



# Article A Modulation Recognition System for Underwater Acoustic Communication Signals Based on Higher-Order Cumulants and Deep Learning

Run Zhang <sup>1,2</sup>, Chengbing He <sup>1,2</sup>, Lianyou Jing <sup>2,3,\*</sup>, Chaopeng Zhou <sup>1</sup>, Chao Long <sup>1</sup> and Jiachao Li <sup>1</sup>

- <sup>1</sup> School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China; run2rui@mail.nwpu.edu.cn (R.Z.); hcb@nwpu.edu.cn (C.H.)
- <sup>2</sup> Research and Development Institute, Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
- <sup>3</sup> Ocean Institute of NPU, Northwestern Polytechnical University, Taicang 215400, China
- \* Correspondence: lyjing@nwpu.edu.cn

Abstract: Underwater acoustic channels, influenced by time-varying, space-varying, frequencyvarying, and multipath effects, pose significant interference challenges to underwater acoustic communication (UWAC) signals, especially in non-cooperative scenarios. The task of modulating and identifying distorted signals faces huge challenges. Although traditional modulation recognition methods can be useful in the radio field, they often prove inadequate in underwater environments. This paper introduces a modulation recognition system for recognizing UWAC signals based on higher-order cumulants and deep learning. The system achieves blind recognition of received UWAC signals even under non-cooperative conditions. Higher-order cumulants are employed due to their excellent noise resistance, enabling the differentiation of OFDM signals from PSK and FSK signals. Additionally, the high-order spectra differences among signals are utilized for the intra-class recognition of PSK and FSK signals. Both simulation and lake test results substantiate the effectiveness of the proposed method.

**Keywords:** underwater acoustic communication; modulation recognition; higher-order cumulants; deep learning

# 1. Introduction

The ocean contains abundant untapped resources. To safeguard maritime sovereignty and maritime rights and interests, the demand for underwater information transmission is becoming increasingly urgent. However, the propagation of electromagnetic waves in underwater media suffers from severe attenuation, which greatly limits the direct application of wireless technologies in underwater communication. Although there are technologies such as photonic communication for underwater communication, they have higher transmission bandwidth, data transfer rates, lower link delays, and are difficult to detect with high security. However, limitations, such as losses, caused by complex turbulence, bubbles, seawater absorption, and light scattering in the underwater environment restrict their transmission distance and reliability [1,2]. As a result, underwater acoustic communication currently serves as the primary form of underwater information transmission and is gaining more importance.

In underwater acoustic communication (UWAC), information is transmitted via sound waves, and information is loaded by mapping parameters, such as phase, frequency, and amplitude. In practical applications, there are two scenarios for the underwater communication receiver to receive and process signals. In a cooperative communication scenario, prior to the accurate restoration of the data encompassed within the signal, the receiver must identify the modulation method, also known as the information loading



Citation: Zhang, R.; He, C.; Jing, L.; Zhou, C.; Long, C.; Li, J. A Modulation Recognition System for Underwater Acoustic Communication Signals Based on Higher-Order Cumulants and Deep Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 1632. https://doi.org/10.3390/ imse11081632

Academic Editor: Rouseff Daniel

Received: 29 June 2023 Revised: 16 August 2023 Accepted: 18 August 2023 Published: 21 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). method; in a non-cooperative communication scenario, the reconnaissance party needs to quickly and accurately classify and recognize the intercepted signals' modulation methods, which is a prerequisite for further analyzing and mastering the enemy's communication information. Even if our side cannot directly decipher the information, we can send deceptive signals to interfere with the enemy's communication signals by identifying their modulation methods, which can provide a new idea for underwater communication countermeasures technology. Building a smooth and secure underwater communication network in both civilian and military fields requires the automatic modulation classification (AMC) of UWAC signals. Therefore, AMC holds great significance.

Figure 1 displays the two principal classifications of modulation identification methods: likelihood-based (LB) hypothesis testing based on decision-theory and feature-based (FB) modulation classification based on feature extraction. The FB modulation classification can be further segregated into three distinct types: statistical pattern analysis utilizing feature extraction, modulation classification based on machine learning, and modulation classification based on deep learning.



Figure 1. Common modulation recognition methods.

The hypothesis testing method based on decision theory, which is likelihood-based, treats the modulation recognition problem as a probability-based hypothesis testing problem, utilizing hypothesis testing-related theory. It establishes the maximum likelihood discrimination formula of the signal, deduces the optimal decision threshold of the signal, compares a certain statistical quantity of the signal with the threshold value, and judges the signal modulation type [3]. Despite the robust theoretical foundation of modulation recognition technology based on the maximum likelihood ratio hypothesis testing algorithm, this approach necessitates prior knowledge, such as the signal's mean, variance, and covariance, which can be challenging to acquire in typical non-cooperative communication settings. This results in practical difficulties in application, insufficient generalization and robustness, and high computational complexity. Therefore, this research method is not commonly used at present [4].

The feature-based statistical pattern analysis method selects and extracts the signal's modulation-type features, transferring them from the object space to the feature space. Subsequently, this method analyzes the distribution properties of the various features, identifies the clustering patterns of different features, partitions the feature space, and then maps it from the feature space to the decision space to make classification decisions or judgments, ultimately accomplishing modulation recognition [5]. The disadvantage is that the recognition framework of this method currently lacks a unified and complete theoretical basis as support, and the recognition system is relatively complex. The algorithm usually extracts features based on specific signal samples, and the decision threshold is set artificially based on experience, so the recognition effect is greatly affected by changes in noise and environment. When entering unfamiliar waters and the channel is not ideal, the features will become blurred or even invalid, and the system's robustness will be insufficient [6].

