

Deep Learning and Data Randomness based Blind Recognition of Channel Codes

Haifeng Peng^{1,2}, Chunjie Cao^{1,2}, Yang Sun^{1,2}, Haoran Li^{1,2}, Kangrui Ye^{1,2},

¹School of Cyberspace Security (School of Cryptology), Hainan University, Haikou, China

²Key Laboratory of Internet Information Retrieval of Hainan Province, Hainan University, Haikou, China

Email: peng_hf@163.com, chunjie_cao@126.com, sunydat@foxmail.com, {flutterblu, 1660584637}@qq.com

Abstract—Channel coding is an efficient approach for data transmission in the wireless network, blind recognition of channel codes plays a pivotal role in non-cooperative communication and spectral scrutiny. But there are some difficulties in the low accuracy and robustness of the current research, meanwhile, most of them analyze this problem to take the coding type as the prior knowledge, which may be unknown in the actual condition. This paper proposes a multi-dimensional feature extraction algorithm, that identifies channel coding types and parameters by the neural network. A data randomness based multi-dimensional feature extraction algorithm and efficient channel attention for deep convolution neural network (ECA-Net) are used in the proposed recognizer, it had an outstanding performance in the identification of coding types, coding lengths, and coding rates. Simulation experiments showed that the proposed recognizer could realize the recognition of coding types, when the signal-to-noise ratio (SNR) is not lower than 10dB, it could achieve above 70% of identification accuracy. On the other hand, it can improve the probability of detection by 10% to 60% compared with previous work for coding lengths and coding rates, which testifies the availability of this approach.

Index Terms—Physical layer security, channel codes, blind identification, coding type.

I. INTRODUCTION

The wireless network is accessorial applied nowadays, and it causes the lack of the frequency spectrum resource [1]. Arising cognitive radio alters the status quo, and remits the problem to a certain degree. Accompanying cognitive radio networks [2] demands makes wireless security a key research area, such as channel codes [3], modulation types [4], medium access control (MAC) protocols [5], [6], etc.

In cognitive radio networks, blind identification of channel codes [7]–[9] is a research field to estimate the channel coding scheme of the master user, and it possesses the significance of wireless physical layer security [10]. Existent algorithms to distinguish channel codes need relative rich professional skills and experience, e.g., Xinyi Wang [11] et al. make use of nested construction in blind detection of polar codes by selecting the candidate of the target physical downlink control

channel (PDCCH). They attempt to decode PDCCH candidates with a lower aggregation level (AL), if decoding fails, iteratively selects a higher AL to decode. But this approach to blind detection of polar codes significantly depends on the experience of the manipulator, and it needs to try to decode codewords. Although compared to exhaustive decoding reduces the number of attempts to decode, if the coding length is considerable, the computational complexity of this algorithm will increase dramatically.

Currently, most of the research about blind channel coding recognition focuses on identifying coding parameters [12]–[17], i.e., coding length and rate. These researchers regard coding type as prior knowledge, still, in the factual information countermeasure and information interception scenarios, a non-cooperative transmitter may be utterly blind to a non-cooperative receiver. That is to say, the coding type of the intercepted codeword is unknown. Therefore, in addition to coding parameters, including coding lengths and rates, it is essential to consider the identification of coding types in a total blindness environment [18].

Over the last few years, the evolution of deep learning makes itself embody functionality in image classification [19] and natural language processing [20], deep learning has shown excellent properties in various domains by data-driven. In the meantime, because of many data in wireless communication systems, deep learning is comprehensively used in communication territories, such as modulation recognition [21], channel estimation [22], and radio frequency fingerprint (RRF) identification [23]. In the field of blind channel coding recognition, Boxiao Shen et al. [24] considered distinguishing coding parameters of channel codes, based on the recurrent neural network (RNN), attention mechanism, and residual neural network (ResNet), proposed three universal channel coding recognizers. Sepehr Dehdashtian et al. [25] based on deep learning, researched blind recognition of channel codes under the AWGN channel and the Rayleigh fading channel, recovered channel coding parameters on the candidate set by the received signal. Still, there is a significant difference between the actual fading circumstance and the Rayleigh fading model [26].

In this paper, we propose a recognizer based on efficient channel attention for deep convolutional neural networks (ECA-Net) [27], which could blindly identify coding types, lengths, and rates using deep learning. Meanwhile, a multi-

The work described in this paper was funded by the Hainan Province Science and Technology Special Fund (No. ZDYF2021GXJS216), the Hainan Province Major Science and Technology Project (No. ZDKJ2021052), the National Natural Science Foundation of China (No. 62162020). The authors would like to thank anonymous reviewers, Associate Editor and Editor for their constructive comments and for improving the manuscript. (*Corresponding author: Chunjie Cao*)

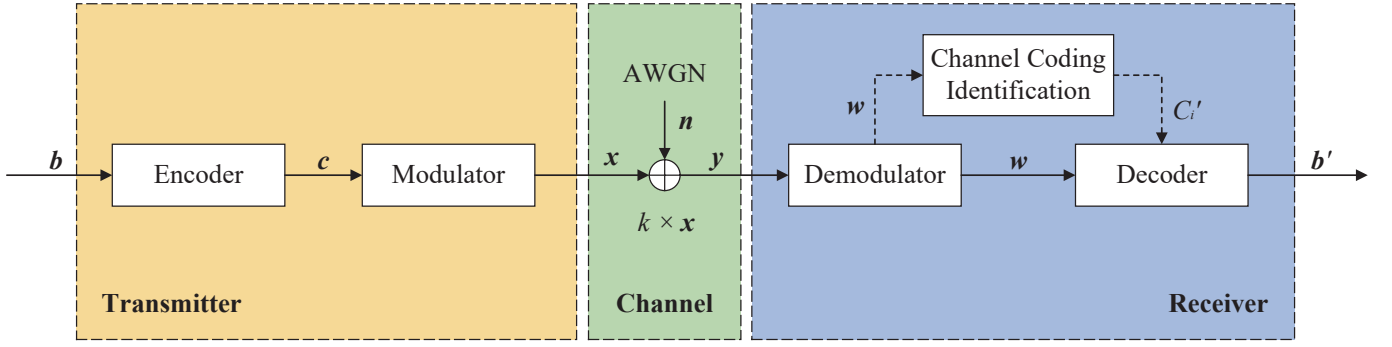


Fig. 1. Structure of the communication link.

dimensional feature extraction algorithm based on data randomness [28] is proposed to dispose of the received codewords, and this recognizer could distinguish channel codes in intercept codewords that are coded with different coding types and parameters. Through enough experiments, we have proved that our recognizer based on deep learning architecture could achieve high performance on blind identification of channel codes. Even without prior knowledge of coding types and parameters, i.e., this recognizer could realize completely blind identification of channel codes.

