D-GF-CNN Algorithm for Modulation Recognition

Ruiyan Du1,2 · Fulai Liu1,2 · Jialiang Xu2 · Fan Gao2 · Zhongyi Hu2 · Aiyi Zhang2

Accepted: 14 November 2021 / Published online: 3 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
This paper presents a novel modulation recognition algorithm based on dilated convolutional neural network with a new defined GF regularization function, named as D-GF-CNN algorithm. Firstly, an asynchronous delay sampling (ADS) technique is introduced. Via the defined ADS, the received signal is converted into an asynchronous delay histogram (ADH). The ADH of different modulation signals has distinct characteristics, which provides great convenience for the neural network to identify the modulation mode. Then, the pixel point matrix of the ADH is convolved with the dilated convolution kernel of the convolutional neural network, and the automatic extraction of signal features is completed so that the manual feature extraction processing can be effectively avoided. Finally, a novel GF regularization function is given, which can improve the constraint ability of the loss function on the weight and effectively weaken the influence of network over-fitting on the modulation recognition accuracy. Theoretical analysis and simulation experiments show that the proposed algorithm provides several advantages, for example: (1) automatically extract features; (2) effectively prevent network over-fitting; (3) significantly improve recognition accuracy in the lower SNR scenarios.

Keywords Modulation recognition · Asynchronous delay sampling · Convolutional neural networks · Dilated convolution · Regularization

1 Introduction

Modulation recognition, as an important technology of wireless communication, is widely used in military and civil domains. In military applications, it can be employed for signal confirmation, spectrum detection, emitter interception and so on. The wide range of
civilian applications includes signal verification, cognitive radio, interference identification, etc. In 1969, modulation recognition technology is first proposed in [1], which presents the preliminary results of automatic modulation recognition on a high-frequency radio signal. After that, modulation recognition is extensively studied in different ways.

Generally, modulation recognition can be categorized into two main approaches, including Likelihood-Based algorithm (LB-based) and Feature-Based method (FB-based). In [2], a Likelihood-Ratio function and approximations thereof are proposed for the problem of classifying multiple phase shift keying (MPSK) modulations in additive white Gaussian noise. On the basis of Likelihood-Ratio function, a maximum-likelihood (ML) is applied to the classification of digital quadrature modulations and obtains a generic formula for the error probability of the ML classifier [3]. In [4], ML is exploited to design General Maximum Likelihood Classifier (GMLC), which doesn’t need any restriction on the baseband pulse and provides a general theoretical framework for many pattern recognition. However, the GMLC has less accurate for the classification of non-constant envelope modulations. To solve this problem, Hybrid Likelihood Ratio Test (HLRT) is proposed [5], this method can achieve significant performance gains for the classification of non-constant envelope modulations. It is clear that the above LB-based method has high computational complexity and unsatisfactory robustness. Moreover, when the signal to noise ratio (SNR) decreases, the classification performance will decline sharply. With this backdrop, several FB-based methods have been proposed in recent years. For example, a method based on elementary fourth-order cumulants is proposed for the classification of digital modulation schemes [6]. The approach is robust in the presence of carrier phase and frequency offsets, as well as impulsive non-Gaussian noise. Instead of fourth-order cumulants, higher-order cumulants are used to recognize the modulation types of transmitted signals and prove the feasibility of multiuser modulation classification [7]. In addition, a feature-based blind modulation classification based on the combination of elementary cumulant (EC) and cyclic cumulants is proposed. It has better classification performance over a Rayleigh flat fading channel compared with the other existing methods [8]. However, most cumulant-based classifiers have interference, which affects classification performance. After that, a method based on support vector machine (SVM) is proposed, which is robust in the classification of amplitude modulation (AM), binary phase shift keying (BPSK), quadrature reference phase shift keying (QPSK) and binary frequency shift keying (BFSK), for a large SNR range and in the presence of a multipath fading channel [9]. In order to further improve classification accuracy, Genetic Programming (GP) and K-nearest neighbor (KNN) are combined in [10], the method is robust for a large SNR range and in the presence of a multipath fading channel and improves classification accuracy. However, the FB-based method needs a large number of manual feature extraction, and the representativeness of manual feature extraction directly determines the final classification performance.

Deep learning (DL) is an effective technique to extract various complex features from the original data automatically. Thus DL is widely used in computer vision, image recognition and modulation recognition. OShea applies deep learning to modulation recognition firstly and discusses the critical importance of good datasets for model learning, testing, and evaluation [11]. Mendis employs a deep belief network (DBN) to classify the modulation techniques by learning their features from the associated spectral correlation function (SCF) patterns [12]. Compare with the spectral correlation function, the deep neural network (DNN) based on high-order cumulants is employed to improve the recognition rate, and the overall success rate of the method is over 99 % at the SNR of -2dB [13]. A convolutional neural network (CNN) based on constellation diagrams is designed to recognize modulation modes that are difficult to distinguish such as 16 quadratic-amplitude
modulation (16QAM) and 64 quadratic-amplitude modulation (64QAM) [14]. However, the recognition rate of the existing DL methods such as [11]–[14] degrades sharply in the lower SNR environments. Specifically, when SNR is less than -10 dB, the recognition rate is less than 30% or even less than 10%.