The machine learning-based signal modulation recognition method initially extracts specific signal features and subsequently utilizes machine learning categorizer algorithms to differentiate the signal's modulation type based on the differences in signal features [7]. As the machine learning-based signal modulation recognition method is created manually based on specific standards and then feature selection is determined based on these criteria, the feature extraction procedure and the final prediction model learning are separate. Consequently, the acquired features may not enhance the final model's performance or guarantee the model's capacity for generalization.

The deep learning-based modulation recognition method employs deeper neural networks and data-driven techniques to solve the modulation recognition problem without human intervention [8,9]. It extracts complex features directly from raw data using a feature extraction network and makes effective recognition decisions using a classification network, thereby achieving end-to-end feature extraction and recognition. Compared to conventional methods, it eliminates the need for complex manual feature selection and is capable of achieving synchronous optimization processing of the feature extraction network and classifier under sufficient training samples. This results in excellent communication signal recognition performance and makes it an important technical method for target feature extraction and recognition [10].

Thus, in comparison to the preceding three approaches, deep learning network technology possesses potent representation learning capabilities, enabling the automatic extraction of various intricate features from raw data. In the face of a complex underwater acoustic channel, the deep learning-based modulation recognition classifier has good robustness and can adapt to changes in the underwater environment, even under low signal-to-noise ratio conditions, it can still effectively complete the signal modulation recognition task.

In 1998, Nandi, A.K. et al. systematically expounded the theory and algorithm of automatic recognition of communication signal modulation in reference [11]. The simulation results indicate that the results obtained by the artificial neural network method are better than those obtained by the decision theory method. In 2016, O'Shea, T.J. et al. utilized convolutional neural networks for the first time in communication signal recognition, as referenced in [12]. By constructing an end-to-end convolutional neural network, he successfully identified the modulation methods of 11 different signals. In 2018, the team further demonstrated the ability of deep learning (DL) in radio modulation recognition through experimental research, which is stronger than computer vision (CV) and machine learning (ML). The signal recognition model based on VGG-Net and ResNet networks can identify 24 signal modulation methods [13]. The investigation carried out by Jeong, S. et al. in [14] employed short-time Fourier transform (STFT) for generating a time-frequency diagram of the communication signal. Subsequently, CNN was utilized to capture the signal characteristics present in the time-frequency diagram. The FSK, PSK, and QAM modulation methods were recognized. Zhang, Z. et al. proposed a feature fusion scheme for AMC based on convolutional neural networks in reference [15]. By filtering the pseudo-Wigner-Ville distribution and Born-Jordan distribution, this method transforms the signal into two time-frequency images and employs a finely tuned CNN model to extract image features. The outcomes of the simulation revealed that the scheme can achieve a classification accuracy of 92.5% when the SNR is 4 dB.

The aforementioned accomplishments are some outcomes in the domain of modulation recognition technology in wireless communication. In the complex time-varying and space-varying underwater acoustic channel environment, improvements are needed in wireless technologies to better adapt to the modulation recognition scenarios in UWAC. Cheng, E. et al. proposed a method for recognizing MPSK-like UWAC signals, which contain Gaussian white noise and multipath, in their study. Initially, this technique employs wavelet transform on the signals and utilizes the variance amplitude of the processed signals for inter-class recognition to ascertain whether they are PSK-like modulation signals. Eventually, the fourth-order cumulant of the signals is computed for intra-class recognition of PSK-like signals. The simulation outcomes revealed that, at an SNR of -5 dB, the identification precision for both BPSK and QPSK was 100% [16]. To address the problem of acquiring high-quality labeled data in time-varying and space-varying underwater channels, Xu, Z. et al. presented a semi-supervised learning-based blind modulation recognition technique called SSLUWA in their work [17]. The technique employs linear interpolation to pseudo-label unlabeled signals and trains the classification system using the interpolation consistency principle to extract classification features from unlabeled signals and test the knowledge learned from labeled signals, improving recognition accuracy when labeled data are scarce. Experimental results showed that compared to fully labeled samples, when the labeled samples accounted for only 10%, the recognition accuracy was 99% at an SNR of 2 dB.

In the context of non-cooperative reception, Wang, B. et al. proposed a deep learningbased blind detection technique for underwater acoustic communication signals [18]. The technique utilizes an impulse noise preprocessor (INP) and a generative adversarial network (GAN) to preprocess the received signals for noise reduction, mitigating the adverse effects of underwater impulse noise. Subsequently, an automatic feature extraction technique based on a convolutional neural network (CNN) is employed to differentiate underwater acoustic communication signals from underwater noise. Furthermore, in their subsequent research documented in [19], they introduced a hybrid neural network (HNN) for modulation classification of UWAC signals. The network employs the same impulse noise preprocessor (INP) as in [18] for noise reduction, and then employs a CNN with an attention mechanism to extract signal features for recognition of 2FSK, 4FSK, 8FSK, PSK, OFDM, and other signals. Subsequently, a sparse autoencoder is used for the intra-class recognition of PSK-like signals, distinguishing between BPSK and QPSK. Field trial results showed that except for 8FSK, the recognition accuracy for other signals was above 84.5%. Lastly, in [20], the authors presented their new achievement, an underwater acoustic communication modulation recognition technique suitable for small sample conditions, named IAFNet. It includes an impulse noise pre-processing module (INP), an attention mechanism (AN), and few-shot learning (FSL). The fundamental concept is to extract similarity features from a handful of labeled signals and unlabeled signals, assign weights to the features using an attention network, and then feed them into a similarity comparison module to ascertain the modulation type of the unlabeled signals. The field trial outcomes indicated that, aside from BPSK and QPSK signals, which had a 30% probability of being classified erroneously as OFDM signals, the recognition accuracy for other signals exceeded 93%, demonstrating the effectiveness of the proposed approach.