The residue of this paper is organized as follows. Section II formulates the problems of channel coding recognition and explains data randomness detection. Section III discusses the proposed multi-dimensional feature extraction algorithm and the ECA-Net based recognizer for blind identification of channel codes. Section IV gives the results of simulation experiments. Section V concludes this paper.

II. PRELIMINARIES

A. System Structure

In non-cooperate communications [29], the status of the transmitter is entirely unknown to the receiver, and the receiver can only intercept signals. The goal of blind channel coding recognition is to analyze the type and parameters that the transmitter uses via the received signal.

The communication model under fading conditions is shown in Fig. 1. An information source generates a bit information stream $\mathbf{b} = (b_1, b_2, \dots, b_n)$, where n is the length of the bit sequence. The transmitter codes the bit information stream \mathbf{b} using channel coding parameters C_i , which are from the candidate set $\Theta = C_1, C_2, \dots, C_N$, where N ($1 < i < N$) is the number of channel coding parameters in the set. Then gets the codewords $\mathbf{c} = (c_1, c_2, \dots, c_n)$, where $c_i \in GF(2)$. Codewords would be modulated in BPSK as \mathbf{x} , which equals \mathbf{c}^{BPSK} , and transferred to the channel. The non-cooperate receiver would intercept the signals \mathbf{y} from the channel, and it could be written as

$$\mathbf{y} = k \times \mathbf{c}^{BPSK} + \mathbf{n}, \quad (1)$$

where k is the loss of signal by channel fading, $\mathbf{c}^{BPSK} \in \{+1, -1\}$ is the modulated symbol of the codeword \mathbf{c} , and \mathbf{n}

is the zero-mean AWGN, which is independent and identically distributed with noise power σ^2 . The received signals \mathbf{y} go through demodulation, and become $\mathbf{w} = (w_1, w_2, \dots, w_n)$. The neural network would recognize codewords \mathbf{w} , to get the classification result with coding parameters C'_i . Finally, the receiver will decode \mathbf{w} by the identified C'_i , and get the intercepted bit information \mathbf{b}' .

In cooperative communication, the coding parameters are known by the receiver. But in non-cooperative communication, i.e. electronic warfare, the coding parameters are unknown.

B. Data Randomness Test

Data randomness occupies a very crucial position in cryptography, and the randomness of the output sequence is a significant aspect of evaluating a cryptographic algorithm. A random sequence should be able to through most of the data randomness test [30], each element of the sequence is independent and identically distributed.

Data randomness detection refer to a binary sequence, consider using a series of statistical test method to judge whether it is a really random sequence, and calculate the disparity between them. The data randomness test not only needs to consider the probability of the next bit in the sequence, but also should use some sample sequences to infer whether the sequence is reasonable.

National institute of standards and technology (NIST) [31] suggested the NIST SP800-22 criterion [32], which is the most comprehensive data randomness test method. NIST SP800-22 criterion includes 15 test methods, each generating P by the corresponding test sequence. Then designate significance level α , the test sequence pass the test when $P \geq \alpha$, otherwise do not pass the test. In addition, to make an integrated judgment, it is requisite to select different sample sequences, that would generate different P . Concerning multiple sets of the tests for the different sample sequences, if the test pass rates satisfy

$$r \in \left[P' - 3\sqrt{\frac{P'(1-P')}{m}}, P' + 3\sqrt{\frac{P'(1-P')}{m}} \right], \quad (2)$$

where $P' = 1 - \alpha$, m is the number of the sample sequences, i.e. the times of the tests, and it can determine that the test sequence passes the test. It can also divide the interval $[0, 1]$

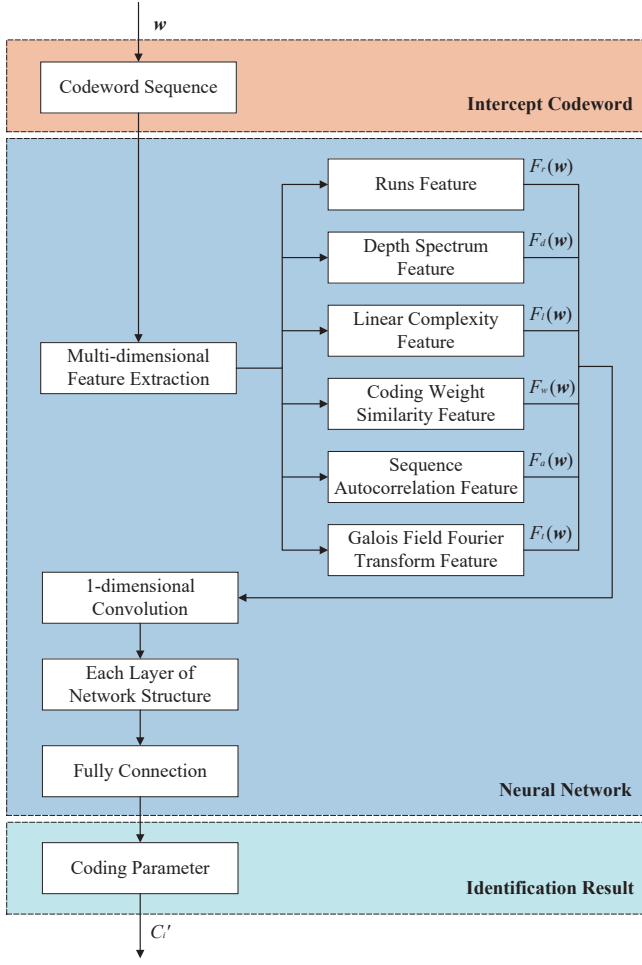


Fig. 2. Scheme of channel coding identification.

into 10 sub-intervals, and calculate the number of P that fall into each sub-interval, then perform the fitting calculation with

$$\chi^2 = \sum_{i=1}^{10} \frac{(n_i - \frac{m}{10})^2}{\frac{m}{10}}, \quad (3)$$

$$P' = \frac{1}{\Gamma(\frac{9}{2})} \int_0^x e^{-t} t^{\frac{9}{2}-1} dt, \quad (4)$$

where n_i is the number of P that fall into the i -th sub-interval. Once more select another significance level β , if it satisfies $P' \geq \beta$, it can tell P obeys uniform distribution. In other words, the test sequence passes the data randomness test.