In order to improve the recognition rate of modulation recognition technology at low SNR. In this paper, a novel D-GF-CNN algorithm is proposed. The main contributions can be summarized as follows:

(1) An ADS technique is introduced for data preprocessing in this paper. The ADS can not only avoid the manual feature extraction, but also capture changes in the amplitude, frequency, and phase of the modulated signal. Therefore, the influence of channel noise and channel time variation on the recognition result can be effectively suppressed. Furthermore, the 1-D modulated signal is converted to the 2-D ADH through ADS, which is used as the input data to the subsequent network.

(2) To solve the problem of low recognition rate caused by network fitting noise under low SNR, the dilated convolution layers are used instead of ordinary convolution layers in the proposed algorithm. It is clear that the weight parameters of the network can be reduced through dilated convolution layers. Therefore, the fitting ability of neural network to noise is weakened.

(3) A new GF regularization function is defined in this paper, which can improve the generalization ability of network. The new GF regularization function is used as the penalty term for the back propagation of network, so that the neural network can not only avoid the over-fitting problem, but also maintain acceptable convergence. Furthermore, an optimal range of weights is given by theoretical derivation for the proposed D-GF-CNN network, which can reduce the over-fitting of the network more quickly in this interval.

2 Data Model

According to [15], the received signal can be given by

\[ r(t) = s(t; \mathbf{u}_i) + n(t) \]  

where \( s(t) \) is the modulated signal. \( \mathbf{u}_i \) represents the vector of unknown quantities on which the noise-free baseband complex envelope depends. \( n(t) \) means the additive white gaussian noise (AWGN).

The expression of \( s(t; \mathbf{u}_i) \) can be written as

\[ s(t; \mathbf{u}_i) = A e^{j(\omega t + \theta)} \sum_{k=0}^{K-1} x_k^{i,j} g(t - kT_s - \epsilon T_s) \]  

where \( \mathbf{u}_i \) is defined as a multi-dimensional parameter set, in which the unknown signal and channel variables are deterministic or stochastic under the \( i-th \) modulation scheme, \( \mathbf{u}_i = \{A, \omega, \theta, \epsilon, T, \{x_k^{i,l}\}_{l=1}^{M_k}\} \). \( A \) represents the signal amplitude. \( \omega \) is the residual carrier angular frequency after carrier removal. \( \theta \) denotes the fixed phase offset introduced by the propagation delay as well as the initial carrier phase. \( \epsilon \) stands for the normalized epoch for timing offset between the transmitter and the receiver. \( T \) represents symbol interval. \( x_k^{i,l} \) is the \( i-th \) constellation point under \( l-th \) modulation scheme where the symbols are equally distributed, \( i \in \{1, \ldots, M_k\}, k \in \{1, \ldots, C\} \). The mark \( M_k \) denotes the symbol number using this modulation, and \( C \) is the number of candidate modulation schemes. \( g(t) \)
denotes the composite effect of the residual channel $h(t)$ and the pulse-shaping function $p(t)$. It is expressed by the equation $g(t) = h(t) \ast p(t)$, where $\ast$ is the convolution operator.

### 3 Algorithm Formulation

In this section, a new network structure called D-GF-CNN is proposed to solve the modulation recognition problem. Firstly, the 1-D modulated signal is converted to the 2-D ADH through the ADS technique, which can effectively avoid the manual feature extraction. Then, a novel CNN structure with the dilated convolution kernel for modulation recognition is developed, which can reduce the weight parameters of the network. Finally, to improve the generalization ability of network, a new GF regularization function is defined.

#### 3.1 Data Preprocessing

Considering that the modulation recognition technology based on the traditional statistical mode method relies on feature extraction, which may lead to the decrease of recognition accuracy. An ADS is used to data preprocessing in the proposed algorithm. Since the ADS does not need statistical signal features, it can effectively reduce the dependence of the recognition algorithm on feature extraction and avoid the complex process of feature extraction.

The ADS uses two adjacent samples in each sampling period to obtain modulated signal features, such as amplitude, frequency and phase [16]. Compare with synchronous sampling, asynchronous sampling does not require prior knowledge of the modulated signal and can acquire more signal information [17]. In the ADS, the signal envelope is sampled in pairs $(p_k, q_k)$ and the sampling rate is lower than synchronous sampling [18]. Both of these features are ideal because they can reduce the cost of implementation. The sampling process is shown in Fig. 1.