Wang, J. et al. proposed a deep fusion neural network model called R&CNN (recurrent and convolutional neural network) for modulation classification of UWAC signals in their work [21]. The model constructs a recurrent layer using a gated recurrent unit (GRU), which can mitigate the interference of the Doppler effect by memorizing and processing signal sequences. In the Yellow Sea zone of China, they assembled a dataset containing measured signals of seven modulation types, namely BPSK, QPSK, BFSK, QFSK, 16QAM, 64QAM, and OFDM. The classification recognition accuracy on the field trial data reached 99.38%, outperforming traditional methods, such as AlexNet8, LSTM, and CNN-LSTM in terms of recognition accuracy.

This paper presents a modulation recognition system designed for the blind identification of received signals in the passband, utilizing modulation techniques, like binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), 2-frequency shift keying (2FSK), 4-frequency shift keying (4FSK), and orthogonal frequency division multiplexing (OFDM) in non-cooperative settings, particularly tailored to complex underwater channel environments. Below, we outline the primary contributions of this paper:

- (1) An improved sixth-order cumulant is proposed for the identification of OFDM signals.
- (2) An improved bispectrum is proposed for the identification of BPSK, QPSK, 2FSK and 4FSK signals.
- (3) Extensive simulations are conducted in a theoretical computer simulation environment and a Bellhop simulation environment. The simulations involve the addition

of realistic in-band colored noise and multipath effects to the transmitted signals. The simulation outcomes validate the effectiveness of the proposed techniques. Moreover, field experiments conducted in an actual lake environment furnish additional proof of the effectiveness and resilience of the proposed system.

#### 2. System Model

## 2.1. Signal Model

Several factors, including specific application scenarios, transmission distances, and bandwidth limitations, need to be considered when selecting modulation schemes in underwater acoustic communication. Different modulation schemes are suitable for different channel conditions. This paper considers common underwater acoustic modulation signals such as BPSK, QPSK, 2FSK, 4FSK, and OFDM. BPSK is generally suitable for poor channel conditions, such as high noise environments or strong fading channels. BPSK's two-phase states provide high tolerance to the phase offset and inter-symbol interference, leading to better error rate performance in low SNR environments. QPSK is suitable for relatively good channel conditions, such as moderate signal-to-noise ratio environments. QPSK divides the signal into four phase states, allowing for the transmission of more information under the same bandwidth and power conditions. Therefore, in good channel conditions, QPSK can provide higher transmission rates. Both 2FSK and 4FSK belong to frequency shift keying modulation and are suitable for channel conditions with pronounced frequency selective fading. As underwater acoustic channels are afflicted by frequency selective fading, FSK modulation schemes adapt to the frequency variations of the channel by changing the carrier frequency, thereby improving communication reliability. Among them, 4FSK can transmit more information under the same bandwidth and power conditions by dividing the signal into four frequency states, thus providing higher data transmission rates compared to 2FSK.OFDM is appropriate for channel conditions characterized by frequency selective fading and multipath propagation. As underwater acoustic channels are susceptible to these issues, OFDM segregates the spectrum into multiple subcarriers and transmits them orthogonally, effectively minimizing inter-symbol interference caused by multipath propagation, thereby enhancing anti-interference performance and spectral efficiency.

When there is no relative movement between the transmitter and receiver, the underwater acoustic channel can be modeled as a coherent multipath UWAC channel. In this scenario, the received signal on a single channel can be expressed as follows:

$$r(t) = h(t) \otimes s(t) + n(t) \tag{1}$$

Here, r(t) denotes the signal received, h(t) represents the response of the system to an impulse of the UWAC channel, s(t) represents the transmitted passband modulated signal,  $\otimes$  denotes the convolution operation, and n(t) represents the ambient noise in the ocean environment.

If the transmitted signal s(t) is the MPSK signal, its mathematical expression is as follows:

$$s(t) = A\cos[2\pi f_c t + 2\pi k/M + \theta]$$
<sup>(2)</sup>

In this equation, *A* represents the magnitude of the signal,  $f_c$  signifies the carrier frequency, *k* is the symbol index, *M* stands for the number of symbols, and  $\theta$  denotes the initial phase.

If s(t) is the MFSK signal, its mathematical expression is as follows:

$$s(t) = A\cos[2\pi f_1 t + \varphi_1(k)] + A\cos[2\pi f_2 t + \varphi_2(k)] + \dots + A\cos[2\pi f_M t + \varphi_M(k)]$$
(3)

where *A* is the signal amplitude,  $f_1, f_2, ..., f_M$  are the carrier frequencies,  $\varphi_1(k), \varphi_2(k), ..., \varphi_M(k)$  represent the phase shifts corresponding to the *k*-th symbol, while *M* denotes the total number of symbols.