Different sequences would get different results in the data randomness test, by parity of reasoning, different channel codes with different coding parameters would also get different results. Therefore, feature vectors of channel codes could be obtained by data randomness detection, and we proposed a multi-dimensional feature extraction algorithm of channel codes based on data randomness.

III. PROPOSED RECOGNITION APPROACH

A. Data Preprocessing

Galois field is the essential theoretical basis of channel codes, and the intercept codewords can be seen as an $N \times K$ matrix, where N is the number of frames, and K is the length of the frame, it should note that N and K are both unknown parameters. The matrix can rearrange to a code stream, and based on Galois field $GF(p^n)$, we can get $p = 2$ of the code stream. Then we can fill the code stream with 0, until the length of it satisfy

$$(p^n - 1) \cdot n > N \cdot K, \quad (5)$$

where $(p^n - 1) \cdot n$ is the length of the coding stream after filling, p^n is the number of elements in Galois field $GF(p^n)$, n is the minimum positive integer that makes the above formula true, and $p = 2$ in a binary sequence. The coding stream can arrange to $(p^n - 1) \times n$ matrix after filling, each of these rows with the length n is a coding block, and every coding block obeys Galois field $GF(p^n)$.

B. Feature Extraction

The multi-dimensional feature extraction algorithm that we proposed is shown in Fig. 2. The intercepted codewords will feed into the neural network, and the first layer of the neural network is the multi-dimensional feature extraction algorithm based on data randomness. Then the eigenvectors would be handled with the other layers of the neural network, output the coding parameters by identification.

The number of transformations for consecutive bits 0 and 1 is differs in various binary sequences, and the number of transformations is run. Calculate the sum of the runs in the entire codeword under test is the emphasis of the runs test, and the runs are defined as

$$R_j = \sum_{i=1}^{n-1} r_i + 1, \quad (6)$$

where j ($0 < j < p^n$) is the index of the coding block, n is the length of codeword w , r_i is 0 when bit w_i equals bit w_{i+1} , and is 1 when bit w_i is not equivalent to bit w_{i+1} , bit w_i and w_{i+1} is the i -th and $(i + 1)$ -th bit in codeword w . Then can obtain the feature of runs $F_r(w) = (R_1, R_2, \dots, R_{p^n-1})$.

Depth spectrum is the feature that is generated in codeword by differential operation, and the differential operation could be operated as

$$f(w) = (w_1 \oplus w_2, w_2 \oplus w_3, \dots, w_{n-1} \oplus w_n), \quad (7)$$

where the differential is xor to adjacent bits in binary codes. It would keep differential operation until there are all-zero results, and the differential time d is codeword depth. The multiple differentials could be formulated as

$$f^d(w) = [0^{n-d}], \quad (8)$$

where d is the differential times, the minimum d that can make the above equation true is the codeword depth. The codeword depth belongs to a codeword, and the depth spectrum could

embody the probability of codeword depth in disparate coding parameters. The codeword depth of the j -th coding block can be expressed as d_j , and the feature of the depth spectrum can be given as $F_d(\mathbf{w}) = (d_1, d_2, \dots, d_{p^n-1})$.

It is similar to random sequences, channel codes are generated by linear feedback shift register (LFSR) in normal circumstances, so the linear complexity can be determined by calculating the length of LSFR. The coding block to be calculated can be expressed as $\mathbf{a} = (a_1, a_2, \dots, a_n)$, initialize $i = 0$, $f_0(x) = 1$, $l_0 = 0$, iterative compute $d_i = f_i(x)a_i$. When $d_i = 0$, make $f_{i+1}(x) = f_i(x)$, $l_{i+1} = l_i$, and $i++$ if $i < n$. When $d_i = 0$ and $l_k = 0$ ($0 \leq k \leq i$), make $f_{i+1}(x) = x^{i+1} + 1$, $l_{i+1} = i + 1$, it is similar that $i++$ if $i < n$. When $d_i = 0$ and $l_k < l_{k+1}$ ($0 \leq k \leq i - 1$), make $f_{i+1}(x) = f_i(x) + x^{i-k}f_j(x)$, $l_{i+1} = \max(l_i, i - l_i + 1)$, this moment i is not changing.

After iterative calculation, the linear complexity L of the coding block can be obtained, and the theoretical expectation of a random sequence can be calculated as

$$\mu = \frac{n}{2} + \frac{9 + (-1)^{n+1}}{36} - \frac{\frac{n}{3} + \frac{2}{9}}{2^n}, \quad (9)$$

where n is the length of the coding block \mathbf{a} . Then compared with the theoretical expectation μ of a random sequence, and the result is

$$T_j = (-1)^n (L_j - \mu), \quad (10)$$

where L_j is the corresponding linear complexity for each coding block \mathbf{a}_j , and the feature of linear complexity can be given as $F_l(\mathbf{w}) = (T_1, T_2, \dots, T_{p^n-1})$.

The information section has a constrained relationship with the parity section in the codeword, and the probability of bits in the random sequence is uniform, so there is a big difference in the coding weight distribution between the codeword and the random sequence. Calculate the probability distribution $\mathbf{x} = (x_0, x_1, \dots, x_n)$ of Hamming coding weight for a coding block, the element of \mathbf{x} represents the occurrence probability of the coding weight, and it is similar to calculating the probability distribution $\mathbf{y} = (y_0, y_1, \dots, y_n)$ of Hamming coding weight for a random sequence. The variance and the covariance of coding weight probability distribution in a codeword and a random sequence such that

$$D(x) = \frac{1}{n+1} \sum_{i=0}^n \left(x_i - \frac{1}{n+1} \right)^2, \quad (11)$$

$$D(y) = \frac{1}{n+1} \sum_{i=0}^n \left(y_i - \frac{1}{n+1} \right)^2, \quad (12)$$

$$COV(x, y) = \sum_{i=0}^n \left(x_i - \frac{1}{n+1} \right) \left(y_i - \frac{1}{n+1} \right), \quad (13)$$

where x_i is the i -th element in code weight \mathbf{x} of the coding block, and y_i is the i -th element in code weight \mathbf{y} of the ran-

dom sequence. The probability distributions of the two vectors are statistically similar, and the coding weight similarity is

$$W_j = \left[\frac{COV(x, y)}{\sqrt{D(x)} \cdot \sqrt{D(y)}} \right]^2, \quad (14)$$

the discrepancy of code weight probability distribution in a codeword and a random sequence could reflect their similarity. Hence the feature of coding weight similarity can be given as $F_w(\mathbf{w}) = (W_1, W_2, \dots, W_{p^n-1})$.

