu, "Analysis of network coverage optimization based on feedback K-means clustering and artificial fish swarm algorithm," *IEEE Access*, vol. 8, pp. 42864–42876, Jan. 2020.
[34] A. Ouaarab, B. Ahiod, and X.-S. Yang, "Discrete cuckoo search algorithm for the travelling salesman problem," *Neural Comput. Appl.*, vol. 24, nos. 7–8, pp. 1659–1669, Jun. 2014.
[35] T. A. Almohamad, M. F. M. Salleh, M. N. Mahmud, I. R. Karas, N. S. M. Shah, and S. A. Al-Gailani, "Dual-determination of modulation types and signal-to-noise ratios using 2D-ASIQH features for next generation of wireless communication systems," *IEEE Access*, vol. 9, pp. 25843–25857, Feb. 2021.
[36] D. Vučić, S. Vukotić, and M. Erić, "Cyclic spectral analysis of OFDM/OQAM signals," *AEU-Int. J. Electron. Commun.*, vol. 73, pp. 139–143, Mar. 2017.
[37] S. Alam, M. Kang, J.-Y. Pyun, and G.-R. Kwon, "Performance of classification based on PCA, linear SVM, and multi-kernel SVM," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Vienna, Austria, Jul. 2016, pp. 987–989.
[38] S. Aly and A. Mohamed, "Unknown-length handwritten numeral string recognition using cascade of PCA-SVMNet classifiers," *IEEE Access*, vol. 7, pp. 52024–52034, Apr. 2019.
[39] A. Zare, A. Ozdemir, M. A. Iwen, and S. Aviyente, "Extension of PCA to higher order data structures: An introduction to tensors, tensor decompositions, and tensor PCA," *Proc. IEEE*, vol. 106, no. 8, pp. 1341–1358, Aug. 2018.
[40] S. Lipovetsky, "PCA and SVD with nonnegative loadings," *Pattern Recognit.*, vol. 42, no. 1, pp. 68–76, Jan. 2009.
[41] X. Lei, X. Yang, and F.-X. Wu, "Artificial fish swarm optimization based method to identify essential proteins," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 2, pp. 495–505, Apr. 2020.
[42] Y. Liu and B. Cao, "A novel ant colony optimization algorithm with Levy flight," *IEEE Access*, vol. 8, pp. 67205–67213, Apr. 2020.

JINGPENG GAO received the B.S., M.S., and Ph.D. degrees in electrical information engineering from Harbin Engineering University (HEU), China, in 2002, 2007, and 2014, respectively. He has been a Lecturer with Harbin Engineering University, since 2002, became a Lecturer, in 2007, and a Master Tutor, in 2015. From 2015 to 2017, he stayed with the State Key Laboratory of Computational Mathematical and Experimental Physics, Beijing Institute of Space Long March Vehicle, as a Postdoctoral Researcher. His research interests include electronic countermeasure, machine learning, radar target recognition, and signal detection. He was a recipient of the first and third prizes of teaching instruments and equipment in China.

XU WANG received the B.S. degree in communication engineering from Anhui University, China, in 2019. She is currently pursuing the M.S. degree in electronics and communication engineering from Harbin Engineering University, China. Her research interests include signal processing, radar, machine learning, deep learning, and image processing.

RUOWU WU received the master's degree from the University of Electronic Science and Technology of China, in 2013. His main research interest includes characteristics and simulation of complex electromagnetic environments.



XIONG XU received the B.S. degree in electronic information science and technology and the Ph.D. degree in physical electronic from the University of Electronic Science and Technology of China, Chengdu, in 2006 and 2012, respectively. From 2013 to 2015, he was a Postdoctoral Research Fellow with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), where he is currently a Research Associate. His research interests include data processing and machine learning.