If the transmitted signal s(t) is an OFDM signal, its mathematical expression is as follows:

$$s(t) = \sum_{k=0}^{N-1} x_k(t) e^{j2\pi f_k t}$$
(4)

In this equation,  $x_k(t)$  signifies the complex-valued symbol transmitted on the *k*-th subcarrier at time *t*,  $f_k$  denotes the frequency of the *k*-th subcarrier, *N* represents the total number of subcarriers, and *j* stands for the imaginary unit.

To obtain  $x_k(t)$  in the time domain, the Fourier inverse transform of the corresponding frequency-domain symbol  $X_k$  is taken:

$$x_k(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_k e^{j2\pi kn/N}$$
(5)

In this equation,  $X_k$  denotes the complex-valued symbol transmitted on the *k*-th subcarrier in the spectral domain.

The OFDM signal is formed by concatenating the time-domain symbols for all subcarriers:

$$s(t) = \sum_{k=0}^{N-1} \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_k e^{j2\pi kn/N} e^{j2\pi f_k t}$$
(6)

Usually, the OFDM signal is transmitted through a frequency-selective channel, where each subcarrier undergoes diverse channel gain and phase shift. At the receiver, the signal is demodulated by performing a Fourier transform to recover the frequency-domain symbols  $X_k$  that are then employed to retrieve the original data.

#### 2.2. Underwater Acoustic Channel Analysis

Frequency selective channel refers to a channel in which the transmitted signal undergoes frequency selective fading. This type of channel is characterized by different degrees of signal attenuation at different frequencies, resulting in changes in the transmitted signal's spectrum. Multipath propagation refers to the phenomenon where the signal reaches the receiver through multiple paths during transmission, with each path having different propagation distances and propagation times. This leads to the generation of multiple different time delays and amplitudes at the receiver. This phenomenon causes the signals to superimpose, resulting in multipath interference. For frequency-selective channels, multipath propagation can cause different degrees of interference on signals at different frequencies, resulting in spectrum expansion and distortion. From the perspective of ray acoustics, the signal is emitted from the sound source and travels along different paths to reach the receiver. At the recipient's end, the composite received signal arises from the combination and interference of the transmitted signals across all conceivable acoustic pathways. For simplicity, we assume that each path has sufficient stability during a certain communication duration. Under this assumption, the underwater acoustic channel can be approximated as a static time-invariant channel, and its corresponding impulse response is a one-dimensional function with a delay  $\tau$ , denoted as follows:

$$c(\tau) = \sum_{p=1}^{P} \alpha_p e^{j\phi_p} \delta(\tau - \tau_p)$$
(7)

In Equation (7),  $\delta(\cdot)$  represents the Dirac impulse function, *P* represents the number of paths in the channel, and  $\alpha_p$ ,  $\phi_p$ ,  $\tau_p$  signify the magnitude, phase, and delay of the *p*-th path, respectively. The relevant path parameters can be obtained by solving the eigenray

equations as mentioned earlier. Then, the channel response output signal is obtained by convolving it with the transmitted signal.

$$r(t) = \int_{-\infty}^{\infty} s(t-\tau)c(\tau)d\tau$$
(8)

By substituting  $c(\tau)$  into the equation above, we obtain

$$r(t) = \sum_{p=1}^{P} \alpha_p \mathbf{e}^{\mathbf{j}\phi_p} s(\tau - \tau_p)$$
<sup>(9)</sup>

In Equation (9), s(t) and r(t) indicate the emitted signal and the received signal, respectively, following transmission through the underwater acoustic channel. The Doppler shift refers to the change in frequency of a signal when either the source or the receiver is in motion. When the source or the receiver moves toward the other, the signal frequency increases; when the source or the receiver moves away, the signal frequency decreases. In the underwater acoustic channel, the motion of a vessel or sonar equipment can cause the Doppler shift. The Doppler shift affects the spectral composition of the transmitted signal, resulting in changes in the received signal's spectrum. When primarily considering this factor, Doppler shift  $\Delta f$  can be expressed by the following formula:

$$\Delta f = \delta f_c = v_r f_c / c \tag{10}$$

In Equation (10),  $v_r$  signifies the relative speed between the sender and receiver, and can be expressed as  $v_T = v_T \cos \theta_T + v_0 \cos \theta_0$ . Here,  $v_T$  and  $v_0$  denote the velocities of the transmitter and receiver, respectively, while  $\theta_T$  and  $\theta_0$  represent the angles between the transmitter and receiver velocities and the direction of sound propagation. Moreover,  $f_c$  denotes the carrier frequency,  $\delta$  signifies the Doppler factor, and c denotes the speed of sound. As the velocity of sound in the ocean is relatively low (approximately 1500 m/s) compared to terrestrial wireless communication systems, the Doppler effect has a more significant influence on underwater acoustic communication. The Doppler effect introduces time-varying characteristics to the channel response. In underwater communication systems, the channel's time-varying nature, to some extent, can prevent the receiver from recovering the carrier, making it impossible to achieve high-rate communication using coherent transmission. Fading refers to the amplitude attenuation that signals experience during transmission. In underwater acoustic channels, fading can be caused by various factors such as sound propagation loss, scattering, and the Doppler effect. Fading leads to changes in the amplitude of the received signal, thereby affecting signal quality and reliability. In summary, frequency-selective channels in underwater communication are affected by multipath propagation, Doppler shift, and fading. Multipath propagation causes signal spectrum expansion and distortion, Doppler shift results in changes to the signal spectrum, and fading causes changes in signal amplitude. These effects have significant impacts on the transmission quality and reliability of signals.