![Asynchronous delay sampling of FSK](image)

**Fig. 1** Asynchronous delay sampling of FSK
As shown in Fig. 1, the asynchronous delay sampling is used in the frequency-shift keying (FSK) signal with period $T$. The abscissa represents the time and the ordinate represents the amplitude of signal. Two sample points $p_k$ and $q_k$ are taken in each sampling period $T_{\text{sampling}}$, and $T_{\text{sampling}}$ is independent of the signal period $T$. When the signal sampling process is completed, two sample sequences $p_k$ and $q_k$ are obtained. This process is given by

$$p_k = S(t_1 + (k-1)T_{\text{sampling}})$$

$$q_k = S(t_1 + \tau + (k-1)T_{\text{sampling}})$$

(3)

where $S(\cdot)$ represents a modulated signal series, $t_1$ stands for the time of the first sampling point, $k \in 1, 2, \ldots, N$, $N$ is positive integer, $T_{\text{sampling}}$ denotes the sampling interval of adjacent $p$ points, $\tau$ stands for the sampling interval between adjacent $p$ and $q$ points.

Using the sequence $p_k$ as the abscissa data and the sequence $q_k$ as the ordinate data, the 2-D ADH is obtained. As shown in Fig. 2, the ADH clearly reflects the amplitude, frequency and phase of the signal and effectively suppress the influence of channel noise and channel time variation on the recognition result.

### 3.2 Network Architecture

In this subsection, a novel network D-GF-CNN is proposed to identify the 2-D ADH of the modulated signal. Figure 3 shows the D-GF CNN structure. The network consists of convolution layers, activation layers, pooling layers, and fully connected layers. The data processing for each layer of the network is described as follow.

#### 3.2.1 Convolution Layers

To effectively capture 2-D ADH information and increase the receptive field size, the dilated convolution is adopted in convolution layers. The ADH is characterized that the ADS scatter has concentrated in the diagonal from the lower left corner to the upper right.
corner. Using traditional convolution kernels will add many background features during the convolution process, which will increase the parameters of the network and the amount of computation.

Dilated convolution is a convolution method which can reduce the resolution of images and information lose in the downsampling [19] [20]. By adding holes in the standard convolution kernel, that is, adding 0 to the position of the hole, makes the receptive field of convolution operation become larger without increasing the computational complexity. This property of dilated convolution solves the problem of data redundancy in the ADS. Using dilated convolution can extract useful feature information in the ADH, reduce the network parameters and computing costs.

According to [19], a 2-D discrete convolution operator $\ast$, which convolves signal $x$ with kernel $\omega$ of size $(2m + 1) \times (2m + 1)$, is defined as

$$ (x \ast \omega)(p) = \sum_{s+t=p} x(s)\omega(t) $$

where $p, s \in \mathbb{Z}^2$ and $t \in [-m, m]^2 \cap \mathbb{Z}^2$ with $\mathbb{Z}$ denotes the set of integers. A dilated version of the operator $\ast$, which is denoted by $\ast_l$, can be written as

$$ (x \ast_l \omega)(p) = \sum_{s+l+t=p} x(s)\omega(t) $$

where $l \in \mathbb{Z}^+$ is the dilation factor. Note that conventional convolutions can be regarded as 1-dilated convolutions.

### 3.2.2 Activation Layers

In order to increase the nonlinearity of the network, the ReLU activation function is introduced in this subsection. The ReLU activation function can be expressed as

$$ y = \begin{cases} 
  x, & \text{if } x \geq 0 \\
  0, & \text{if } x < 0
\end{cases} $$

where $x$ is input data after BN.

### 3.2.3 Pooling Layers

Pooling layer can compress data and reduce network parameters so that avoid over-fitting problem. In this paper, max pooling is adopted to get the maximum value in the adjacent rectangular region. Max pooling can be written as

\[ y = \max_{i,j \in [1,n]} x(i,j) \]
where $p_{m,n}$ indicates the output of the pooling layer. $y_{m,n}$ represents the pooling areas corresponding to $p_{m,n}$, $a_{ij}$ represents the activation value on the $y_{m,n}$, and $i,j$ is the index of $a_{ij}$ on the $y_{m,n}$.

### 3.2.4 Fully Connected Layers

The fully connected layer appears at the end of the network structure. The output of the fully connected layers is data features, and Softmax classifier outputs the probability of the corresponding category based on these features. The forward propagation for the full connection layer can be expressed as

$$z = f\left(\sum_{i=1}^{n} w_i p + b\right)$$

where $p$ represents the input of the neuron. $w$ is the weight. $z$ stands for the output of the neuron. $b$ denotes the bias of the neuron. $f$ is ReLU activation function.