The sequence autocorrelation features are based on the relationship between symbols, which is realized by the test of random sequences. The number of different elements in the original sequence and the shifted sequences with

$$E_j = \sum_{i=0}^{n-k+1} w_i \oplus w_{i+k}, \quad (15)$$

where k is logical shift delay. Shift xor addition is used to calculate the self-correlation feature, and then the feature of sequence autocorrelation can be given as $F_s(\mathbf{w}) = (E_1, E_2, \dots, E_{p^n-1})$.

Taking Galois field Fourier transform (GFFT) [33] of the coding block can get the distribution of the spectral peak, thus detecting the periodic characteristic of the coding block, and it could be seen that the degree of deviation. GFFT is a discrete Fourier transform (DFT) [34] of a coding block \mathbf{a} in the Galois field, and DFT could be defined as

$$f_j = \sum_{k=1}^n a_k \left[\cos \left(\frac{2\pi k j}{n} \right) + \sin \left(\frac{2\pi k j}{n} \right) i \right], \quad (16)$$

where $i = \sqrt{-1}$, $j \in [0, n-1]$. The coding block is subjected to Galois field $GF(p^n)$, the GFFT matrix \mathbf{M} can be solved first, and it is a $(p^n-1) \times (p^n-1)$ matrix. The coding stream after the arrangement is a $(p^n-1) \times n$ matrix, and converts the $1 \times n$ row vector to a decimal number, get the $(p^n-1) \times 1$ column vector \mathbf{c} . Then calculate the eigenvector of GFFT, and it can be given as

$$F_g(\mathbf{w}) = \mathbf{M} \cdot \mathbf{c} \quad (17)$$

and it is also a vector of length (p^n-1) .

The features of runs, depth spectrum, linear complexity, coding weight similarity, sequence autocorrelation, and GFFT joint into a 6-dimensional eigenvector with length p^n-1 . It is worth noting that the GFFT feature includes the real and imaginary parts, and then the neural network would be trained and classified by the eigenvectors.

C. Deep Learning Framework

Deep learning [35] is a significant breakthrough in computer science, and plays a central role in most computer vision systems. In recent years, deep learning is also being applied to other areas, and some encouraging results have been achieved.

The neural network filter convolutes eigenmatrix, generates feature maps, then performs max pooling to each feature map, and records the most significant number in each feature map. In the penultimate floor, neural network joints eigenvalue to

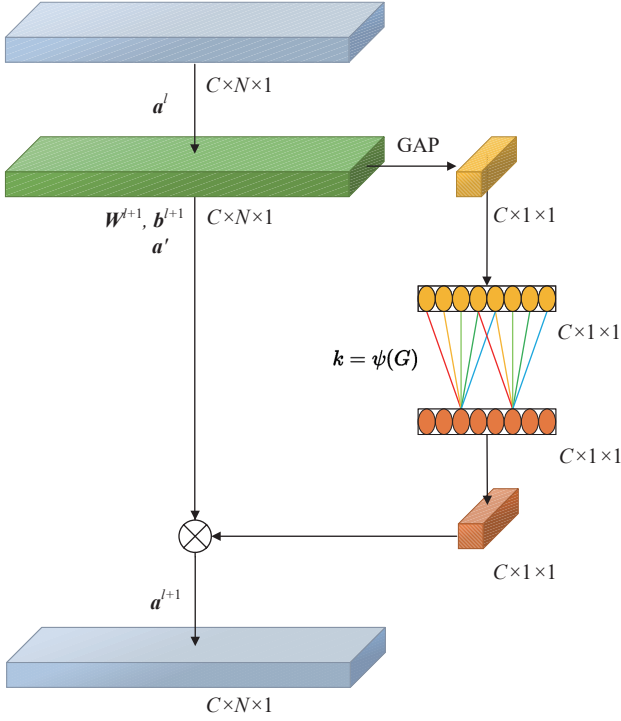


Fig. 3. Operating principle of ECA module.

the eigenvector, finally inputs eigenvector to softmax function, and classifies data by it.

In region size h , there is a filter with weight matrix ω , and $h \times d$ parameters need to be estimated by ω . Eigenmatrix could be denoted as $\mathbf{A} \in \mathbb{R}^{s \times d}$, the sub-matrix from p -th row to q -th row in eigenmatrix \mathbf{A} is $\mathbf{A}[p : q]$. By repeated applying the filter to the sub-matrix in \mathbf{A} , the output $\mathbf{o} \in \mathbb{R}^{s-h+1}$ from the convolution operator could be obtained, and it could be formulated as

$$\mathbf{o}_p = \omega * \mathbf{A}[p : p + h - 1], \quad (18)$$

where p ($1 \leq p \leq s - h + 1$) is the row in \mathbf{A} . So feature mapping \mathbf{z} of this filter could be operated as

$$\mathbf{z} = f(\mathbf{o} + \mathbf{b}), \quad (19)$$

where f is the activate function, and \mathbf{b} is the offset item. The neural network can use filters with identical or different area sizes to learn complementary features from the input eigenmatrix, and achieve the purpose of training.

D. ECA-Net

ECA-Net is a sort of convolutional neural network (CNN) [36] with novel channel attention mechanisms, and it improves channel attention module based on squeeze-and-excitation network (SeNet) [37]. ECA-Net realized a new approach to the adaptive selection of convolutional kernel and local cross-channel interaction policy, it not only boosts the performance of complex attention modules, but also reduces the complexity of the neural network model.

TABLE I
HYPER-PARAMETER SELECTINGS.

Description	Values
Activation	ReLU
Optimizer	Adam
Learning rate	0.01
Batch size	256
Number of iterations	20000

In the ECA module, the operating principle of the channel attention mechanism is shown in Fig. 3. Firstly input the features of original data, proceed with global average pooling (GAP) [38] to the features, and obtain features without dimension reduction. On this basis, the ECA module gets local cross-channel interactions by fast 1-dimensional convolution with size k , and parameter k could be generated by adaptive function with size C , which is the size of the input channel. Later, using the sigmoid function, calculates the weight in each channel, and obtains features with channel attention by combining original input features and channel weights. The neural network that takes in the ECA module could extract discriminant features more efficiently, and achieve better performance based on channel dimension.