# 2.3. Theoretical Analysis of Higher-Order Cumulant

### 2.3.1. Higher-Order Cumulant

Higher-order statistical analysis of signals, often termed non-Gaussian waveform processing, refers to signal analysis utilizing higher-order statistical moments. Classical signal processing approaches employ second-order statistical moments as mathematical tools for analysis, which are represented in the temporal realm as correlation functions and in the spectral domain as power spectra. However, these methods have certain limitations, such as multiplicity or equivalency, and cannot identify non-minimum phase systems. Moreover, they are susceptible to the influence of additive noise and can only handle signal data with additive white noise. To address these limitations, higher-order statistical moments are required. While analysis based on second-order statistical measures can solely

capture the primary information of the signal, namely the outline, higher-order statistical analysis can offer more intricate insights into the signal [22]. Consequently, this paper employs higher-order statistical moments as mathematical tools for modulation recognition.

Reference [22] provides some derivations: Given a random variable *x*, the initial characteristic function (also known as the moment generating function) is characterized as the Fourier inverse transform of its probability density function.

$$\phi(\omega) = \int_{-\infty}^{\infty} f(x) e^{j\omega x} dx = E\{e^{j\omega x}\}$$
(11)

The cumulant generating function, alternatively referred to as the second momentgenerating function, is defined as follows:

$$\psi(\omega) = \ln(\phi(\omega)) \tag{12}$$

Let the *k*-th derivative of the initial characteristic function be denoted as follows:

$$\phi^{k}(\omega) = \frac{d^{k}\phi(\omega)}{d\omega^{k}} = j^{k} \mathbb{E}\{x^{k} \mathrm{e}^{\mathrm{j}\omega x}\}$$
(13)

By setting  $\omega = 0$ , we can derive the *k*-th moment of *x* as follows:

$$m_{k} = \mathbf{E}\{x^{k}\} = (-j)^{k} \frac{d^{k} \phi(\omega)}{d\omega^{k}} \bigg|_{\omega=0} = (-j)^{k} \phi^{(k)}(0)$$
(14)

The delineation of the *k*th-order cumulant of *x* is

$$c_{kx} = (-j)^k \left. \frac{d^k \ln(\phi(\omega))}{d\omega^k} \right|_{\omega=0} = (-j)^k \psi^{(k)}(0)$$
(15)

We consider a Gaussian random variable *x* with a distribution, described as  $x \sim N(0, \sigma^2)$ , where its probability density function is expressed as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{x^2}{2\sigma^2})$$
(16)

Thus, the generating function of moments for the Gaussian random variable *x* can be formulated as follows:

$$\phi(\omega) = \int_{-\infty}^{\infty} f(x)e^{j\omega x} dx = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp(-\frac{x^2}{2\sigma^2} + j\omega x) dx \tag{17}$$

According to the integral transformation formula, the equation can be transformed as follows:

$$\int_{-\infty}^{\infty} \exp(-Ax^2 \pm 2Bx - C)dx = \sqrt{\frac{\pi}{A}}\exp(-\frac{AC - B^2}{A})$$
(18)

Let us consider the values  $A = \frac{1}{2\sigma^2}$ ,  $B = \frac{j\omega}{2}$ , and C = 0. In this case, the moment generating function of *x* can be expressed as follows:

$$\phi(\omega) = e^{-\sigma^2 \omega^2/2} \tag{19}$$

In this case, the various derivatives of  $\phi(\omega)$  can be computed as follows:

$$\phi'(\omega) = -\sigma^2 \omega e^{-\sigma^2 \omega^2/2} \tag{20}$$

$$\phi''(\omega) = (\sigma^4 \omega^2 - \sigma^2) e^{-\sigma^2 \omega^2/2}$$
(21)

$$\phi^{(3)}(\omega) = (3\sigma^4\omega - \sigma^6\omega^3)e^{-\sigma^2\omega^2/2}$$
(22)

$$\phi^{(4)}(\omega) = (3\sigma^4 - 6\sigma^6\omega^2 + \sigma^8\omega^4)e^{-\sigma^2\omega^2/2}.....$$
(23)

Therefore, the *k*-th order statistical moment of the Gaussian random variable *x* can be represented as follows:

$$m_1 = 0 \tag{24}$$

$$m_2 = \sigma^2 \tag{25}$$

$$m_3 = 0 \tag{26}$$

$$m_4 = 3\sigma^4 \tag{27}$$

Consequently, it can be generalized that for any arbitrary integer *k*, the moment of the random variable *x* following a Gaussian distribution is given by

$$m_{k} = \begin{cases} 0, k = \text{Oddnumber} \\ 1 \cdot 3 \cdots (k-1)\sigma^{k}, k = \text{Evennumber} \end{cases}$$
(28)

Subsequently, the cumulative generating function of *x* can be derived as follows:

$$\varphi(\omega) = \ln(\phi(\omega)) = -\frac{\sigma^2 \omega^2}{2}$$
(29)

The derivatives of the cumulative generating function of *x* with respect to its orders are given by

$$\varphi'(\omega) = -\sigma^2 \omega \tag{30}$$

$$\varphi''(\omega) = -\sigma^2 \tag{31}$$

$$\varphi^{(k)}(\omega) = 0, k > 2 \tag{32}$$

Therefore, the *k*th-order cumulative moments of the Gaussian random variable *x* are given by

$$c_1 = 0 \tag{33}$$

$$c_2 = \sigma^2 \tag{34}$$

$$c_k = 0, k > 2 \tag{35}$$

Upon derivation, it becomes apparent that the variance of any zero-mean Gaussian random process is equivalent to its second-order cumulant, which is also equivalent to its variance  $\sigma^2$ . The odd-order statistical moments are consistently zero, while the even-order statistical moments are non-zero. Nevertheless, the higher-order cumulants (third-order and beyond) are always zero. This indicates that the second-order cumulant is susceptible to additive noise, whereas the higher-order cumulants (third-order and beyond) demonstrate anti-noise interference characteristics for Gaussian random processes, such as Gaussian-colored noise.