In the training process of neural network, the loss function is the evaluation standard of the network model. The loss function can be written as follows

$$J(\theta;X,y) = L(\theta;X,y) + \Omega(\theta)$$

where $L(\theta;X,y)$ is cross entropy loss function, the expression is given by

$$L(\theta;X,y) = -\frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{e^{h_{y_i}}}{\sum_{j=1}^{C} e^{h_j}} \right)$$

where $N$ represents number of training samples, $h_{y_i}$ stands for output of the network, $h_j$ denotes real label, $\theta$ is weight of the network. $\Omega(\theta)$ represents the regularization, which will be described in detail in the next subsection.

For briefly, the parameters of the D-GF-CNN network are shown in Table 1. It will be shown in Section 4 that such a network can achieve excellent performance in the modulation recognition system.

### 3.3 Regularization

To reduce the over-fitting of network, a GF regularization function is proposed in this subsection. The proposed GF regularization function can constrain weight parameters $w$ in the network to a smaller range, thus prevent network over-fitting.

**Definition 1** For the weight parameters $w$ during network training, the GF regularization function $\Omega(\omega)$ is defined as follows

$$\Omega(\omega) = \frac{\omega e^\omega - \omega}{e^\omega + 1}$$

Based on the above definition, the properties of $\Omega(\omega)$ can be proved as follow.
Theorem 1  Assume that \( L(\omega;X,y) \) is the original loss function of network. Introduce the \( \Omega(\omega) \) into the loss function then \( \tilde{J}(\omega;X,y) = L(\omega;X,y) + \frac{\lambda}{n} \Omega(\omega) \). \( \omega \) are constantly trained by \( \tilde{J}(\omega;X,y) \). Through the processing of \( \Omega(\omega) \), \( \omega \) is constrained to a smaller range. The network consisting of smaller weights has stronger generalization performance and prevents over-fitting.

Proof  In the training stage, the parameter of network is gradually reduced with gradient descent method until the network weight is stable. This process is given by

\[
\omega \rightarrow \omega - \eta \frac{\partial L(\omega;X,y)}{\partial \omega} \tag{12}
\]

After adding the \( \Omega(\omega) \) function, the network weight update process can be rewritten as

\[
\omega \rightarrow \omega - \eta \frac{\partial L(\omega;X,y)}{\partial \omega} - \frac{\eta \lambda}{n} \frac{\partial \Omega(\omega)}{\partial \omega} \tag{13}
\]

where \( \lambda \) is regularization rate, \( n \) is number of samples.

Since (12) is the process of updating \( \omega \) with gradient descent algorithm and this process determines whether the network is over-fitting. From (13) the following equation can be obtained

| Table 1  Network parameters                  | Layer                      | output                  |
|------------------------------------------------|----------------------------|-------------------------|
| Input                                          |                            | 90 x 90 x 3            |
| Dilated Conv2D (filters=96, size=3x3, dilation-rate=3) | +ReLU                      | 84 x 84 x 96           |
| BatchNormalization                             |                            |                         |
| MaxPooling(size=2,stride=2)                    |                            |                         |
| Dilated Conv2D (filters=128, size=3x3, dilation-rate=3) | +ReLU                      | 36 x 36 x 128          |
| Tensor                                         |                            |                         |
| Dense+ReLU+GF regularization                   |                            |                         |
| Dense+ReLU+GF regularization                   |                            |                         |
| Dense+ReLU+GF regularization                   |                            |                         |
| Softmax classification                         |                            |                         |

\( \otimes \) Springer
\[
\omega \rightarrow \omega - \frac{\eta \lambda}{n} \frac{\partial \Omega(\omega)}{\partial \omega} \\
= \omega - \frac{\eta \lambda}{n} \left( \frac{\partial \omega}{\partial \omega} \right) \\
= \omega - \frac{\eta \lambda}{n} e^{2\omega} + 2\omega e^{\omega} - 1 \\
= \omega \left(1 - \frac{\eta \lambda}{n} e^{2\omega} + 2\omega e^{\omega} - 1 \right)
\]  

(14)

To facilitate theoretical derivation in (14), let \(\frac{\eta \lambda}{n} = 1\), and \(g(\omega) = \frac{\partial \Omega(\omega)}{\partial \omega} = \frac{e^{\omega} + 2\omega e^{\omega} - 1}{\omega(\omega + 1)^2}\). Then (14) is written as

\[
\omega (1 - g(\omega))
\]

(15)