We establish ECA-Net based on ResNet [39], input features X and convolution features $F(X)$ directly correspond to the addition of dimensions, i.e., $F(X) + X$. It avoids the gradient dissipation problem that the depth of the neural network is too deep, making the performance of the global neural network would not have to deviate as the depth increases. In residual base block, the neural network cuts in the ECA module after the third convolution, and the preponderance of channel attention mechanism perfects ResNet. It makes the neural network focus on discriminant regions of data, thus improving the performance of the neural network.

Overall models of the neural network all use cross entropy as the loss function, and the loss function of each branch could write as

$$L_{raw} = -\log(P_r(c)), \quad (20)$$

$$L_{object} = -\log(P_o(c)), \quad (21)$$

$$L_{part} = -\sum_{n=0}^{N-1} \log(P_{p(n)}(c)), \quad (22)$$

where L_{raw} , L_{object} , L_{part} are respectively the loss functions of each branch in raw codewords, object-level codeword, part-level codeword, c is the actual label of input codeword, $P_r(\cdot)$, $P_o(\cdot)$ are respectively classification probability values that softmax function takes raw codeword, object level-codeword after training by the neural network, $P_{p(n)}(\cdot)$ is the classification that softmax function takes the n -th codeword in part level codewords after training by the neural network, and N

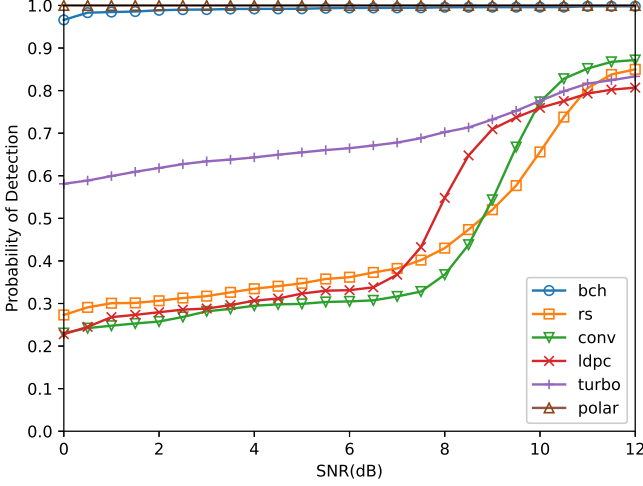


Fig. 4. Probability of detection among different coding types.

is the number of part-level codewords. The loss function of overall neural network L_{total} is defined as

$$L_{total} = L_{raw} + L_{object} + L_{part}, \quad (23)$$

in the base of L_{total} , the model makes overall neural network train fitting by constantly training to optimize.

E. Hyper-Parameter Selection

Hyper-parameter adjustment is the essential step in deep learning, and we select the optimal hyper-parameters after multiple experimental adjustments. To choose the best hyper-parameter, we conducted the grid search in the range of 3 to 9 for the number of residual modules, 64 to 256 for the number of convolutional kernels, 2 to 9 for the size of convolutional kernels, 0.1 to 0.001 for the learning rate, 64 to 256 for the batch size.

The number of residual modules is selected as 4, in coding type recognition, the number of convolution kernel is 64, with the sizes of 3, 5, 5, 7, and in coding parameter recognition, the number of convolution kernel is 32, with the sizes of 3, 3, 3, 3. The softmax function on the last floor, and the training hyper-parameters in the neural network are given in Table I.

Based on the parameters in Table I, we could train a neural network model, and measure the accuracy of the trained model in blind channel coding identification.

IV. PERFORMANCE EMULATION

The data set we use is from [24], and we consider identification of coding type, coding length, and coding rate in the experiments. In each module of the experiments, the numbers of the codewords in training datasets, validation datasets, test datasets are 1,200,000, 400,000, 400,000, and the range of SNR is from 0dB to 12dB in the circularly symmetric complex Gaussian noise (CSCG) channel.

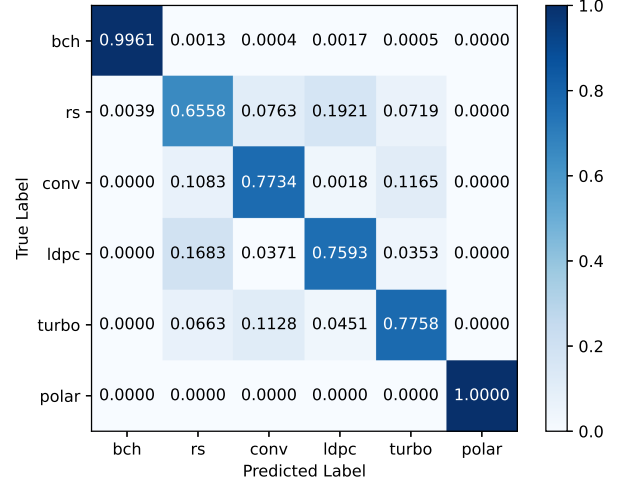


Fig. 5. Confusion matrix for coding types recognition, SNR = 10dB

We employ Windows Server 2019, Intel Xeon Gold 5117, and Nvidia RTX 2080Ti in experiments, and the experiment environment is CUDA 11.6.0, python 3.7.9, and torch 1.8.1, which is in accord with the requirement of many experiments. Because the received codewords are suffered from channel interference, including noise and fading, it would be dealt with the feature extraction algorithm at first, then identify the coding parameters C'_i by the neural network.

A. Coding Type Recognition

We choose Bose-Chaudhuri-Hocquenghem (BCH) code, Reed-Solomon (RS) code, convolutional code, low-density parity-check (LDPC) code, turbo code, and polar code in the candidate set. In other words, we set the parameters of coding types as

$$\mathcal{T} = \{bch, rs, conv, ldpc, turbo, polar\}, \quad (24)$$

these coding types are all widely used in wireless communication and satellite communication. The goal of coding type recognition in our experiment is not just concerned with identifying a particular type of all the codewords, but regard it as a multi-classification problem, identifying the corresponding types of all the codewords.