High-order statistical moments refer to moments greater than the second order, while high-order cumulative moments refer to cumulative moments greater than the second order. We denote the operation of calculating moments as  $mom(\cdot)$  and the operation of

calculating cumulative moments as  $cum(\cdot)$ . According to the definitions, we can derive some properties of high-order moments and high-order cumulative moments:

According to the reference [22], in the presence of additive Gaussian-colored noise, the high-order cumulant moments of the observation process mirror the high-order cumulant moments of the non-Gaussian signal. In other words, the high-order cumulant moments are insensitive or resistant to Gaussian-colored noise. While due to the absence of semi-invariance in high-order moments, the high-order statistical moments of the observation process might differ from those of the non-Gaussian signal, suggesting that high-order moments are susceptible to Gaussian noise. Therefore, in higher-order statistical analysis, high-order cumulant moments are commonly used as tools for analyzing and processing non-Gaussian signals, rather than high-order moments. In signal analysis, high-order cumulant moments provide more comprehensive information about signal characteristics compared to high-order moments, making them significant for the analysis and processing of non-Gaussian signals.

Common high-order cumulative moments include the third-order cumulative moment (skewness) and the fourth-order cumulative moment (kurtosis). Skewness reflects the degree of asymmetry in the signal distribution and can be used to describe its asymmetric characteristics. Kurtosis reflects the peakedness of the signal distribution and can be used to describe its peakedness characteristics.

In signal analysis, high-order cumulative moments can be used to identify and differentiate different types of signals. By calculating the high-order cumulative moments of different signals, non-Gaussianity and nonlinearity features can be extracted, enabling signal classification and recognition. Additionally, high-order cumulative moments can be used to detect abnormal conditions in signals, such as abnormal pulses or noise, as these abnormal signals often lead to significant changes in high-order cumulative moments.

In conclusion, high-order cumulative moments involve the accumulation computation of higher-order statistical quantities of signals, providing more detailed and accurate information about signal characteristics. They have a crucial role in the examination and manipulation of signals that are non-Gaussian and nonlinear. The application of high-order cumulative moments in signal processing, pattern recognition, and anomaly detection fields can help us better understand and utilize signal properties.

The *p*th-order mixed moment of the stationary stochastic process x(t) is defined as follows [22]:

$$M_{pq} = E[x(k)^{p-q} x^*(k)^q]$$
(36)

The second to eighth-order cumulants of x(t) can be obtained as follows [23]:

$$C_{20} = \operatorname{Cum}(x, x) = M_{20} = \operatorname{E}[x(k)^2]$$
(37)

$$C_{21} = \operatorname{Cum}(x, x^*) = M_{21} = \operatorname{E}[|x(k)|^2]$$
(38)

$$C_{40} = \operatorname{Cum}(x, x, x, x) = M_{40} - 3M_{20}^2$$
(39)

$$C_{41} = \operatorname{Cum}(x, x, x, x^*) = M_{41} - 3M_{20}M_{21}$$
(40)

$$C_{42} = \operatorname{Cum}(x, x, x^*, x^*) = M_{42} - |M_{20}|^2 - 2M_{21}^2$$
(41)

$$C_{60} = \operatorname{Cum}(x, x, x, x, x, x) = M_{60} - 15M_{40}M_{20} + 30M_{20}^3$$
(42)

$$C_{61} = \operatorname{Cum}(x, x, x, x, x, x^*) = M_{61} - 5M_{40}M_{21} - 10M_{20}M_{41} + 30M_{20}^3M_{21}^2$$
(43)

$$C_{63} = \operatorname{Cum}(x, x, x, x^*, x^*, x^*) = M_{63} - 6M_{41}M_{20} - 9M_{42}M_{21} + 18M_{20}^2M_{21} + 12M_{21}^3$$
(44)

$$C_{80} = \operatorname{Cum}(x, x, x, x, x, x, x, x) = M_{80} - 28M_{20}M_{60} - 35M_{40}^2 + 420M_{20}^2M_{40} - 630M_{20}^4$$
(45)

In the event that x(t) is a Gaussian process, the cumulants beyond the second order will always be zero, resulting in the higher-order cumulants being capable of efficiently attenuating Gaussian noise. The theoretical values of higher-order cumulants for various modulation techniques of baseband signals are detailed in Table 1 [24,25].

Table 1. Theoretical values of higher-order cumulants for various types of digital modulation signals.