It can be seen from (15) that the decrease of \(\omega\) is related to \(g(\omega)\). Calculating the derivative for \(g(\omega)\). The result can be represented as

\[
\begin{cases}
g'(\omega) > 0 & \omega < 0 \\
g'(\omega) < 0 & \omega > 0
\end{cases}
\]

(16)

However, it can be seen that when \(\omega=0\), \(g(\omega)\) has no meaning, so the limit is taken for \(g(0)\), and the process can be expressed as

\[
\lim_{\omega \to 0} g(\omega) = \lim_{\omega \to 0} \frac{e^{\omega} + 2\omega e^{\omega} - 1}{\omega(\omega + 1)^2} = 1
\]

(17)

According to (16) and (17), when \(\omega \to 0\), \(\lim_{\omega \to 0} g(\omega) = 1\), which explained that \(\omega = 0\) is the upper bound of \(g(\omega)\). So \(g(\omega) < 1\), and in practice, \(\frac{\eta \lambda}{n} < 1\), then \(1 - g(\omega)\) can be written as following

\[
1 - g(\omega) = 1 - \frac{\eta \lambda}{n} \frac{e^{2\omega} + 2\omega e^{\omega} - 1}{\omega(\omega + 1)^2} < 1
\]

(18)

For the neural network without GF regularization, the coefficient of \(\omega\) is 1 in the derivative result. It is worth noting that when the network adopts GF regularization, the coefficient of \(\omega\) is \(1 - g(\omega)\), which reduces \(\omega\). In the gradient descent method, \(\omega\) is related not only to itself, but also to \(1 - g(\omega)\). Thus \(1 - g(\omega) < 1\) reduces the sensitivity of the network, prevents over-fitting of the network and achieves the purpose of regularization.

This concludes the proof. 

After proving the principle of GF regularization function, the properties of its derivative function will be discussed as follows. From gradient perspective, let \(h(\omega) = 1 - g(\omega)\).

As can be seen from Fig. 4, the slope of \(h(\omega)\) represents the transformation rate of \(\omega\). So in \((-\infty, +\infty)\), there is an interval in which the transformation rate of \(\omega\) is faster than the point outside the interval. So the significance of this interval is that \(\omega\) is faster, and it is less likely to over-fitting the network, so the next step is to find this interval.

Firstly, calculate the derivative of \(h(\omega)\) to get \(h'(\omega)\). As can be seen from Fig. 5, when \(\omega > 0\), \(h'(\omega)\) has a maximum value in \((0, 2)\), and \(\omega\) is smaller than this maximum, transformation rate is fast. In the same way, when \(\omega < 0\), \(h'(\omega)\) has a minimum value in \((-2, 0)\). In order to find this value, calculate the derivative of \(h'(\omega)\) to get \(h''(\omega)\).
As can be seen from Fig. 6, when $h''(\omega) = 0$, $\omega \approx \pm 1.42$, $h'(\omega)$ takes the maximum and minimum. So $\omega \approx \pm 1.42$ is the inflection point of $h(\omega)$. When $\omega \in (-1.42, 1.42)$, the convergence speed of the loss function can be closer to 0, and the purpose of regularization is to reduce the weight parameter in the network, and solve the problem of over-fitting. Therefore, when the GF regularization function is used in the network, let $-1.42 < \omega < 1.42$, which can further prevent overfitting.
Summary of the Proposed Method

The procedure of the proposed D-GF-CNN method can be summarized as follows.

**Step 1** In data preprocessing, the ADS technique is adopted and according to (3), modulated signals are mapped into ADH with $p_k$ as the abscissa and $q_k$ as the ordinate.

**Step 2** The ADH is taken as the input of network. According to (4)∼(5), the redundant information of ADH is removed by using dilated convolution.

**Step 3** According to (6)∼(10), processed signals in step 2 are respectively treated through pooling layers and fully connected layers.

**Step 4** According to (11)∼(18), the GF regularization function is used to optimize the weight $\omega$ of network, prevent over-fitting of the network and obtain better classification results.