We assess the proposed recognizer for blind recognition veracity in experiments, Fig. 4 demonstrate the experimental results. As the figures show, the proposed recognizer could achieve good identification performance for BCH, RS, and polar codes in a wide range of SNR. To be specific, this recognizer can approximate 70% for each channel code when SNR is not lower than 10dB, at this moment, it can effectively distinguish all channel coding types. We can find that the probability of detection for BCH codes and polar codes is very high, this is because BCH code is a relatively simple cyclic code, there is a close relationship between the generator polynomial and the minimum coding distance, and the coding

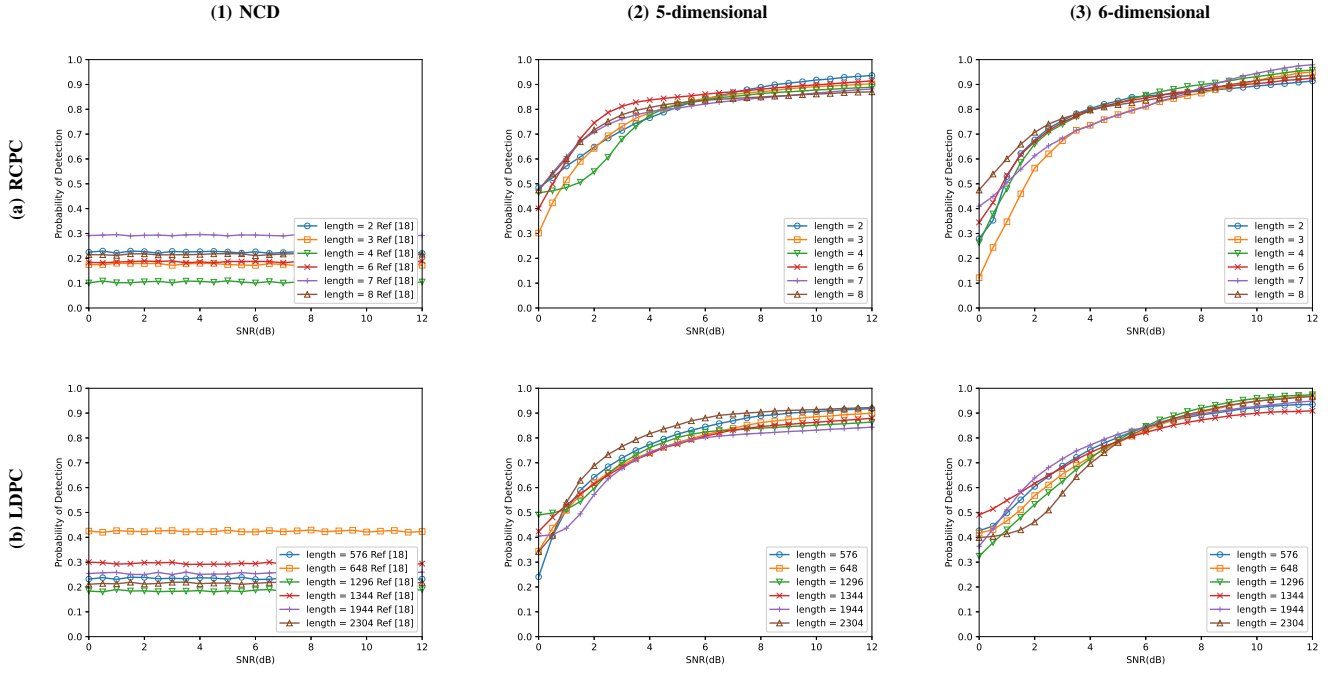


Fig. 6. Probability of detection for different coding lengths of (a) RCPC codes, (b) LDPC codes, by (1) NCD algorithm, (2) 5-dimensional feature, (3) 6-dimensional feature including GFFT, at different SNR ranges, the NCD algorithm is from [18].

theory of polar code is based on channel polarization theory, it is different significantly from other coding types.

For analyzing the misjudgment relationship between different channel coding types, when SNR = 10dB, the confusion matrix for blind recognition of coding types are shown in Fig. 5. The vertical axis is the actual label of the codeword, the horizontal axis is the recognition result bt the proposed recognizer, and the diagonal is the categories that are identified accurately. It indicates that the recognition accuracy all reaches 65% for different channel coding types at SNR 10dB, and it can fully recognize BCH code and Polar code. It could be seen from the above results that there is relatively large confusion between RS code and LDPC code.

B. Coding Length Recognition

The coding length is the length of the output codeword from the encoder, the coding rate is the quotient of the information bit length and the coding length, it also reflects the constraint relation between the information bit and the check bit in the codeword. In the total blindness environment, the feature of coding length is more evident than the coding rate. Therefore, we consider the recognition of the coding length at first, and then consider the coding rate.

We select rate-compatible punctured convolutional (RCPC) codes and LDPC codes in coding length and rate recognition experiments, and there are 6 coding lengths in the candidate set for each channel code. For RCPC codes, we set the parameters of coding lengths as

$$\mathcal{L} = \{2, 3, 4, 6, 7, 8\}, \quad (25)$$

and the generator polynomials \mathcal{G} are from the DVB-RCS criterion [40]. Analogously, for LDPC codes, we set the parameters of coding lengths as

$$\mathcal{L} = \{576, 648, 1296, 1344, 1944, 2304\}, \quad (26)$$

the parity-check matrixes \mathcal{G} are from IEEE 802.11 criterion [41] and IEEE 802.16 criterion [42]. At the same time, considering the computation complexity of GFFT is relatively more significant than other dimensional features, we compare 5-dimensional features without GFFT and 6-dimensional features including GFFT. In addition, the number of classes deficiency (NCD) algorithm of previous work [18] is also being compared.

To demonstrate the performance of our recognizer, we take simulation experiments, and Fig. 6 illustrates the results. Fig. 6(a) is the probability of detection for RCPC coding lengths, and Fig. 6(b) is the probability of detection for LDPC coding lengths. In Fig. 6(2) and Fig. 6(3), we can find that the boost of SNR advances recognition precision, especially recognition precision that reaches above 70% when SNR is more significant than 4dB, the probability of detection for coding length is much improved. Meanwhile, the result of identification by the NCD algorithm is shown in Fig. 6(1), it could be seen that although the previous work is not affected by SNR, the performance is relatively lower than the proposed recognizer in this paper, the recognition accuracy is only about 20%, which is significantly lower than our recognizer.

In addition, comparing the 5-dimensional features without GFFT to the 6-dimensional features including GFFT, it shows that the 6-dimensional features are approximately 5% better than 5-dimensional features when SNR is comparatively high.