Modulation Mode	<i>C</i> <sub>20</sub>	$ C_{21} $	$ C_{40} $	$ C_{41} $	$ C_{42} $	<i>C</i> <sub>60</sub>	$ C_{61} $	<i>C</i> <sub>63</sub>	<i>C</i> <sub>80</sub>
2ASK	E *	Е	2E <sup>2</sup>	2E <sup>2</sup>	2E <sup>2</sup>	16E <sup>3</sup>	16E <sup>3</sup>	13E <sup>3</sup>	$35E^4$
4ASK	E	Е	1.36E <sup>2</sup>	1.36E <sup>2</sup>	1.36E <sup>2</sup>	8.32E <sup>3</sup>	8.32E <sup>3</sup>	9.16E <sup>3</sup>	/
BPSK	E	Е	$2E^2$	$2E^2$	$2E^2$	16E <sup>3</sup>	16E <sup>3</sup>	13E <sup>3</sup>	$272E^{4}$
QPSK	0	Е	$E^2$	0	$E^2$	0	$4E^3$	$4E^3$	$34E^4$
8PSK	0	Е	0	0	$E^2$	0	0	$4E^3$	/
2FSK	0	Е	0	0	E <sup>2</sup>	0	0	$4E^3$	0
4FSK	0	Е	0	0	E <sup>2</sup>	0	0	3E <sup>3</sup>	/
8FSK	0	Е	0	0	E <sup>2</sup>	0	0	$3E^3$	/
16QAM	0	Е	0.68E <sup>2</sup>	0	0.68E <sup>2</sup>	0	2.08E <sup>3</sup>	2.08E <sup>3</sup>	$14E^4$
64QAM	0	Е	0.619E <sup>2</sup>	0	0.619E <sup>2</sup>	0	1.80E <sup>3</sup>	1.80E <sup>3</sup>	/
OFDM	0	0	0	0	0	0	0	0	Ö

\* where E is the power of the signal.

In OFDM systems, the spectrum is divided into multiple subcarriers, and modulation and transmission are performed on each subcarrier. To simplify system design and implementation complexity, OFDM assumes that the modulation sequences on each subcarrier are independent and identically distributed complex random sequences. This assumption is made to simplify the system design and implementation complexity. If the modulation sequences on each subcarrier are mutually independent and identically distributed complex random sequences, the same modulator and demodulator can be used to process each subcarrier, thus reducing system complexity. Additionally, this assumption simplifies the implementation of key techniques, such as channel estimation and equalization. The design and implementation of OFDM systems are closely related to the independence and identical distribution of the modulation sequences on each subcarrier. When designing an OFDM system, parameters such as the number of subcarriers, frequency spacing between subcarriers, and modulation scheme need to be determined. By assuming that the modulation sequences on the subcarriers are independent and identically distributed complex random sequences, system performance analysis and optimization can be conducted to select appropriate parameter settings. In the implementation of an OFDM system, the same modulator and demodulator can be used to process the modulation sequences on each subcarrier, simplifying the hardware design and software implementation of the system.

OFDM is a signal composed of multiple orthogonal subcarriers, with each subcarrier transmitting independent data. The frequency spacing between subcarriers in an OFDM system is orthogonal to each other, which helps avoid interference between subcarriers. The time-domain waveform of an OFDM signal is formed by adding multiple sinusoidal waves together. According to the central limit theorem, when many independent random variables with finite variances are linearly summed, the distribution of the sum is inclined to a normal distribution. In an OFDM signal, the data on each subcarrier can be considered as independent random variables, and due to the orthogonal frequency spacing between subcarriers, their linear sum can be approximated as a Gaussian distribution. Furthermore, OFDM signals are periodic in the time domain, while a Gaussian distribution is a stationary process that has infinite bandwidth in the spectral domain. As a result, the spectrum of an OFDM signal remains flat near the center frequency of each subcarrier and decays rapidly

at the frequency spacing between subcarriers. This spectral characteristic contributes to the improved spectral efficiency and resistance to multipath fading in OFDM systems. In conclusion, the orthogonality between subcarriers and the application of the central limit theorem results in the Gaussian distribution characteristics of OFDM signals in the spectral domain.

Regarding the received OFDM signal at the receiver, since the OFDM signal is a superposition of multiple orthogonal subcarriers, it is assumed that the modulation sequences on individual subcarriers follow independent and identically distributed complex random sequences. According to the central limit theorem of statistics, the OFDM communication signal follows an asymptotic complex Gaussian distribution. Furthermore, the Gaussian nature of OFDM is affected by the number of subcarriers, with stronger Gaussianity observed as the number of subcarriers increases [23]. This property is independent of the modulation scheme used on the subcarriers, whether it is BPSK or QPSK. For single-carrier modulation signals, they do not exhibit normal distribution characteristics, so their  $C_{42}$  and  $C_{63}$  are non-zero.

This paper constructs

$$\tilde{C}_{42}(x(n)) = \left| \frac{C_{42}}{C_{21}^2} \right|$$
(46)

$$C_{42} = 10\log_{10}(\tilde{C}_{42}(x(n)) + 1)$$
(47)

$$\tilde{C}_{63}(x(n)) = \left| \frac{C_{63}}{C_{21}^3} \right|$$
(48)

$$C_{63} = 10\log_{10}(\tilde{C}_{63}(x(n)) + 1)$$
(49)

In the baseband, the theoretical values of  $\tilde{C}_{42}(x(n))$  for BPSK, QPSK, 2FSK, 4FSK, and OFDM signals are 2, 1, 1, 1, and 0, respectively. The theoretical values of  $\tilde{C}_{63}(x(n))$  are 13, 4, 4, 3, and 0, respectively. The aim of creating  $C_{42}$  and  $C_{63}$  is to augment the separation between the data that correspond to the higher-order cumulants of distinct modulation signals. This is conducted to more effectively differentiate between various modulation schemes in the presence of noise. Compared to  $\tilde{C}_{42}(x(n))$  and  $C_{42}$ , due to the complex noise effects in the underwater acoustic channel, using the sixth-order cumulant will have higher robustness than employing the fourth-order cumulant, and  $C_{63}$  is used in this paper to distinguish between OFDM and other modulation schemes.