4 Simulation Results

In this section, several simulation results are provided to illustrate the performance of D-GF-CNN algorithm. The modulated signal data set used in this paper is generated by matlab simulation. The data set contains 11 modulation types: 2ASK, 2PSK, 2FSK, 4PSK, AM, SSB, DSB, VSB, FM, 16QAM, 64QAM. The signal symbol is generated randomly by matlab, and the noise of signal is AWGN. The signal source is generated randomly and the baseband signal is obtained after spread spectrum [23] [24]. The range of signal to noise ratio(SNR) is -20 dB to 10 dB, and interval is 1 dB. Each modulated signal generates 40 samples at each SNR, that is, a total of 88000 samples. Among them, 70400 samples are used to train the network model and 17600 samples are used to test the model. 500 Monte Carlo experiments are carried out based on the above simulation conditions. The software environment used in the training model is Python 3.6 and Tensorflow-gpu 1.3.0. The hardware environment is CPU Intel® Xeon® E5-2560, GPU NVIDIA Tesla K80, RAM 64GB.
Figure 7 shows the ADH of 11 modulated signals without noise interference. It can be seen that the ADH of different modulated signals have obvious differences. For example, in the ADH of analog modulated signal, the SSB is a simple ellipse, AM is an ellipse that expands from center to periphery, DSB is an ellipse with two intersecting centers, FM is an ellipse with a right angle to the diagonal, and VSB is a scatter that spreads around the center. In the ADH of digitally modulated signal, 2FSK is a diagonal of four arcs, 2ASK is an ellipse and there are 4 independent points on the upper right of the image. 2PSK is similar to 2ASK, four independent points are distributed in the four corners of the image. 16QAM distributes some scatter points in an ellipse, and 64QAM has dense scatter points in an ellipse. The different characteristics of ADH are the basis for subsequent network processing.
Figure 8 shows the ADH of the modulated signal at different SNR. From the first line to the third line, the modulation type of the signal are 2FSK, AM, and 16QAM, respectively. The SNR of the first ADH in each row is -10dB, the SNR of the second ADH is 0 dB, and the SNR of the third ADH is 10 dB. It can be seen that with the increase of SNR, the characteristics of each modulation signal ADH are more obvious. For example, with the increase of SNR, the ADS histogram of AM signal gradually changes from circular to elliptical. When SNR is fixed, the ADH of different signals have significant differences. For example, when SNR is 0 dB, the 2FSK signal is a rounded rectangle, the AM signal is an ellipse, and the 16QAM signal is a scatter with uneven distribution. The different characteristics of different ADH provide the basis for network identification of modulation types.

Figure 9 portrays the influence curve of the proposed GF regularization function on network loss when \( \text{SNR} = 0 \text{dB} \). X-axis is iteration and Y-axis means network loss. It can be seen that the neural network with the GF regularization function can remain stable even if it is iterated 1000 times. In contrast, the neural network without GF function has been overfitting after iterative training for about 100 times. Such a phenomenon clearly demonstrates the proposed regularization algorithm can effectively prevent over-fitting of network.

Figure 10 shows the relationship between iteration and recognition rate in the case of a test data set with SNR of -10 dB, -5 dB, 0 dB, 5 dB and 10 dB in the proposed network.
X-axis is iteration and Y-axis is recognition rate. It can be seen that under the same SNR, with increase of the iteration, recognition rate of test data sets increases gradually and tends to be stable, and there is no over-fitting, which verifies the proposed regularization algorithm can effectively prevent over-fitting of network. And with the increase of SNR, the recognition accuracy of test data set is improved.

Figure 11 shows the relationship between the iteration and network loss in the case of a test data set with SNR of $-10\,\text{dB}$, $-5\,\text{dB}$, $0\,\text{dB}$, $5\,\text{dB}$ and $10\,\text{dB}$ in the proposed network. It can be seen from the figure that with increase of the number of iterations, network loss decreases gradually and tends to be stable. When the number of iterations is constant, the larger SNR, the faster the network loss reaches the stable value.

Figure 12 shows the relationship between the dilated rate and the network recognition accuracy under five different SNR. The abscissa is the dilated rate, and the ordinate...
is the recognition accuracy. The five curves in the figure represent the recognition rate at −10 dB, −5 dB, 0 dB, 5 dB, 10 dB respectively. It can be seen that with the increase of the dilated rate, the recognition accuracy of network increases at first and then decreases. When the dilated rate is 3, the recognition accuracy of the network is the highest. Therefore, in the proposed network, the convolution kernel with a dilated rate of 3 is used. And as can be seen from the figure, with the decrease of SNR, the bending degree of different SNR curves is different. The lower SNR, the greater the bending
degree of the curve. It is proved that for the modulation signal with low SNR, the recognition rate can be further improved by dilated convolution.

Figure 13 shows the confusion matrix under the test data set of $-10 \text{ dB}, -5 \text{ dB}, 0 \text{ dB}, 5 \text{ dB}, 10 \text{ dB}, 15 \text{ dB}$. X-axis is predict labels and Y-axis is real labels. It is evident that when SNR is low, signal is submerged in noise, and predicted results fluctuates around

![Confusion Matrix in different SNR](image)

**Fig. 13** Confusion matrix in different SNR
the diagonal of figure, but the fluctuation is not very large. With the increase of SNR, the accuracy of recognition gradually increases, and the number of points except diagonal lines decreases gradually in figure. Finally, the complete recognition is achieved, there are only elements in the diagonal position in figure.