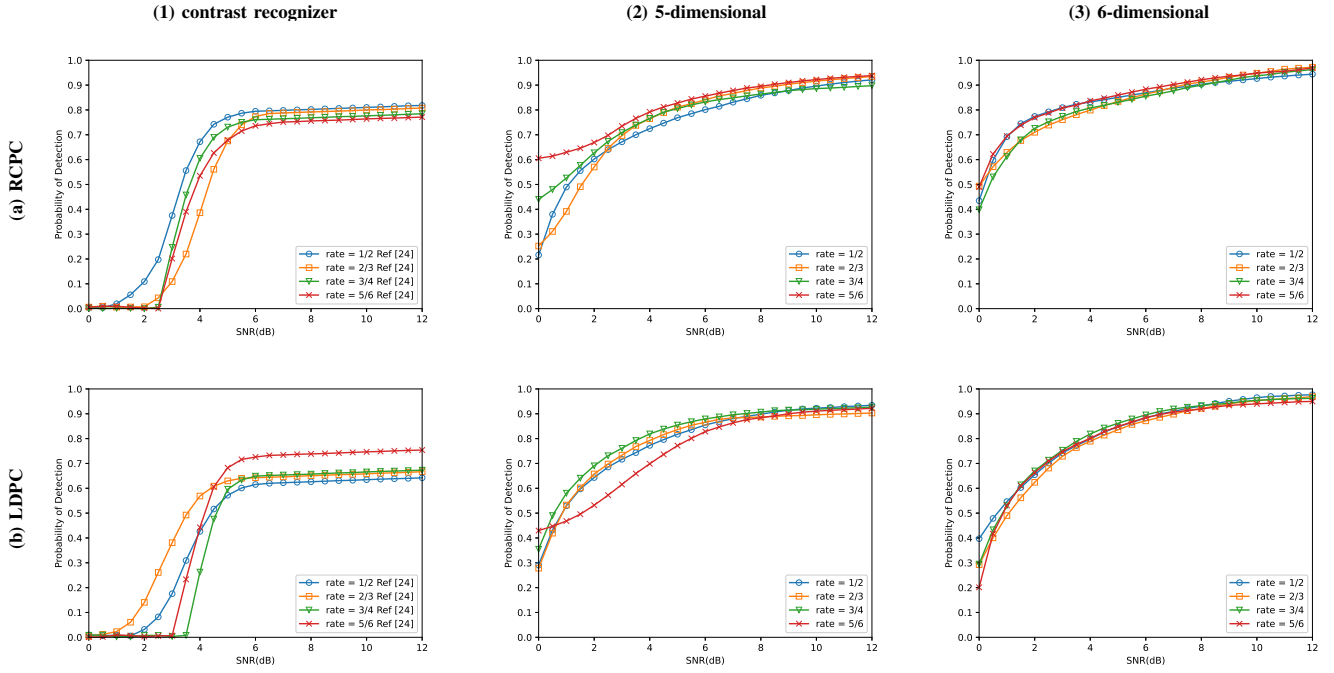


Fig. 7. Probability of detection for different coding rates of (a) RCPC codes, (b) LDPC codes, by (1) contrast recognizer, (2) 5-dimensional feature, (3) 6-dimensional feature including GFFT, at different SNR ranges, the contrast recognizer is from [24].

Therefore, although GFFT increases computational complexity, it can improve the recognition accuracy to a certain extent, and can be selected according to requirements in practical application. The results of simulation experiments testify to the effectiveness of the proposed recognizer for blind identification of coding length.

C. Coding Rate Recognition

Next, we evaluate the performance of the proposed recognizer in coding rate recognition, and compare it with the NCD algorithm. In experiments, the channel codes selected are the same as in the previous subsection. For RCPC codes and LDPC codes, we set the parameters of coding rates as

$$\mathcal{R} = \left\{ \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5} \right\}, \quad (27)$$

the generator polynomials \mathcal{G} of RCPC codes are also from the DVB-RCS criterion, and the parity-check matrixes \mathcal{G} are also from IEEE 802.11 criterion and IEEE 802.16 criterion. Similar to the last subsection, we compare 5-dimensional features without GFFT and 6-dimensional features including GFFT. In addition, we also select the contrast recognizer of previous work [18] to proceed with performance comparison.

Meanwhile, to illustrate the performance of this recognizer, the experiment results are shown in Fig. 7. Thereinto, Fig. 7(a) shows the probability of detection for RCPC coding rates, and Fig. 6(b) shows the probability of detection for LDPC coding rates. In Fig. 7(2) and Fig. 7(3), we can find that the identification accuracy exceeds 70% when SNR is not lower than 5dB, the proposed recognizer can effectively distinguish the coding rate at the moment. As shown in Fig. 7(1), the

recognition precision of the contrast recognizer is a little lower than our recognizer. The recognition is about 10% lower than our recognizer when the SNR is greater than 9db, and about 20% lower when the SNR is less than 9db.

Similarly, when SNR is comparatively high, the 6-dimensional features without GFFT are close to 5% better than 5-dimensional features including GFFT, computational complexity and identification accuracy can be tradeoffs in a real environment. The above experimental results prove that the proposed channel coding recognizer in this paper can effectively distinguish different coding rates.

V. CONCLUSION

This paper illustrates the feasibility of deep learning based channel coding recognition. This approach avoids manual feature extraction, which needs experience and professional skill of the handlers, meanwhile enormously reducing computation complexity. A multi-dimensional feature extraction method based on data randomness detection was proposed to extract the features of the received codewords. In addition, a novel ECA-Net based channel coding recognizer was presented, it shows a predominant property in the identification of coding types, coding length, and coding rates, simultaneously, this recognizer also has a decent generalization ability.

REFERENCES

- [1] Y. Wang, Y. Zeng, S. Sun, and Y. Nan, "High accuracy uplink timing synchronization for 5g nr in unlicensed spectrum," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 604–608, 2020.
- [2] V. Shakhov and I. Koo, "Analysis of a network stability-aware clustering protocol for cognitive radio sensor networks," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12476–12477, 2021.

- [3] L. Zhang, Y. Zhang, W. Yan, and L. Ma, "Blind identification of space-time block codes based on deep learning," *Chinese Journal of Aeronautics*, vol. 35, no. 1, pp. 426–435, 2022.
- [4] P. Chu, L. Xie, C. Dai, and Y. Chen, "Automatic modulation recognition for secondary modulated signals," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 962–965, 2021.
- [5] J. Gao, W. Zhuang, M. Li, X. Shen, and X. Li, "Mac for machine-type communications in industrial iot—part i: Protocol design and analysis," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9945–9957, 2021.
- [6] J. Gao, M. Li, W. Zhuang, X. Shen, and X. Li, "Mac for machine-type communications in industrial iot—part ii: Scheduling and numerical results," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9958–9969, 2020.
- [7] H. Sun, E. Viterbo, and R. Liu, "Efficient blind detection scheme based on simplified decoding of polar codes," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 864–868, 2020.
- [8] A. Jalali and Z. Ding, "Joint detection and decoding of polar coded 5g control channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2066–2078, 2020.
- [9] H. Ju, H. Lee, M. Jang, and S.-H. Kim, "Path-metric-based false alarm reduction in blind detection of polar codes," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2020, pp. 1–6.
- [10] J. D. V. Sánchez, P. Ramírez-Espinosa, and F. J. López-Martínez, "Physical layer security of large reflecting surface aided communications with phase errors," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 325–329, 2020.
- [11] X. Wang, C. Cao, C. Sun, Z. Fei, J. Yuan, and M. Xiao, "Nested construction of polar codes for blind detection," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 711–715, 2020.
- [12] M. Song, J. Kim, and D.-J. Shin, "Blind reconstruction of bch and rs codes using single-error correction," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5120–5133, 2021.
- [13] Y. Liu and F. Wang, "Blind data detection with unknown channel coding," *IEEE Communications Letters*, vol. 24, no. 4, pp. 758–761, 2020.
- [14] Z. Wu, Z. Zhong, and L. Zhang, "Blind recognition of cyclic codes based on average cosine conformity," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2328–2339, 2020.
- [15] S. Kwon and D.-J. Shin, "Analysis of blind reconstruction of bch codes," *Entropy*, vol. 22, no. 11, p. 1256, 2020.
- [16] T. R. Dean, M. Wootters, and A. J. Goldsmith, "Blind joint mimo channel estimation and decoding," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2507–2524, 2018.
- [17] J. Perlstein, T. Dean, M. Wootters, and A. Goldsmith, "Fast blind mimo decoding through vertex hopping," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 148–154.
- [18] A. Bonvard, S. Houcke, R. Gautier, and M. Marazin, "Classification based on euclidean distance distribution for blind identification of error correcting codes in noncooperative contexts," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2572–2583, 2018.
- [19] T. Chen, Y. Zhao, and Y. Guo, "Sparsity-regularized feature selection for multi-class remote sensing image classification," *Neural Computing and Applications*, vol. 32, no. 11, pp. 6513–6521, 2020.
- [20] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 604–624, 2020.
- [21] S. Lin, Y. Zeng, and Y. Gong, "Learning of time-frequency attention mechanism for automatic modulation recognition," *IEEE Wireless Communications Letters*, vol. 11, no. 4, pp. 707–711, 2022.
- [22] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1132–1142, 2021.
- [23] G. Shen, J. Zhang, A. Marshall, and J. R. Cavallaro, "Towards scalable and channel-robust radio frequency fingerprint identification for lora," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 774–787, 2022.
- [24] B. Shen, C. Huang, W. Xu, T. Yang, and S. Cui, "Blind channel codes recognition via deep learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2421–2433, 2021.
- [25] S. Dehdashtian, M. Hashemi, and S. Salehkalebar, "Deep-learning-based blind recognition of channel code parameters over candidate sets under awgn and multi-path fading conditions," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1041–1045, 2021.
- [26] K.-L. Besser and E. A. Jorswieck, "Reliability bounds for dependent fading wireless channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5833–5845, 2020.
- [27] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 11 534–11 542.
- [28] Y. Liu, Y. Zhou, S. Sun, T. Wang, R. Zhang, and J. Ming, "On the security of lattice-based fiat-shamir signatures in the presence of randomness leakage," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1868–1879, 2020.
- [29] J. Dai, J. Gao, and K. Niu, "Learning to mitigate the far in polar code blind detection," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1692–1695, 2019.
- [30] Z. Qu, S. Zhao, J. Xu, Z. Lu, and Y. Liu, "How to test the randomness from the wireless channel for security?" *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3753–3766, 2021.
- [31] C.-Y. Lee, K. Bharathi, J. Lansford, and S. P. Khatri, "Nist-lite: Randomness testing of rngs on an energy-constrained platform," in *2021 IEEE 39th International Conference on Computer Design (ICCD)*. IEEE, 2021, pp. 41–48.
- [32] A. Iwasaki, "Deriving the variance of the discrete fourier transform test using parseval's theorem," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1164–1170, 2019.
- [33] S. Lin, K. Abdel-Ghaffar, J. Li, and K. Liu, "A scheme for collective encoding and iterative soft-decision decoding of cyclic codes of prime lengths: Applications to reed-solomon, bch, and quadratic residue codes," *IEEE Transactions on Information Theory*, vol. 66, no. 9, pp. 5358–5378, 2020.
- [34] M. Chen, H. Chen, L. Fan, S. Zhu, W. Xi, and D. Feng, "A new discrete fourier transform randomness test," *Science China Information Sciences*, vol. 62, no. 3, pp. 1–16, 2019.
- [35] L. Liao, Z. Huang, and W. Wang, "A statistical learning model with deep learning characteristics," in *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE, 2021, pp. 137–140.
- [36] H. Zhang, L. Yu, Y. Chen, and Y. Wei, "Fast complex-valued cnn for radar jamming signal recognition," *Remote Sensing*, vol. 13, no. 15, p. 2867, 2021.
- [37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [38] Q. Wang, J. Xie, W. Zuo, L. Zhang, and P. Li, "Deep cnns meet global covariance pooling: Better representation and generalization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 8, pp. 2582–2597, 2020.
- [39] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on resnet-50," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6111–6124, 2020.
- [40] *Interaction Channel for Satellite Distribution Systems*. document TSI EN 301 790 V1.5.1, Digital Video Broadcasting (DVB), 2009.
- [41] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks-Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. Standard 802.11-2016 (Revision of IEEE Std 802.11-2012), 2016.
- [42] *IEEE Standard for WirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems—Amendment 2: Higher Reliability Networks*. Standard 802.16.1a-2013 (Amendment to IEEE Std 802.16.1-2012), 2013.