Figure 14 shows the recognition rate curve between the proposed method and the existing modulation recognition algorithms under different SNR. X-axis is SNR, range is -20dB ∼ 12dB, and Y-axis is recognition rate. Based on the 11 modulation signal simulation parameters mentioned, this paper compare the proposed method with the several existing methods. In GLRT [25], assuming that the prior probabilities of the eleven modulated signals are the same, the threshold of the classified classification is set to zero, and the probability of correct classification is the average of 1000 independent experiments. In HOC [6] based classifiers, the mean and variance of the modulating signal statistic are calculated in each Monte Carlo experiment. The fourth-order cumulants and the optimal threshold of the signal are calculated by mean and variance, which are compared and identified. In GP-KNN [10], the number of generations used in each run of GP is 100 while the number of individuals in each generation is 25. The genetic operator crossover has a probability of 90% and the probability of mutation is 10%. Each value of SNR and number of samples, 10,000 realization are produced. These 10,000 realizations are tested with the best tree and results are summarized. In LSVM, the amplitude, phase, real part, and imaginary part of the signal are respectively calculated to identify the modulated signal. In BPNN [26], a two layers neural network is adopted and 50 nodes are used in the hidden layer. It can be observed that as SNR increases, the recognition rate of the proposed method is increasing. When SNR exceeds 5 dB, the accuracy of proposed method can reach up to more than 90%. Compared with other existing algorithms, the proposed method has a higher recognition rate at −20 dB ∼ 2 dB, and demonstrates a nearly 10% improvement. Therefore, the proposed method has a better recognition rate in the low SNR environment.

![Classification accuracy in different SNR](image)

**Fig. 14** Classification accuracy in different SNR
Figure 15 shows the loss of network using different regularization methods when SNR is 10 dB. As can be seen from the figure, with the increase of iterations, the loss of network is gradually decreasing. When different regularization methods are used, the loss reduction rate and the loss value when reaching stability are different. As can be seen from the figure, the proposed regularization method can achieve convergence values faster than existing methods including L1 regularization, L2 regularization and the dropout parameter is 0.5. Moreover, the proposed method can achieve a lower loss in network, which makes network more accurate.

5 Results and Discussion

Most existing modulation recognition algorithms have the lower recognition rate at low SNR. In order to solve this problem, this paper proposes a modulation recognition algorithm based on D-GF CNN. In data preprocessing, the ADS technology is used to avoid artificial feature extraction. In the proposed network, dilated convolution method is adopted to process excessive parameters in the network. Furthermore, the proposed GF regularization method can effectively solve the over-fitting problem of the proposed network. The regularization method not only theoretically can be used to reduce the weight of the network, but also gives a reference interval. In this interval, the generalization performance of the network is better. Simulation results show that compared with the existing modulation recognition algorithms, the proposed method has higher recognition accuracy, especially in the low SNR environment.

The ADS method proposed in this paper transforms the modulated signal, but does not eliminate the influence of noise on the recognition accuracy. In the future work, we will consider how to remove the influence of noise.
Acknowledgements. This work was supported by the National Natural Science Foundation of China under Grant No. 61971117, by the Natural Science Foundation of Hebei Province (Grant No. F2020501007).

References

1. Weaver, C. S., Cole, C. A., & Krumland, R. B. (1969). The automatic classification of modulation types by pattern recognition. Stanford electronic laboratories: Stanford University.
2. Huan, C. Y., & Polydoros, A. (1995). Likelihood methods for MPSK modulation classification. IEEE Transactions on Communications, 43(234), 1504.
3. Wei, W., & Mendel, J. M. (2000). Maximum-likelihood classification for digital amplitude-phase modulations. IEEE Transactions on Communications, 48(2), 189–193.
4. Boiteau, D., & Martret, C. L. (1998). A general maximum likelihood framework for modulation classification. In IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 2165–2168).
5. Panagiotou, P., Anastasopoulos, A., & Polydoros, A. (2000). Likelihood ratio tests for modulation classification. In IEEE Milcom Century Military Communications Conference.
6. Swami, A., & Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. IEEE Transactions on Communications, 48(3), 416–429.
7. Zaerin, M., & Seyfe, B. (2012). Multiuser modulation classification based on cumulants in additive white Gaussian noise channel. IET Signal Processing, 6(9), 815–823.
8. Majhi, S., Gupta, R., & Xiang, W. (2017). Hierarchical hypothesis and feature based blind modulation classification for linearly modulated signals. IEEE Transactions on Vehicular Technology, 99, 1–1.
9. Freitas, L. C., Cardoso, C., & Muller, F. C. B. F. (2009). Automatic modulation classification for cognitive radio systems: Results for the symbol and waveform domains. In IEEE Latin-American Conference on Communications.
10. Aslam, M. W., Zhu, Z., & Nandi, A. K. (2012). Automatic modulation classification using combination of genetic programming and KNN. IEEE Transactions on Wireless Communications, 11(8), 2742–2750.
11. Oshea, T. J., West., N. (2016). Radio machine learning dataset generation with gnu radio. Proceedings of the GNU Radio Conference, 1, 1.
12. Mendis, G.J., Wei, J., Madanayake, A. (2016). Deep learning-based automated modulation classification for cognitive radio. IEEE International Conference on Communication Systems.
13. Xie, Wenwu, Sheng, Hu., Chao, Yu., & Zhu, Peng. (2019). Deep learning in digital modulation recognition using high order cumulants. IEEE Access, 7, 63760–63766.
14. Wang, Y. & Liu, M. (2019). Data-driven deep learning for automatic modulation recognition in cognitive radios. IEEE Transactions on Vehicular Technology, 68(4), 4074–4077.
15. Meng, F., Chen, P., & Wu, L. (2018). Automatic modulation classification: A deep learning enabled approach. IEEE Transactions on Vehicular Technology, 67(11), 10760–10772.
16. Dods, S. D. Anderson, T. B. (2006). Optical performance monitoring technique using delay tap asynchronous waveform sampling. Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference.
17. Khan, F. N., Zhou, Y., & Sui, Q. (2014). Non-data-aided joint bit-rate and modulation format identification for next-generation heterogeneous optical networks. Optical Fiber Technology, 20(2), 68–74.
18. Khan, F. N., Teow, C. H., & Kiu, S. G. (2015). Automatic modulation format/bit-rate classification and signal-to-noise ratio estimation using asynchronous delay-tap sampling. Computers & Electrical Engineering, 47, 126–133.
19. Yu, F., Koltun, V. (2016) Multi-Scale Context Aggregation by Dilated Convolutions. ICLR.
20. Tan, K., Chen, J., & Wang, D. L. (2018). Gated residual networks with dilated convolutions for monaural speech enhancement. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 1, 1.
21. Liu, Y., Liu, Y., & Yang, C. (2020). Modulation recognition with graph convolutional network. IEEE Wireless Communication Letters, 99, 1–1.
22. Subbarao, M. V., & Samundiswary, P. (2020). Performance analysis of modulation recognition in multipath fading channels using pattern recognition classifiers. Wireless Personal Communications, 115(2), 1.
23. Shi, W., Liu, D., Cheng, X., Li, Y., & Zhao, Y. (2019). Particle swarm optimization-based deep neural network for digital modulation recognition. IEEE Access, 7, 104591–104600.
24. Abdelbar, M., Tranter, W. H., & Bose, T. (2018). Cooperative cumulants-based modulation classification in distributed networks. *IEEE Transactions on Cognitive Communications and Networking, 4*(3), 446–461.

25. Li, L., Ding, Y., & Zhang, J. K. (2012). Blind detection with unique identification in two-way relay channel. *IEEE Transactions on Wireless Communications, 11*(7), 2640–2648.

26. Han, L., Xue, H., & Gao, F. (2017). Low complexity automatic modulation classification based on order statistics. In *IEEE Vehicular Technology Conference*.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ruiyan Du received her B.S. degree from Hebei Normal University, Shijiazhuang, China, in 1999, the M.S. degree from Yanshan University, Qinhuangdao, China, in 2006, and the Ph.D. degree from Northeastern University, Shenyang, China, in 2012. Since 2012, she is an associate professor in Northeastern University, Qinhuangdao, China. Her research interests include wireless communications, signal processing for communications.

Fulai Liu received the M.S. degree and Ph.D. degree from Northeastern University, Shenyang, China, in 2002 and in 2005, respectively. Since 2010, he is a professor in Northeastern University, Qinhuangdao, China. His research interests include array signal processing and its applications, cognitive radio, millimeter wave MIMO system, etc.
Jialiang Xu received the B.S. degree from Shandong University of Technology, Zibo, China, in 2017. He is currently pursuing the M.S. degree at Northeastern University. His research interests include artificial intelligence and neural network.

Fan Gao received the B.S. degree from Shenyang University of Technology, Shenyang, China, in 2013 and the M.S. degree from Northeastern University, Shenyang, China, in 2019. His research interests include artificial intelligence and neural network.

Zhongyi Hu received the B.S. degree from Northeastern University, Shenyang, China, in 2017. He is currently pursuing the M.S. degree at Northeastern University. His research interests include artificial intelligence and neural network.
Aiyi Zhang received the M.S. degree from Kunming University of Science and Technology, Kunming, China, in 2017. Currently, she is pursuing the PhD at the School of Computer Science and Engineering, Northeastern University, Shenyang, China. Her research interests include tensor analysis, machine learning, signal processing and wireless communications.