#### 2.3.2. High-Order Spectral Analysis

The higher-order spectrum, also known as the polyspectrum, refers to the spectrum containing multiple frequencies. More precisely, the third-order spectrum  $S_{3x}(\omega_1, \omega_2)$  is known as the bispectrum, while the fourth-order spectrum  $S_{4x}(\omega_1, \omega_2, \omega_3)$  is commonly referred to as the trispectrum. This is because they respectively represent the energy spectrum of two and three frequencies. It is customary to use  $B_x(\omega_1, \omega_2)$  to represent the bispectrum and  $T_x(\omega_1, \omega_2, \omega_3)$  to express the trispectrum. The implementation process of the bispectrum is derived below [22].

Consider zero-mean observed samples x(0), x(1), ..., x(N-1) with a sampling frequency of  $f_s$ . The data to be detected are partitioned into K segments, each with a length of M, denoted as  $x^{(k)}(0), x^{(k)}(1), ..., x^{(k)}(M-1)$ , where k = 1, ..., K, and overlapping between adjacent data segments is permitted.

We compute the coefficients of the discrete Fourier transform (DFT), where  $\lambda = 0, 1, ..., M/2$ and k = 1, ..., K.

$$X^{(k)}(\lambda) = \frac{1}{M} \sum_{n=0}^{M-1} x^{(k)}(n) \mathrm{e}^{-\mathrm{j}2\pi n\lambda/M}$$
(50)

We compute the triple correlation of DFT coefficients, where  $\Delta_0 = f_s / N_0$ , and  $N_0$  and  $L_1$  must meet the condition that  $M = (2L_1 + 1)N_0$ :

$$\hat{b}_{k}(\lambda_{1},\lambda_{2}) = \frac{1}{\Delta_{0}^{2}} \sum_{i_{1}=-L_{1}}^{L_{2}} \sum_{i_{2}=-L_{1}}^{L_{1}} X^{(k)}(\lambda_{1}+i_{1})X^{(k)}(\lambda_{2}+i_{2})X^{(k)}(-\lambda_{1}-\lambda_{2}-i_{1}-i_{2})$$
(51)

where  $k = 1, \cdots, K$ ;  $0 \le \lambda_2 \le \lambda_1, \lambda_1 + \lambda_2 \le f_s/2$ .

The average value of K-segment bispectrum estimation is the final bispectrum estimation value of the sample data [22].

$$\hat{B}_{D}(\omega_{1},\omega_{2}) = \frac{1}{K} \sum_{k=1}^{K} \hat{p}_{k}(\omega_{1},\omega_{2})$$
(52)

where  $\omega_1 = \frac{2\pi f_s}{N_0} \lambda_1, \omega_2 = \frac{2\pi f_s}{N_0} \lambda_2.$ 

#### 3. Proposed Methods

#### 3.1. System Framework

The configuration of the system for recognizing underwater communication signal modulation, which is based on higher-order cumulant theory and deep learning network, is illustrated in Figure 2. The system first utilizes the improved higher-order cumulant  $C_{63}$  value of OFDM, which is zero, while the  $C_{63}$  values of other phase modulation and frequency type signals are non-zero. By using this difference, OFDM and other modulation-type signals can be distinguished. The theoretical simulation results without noise are shown in Figure 3, and it can be seen that the feature distance between  $C_{42}$  and  $C_{63}$  is larger than that between  $\tilde{C}_{42}(x(n))$  and  $\tilde{C}_{63}(x(n))$ , resulting in a better distinguishing effect. Moreover, the distinguishing effect of  $C_{63}$  is better than that of  $C_{42}$ . Therefore, it is more appropriate for the system to use  $C_{63}$  to distinguish OFDM and other modulation types.



Figure 2. Block diagram of the modulation recognition system.

Then, using the theory of higher-order spectra characteristics of the signal, the bispectrum of signals with different modulation methods is plotted. Figure 4 shows the bispectrum generated under an SNR of 10 dB.



**Figure 3.** Higher-order cumulants under a theoretical simulation environment without noise. (a) The trend of  $\tilde{C}_{42}(x(n))$  with respect to the SNR; (b) the trend of  $C_{42}$  with respect to the SNR; (c) the trend of  $\tilde{C}_{63}(x(n))$  with respect to the SNR; (d) the trend of  $C_{63}$  with respect to the SNR.



Figure 4. The bispectrum of the signal. (a) BPSK; (b) QPSK; (c) 2FSK; (d) 4FSK.

Due to the fact that the bispectrum directly generated does not have prominent features and is easily obscured by noise, this paper has made improvements based on this and squared the signal before estimating its bispectrum. The enhanced bispectrum of the squared signal is presented in Figure 5. It is apparent that the distinctions between the spectra of various signals are more pronounced, making it suitable for identifying the modulation type in image recognition.



Figure 5. Squared bispectrum of the signal. (a) BPSK; (b) QPSK; (c) 2FSK; (d) 4FSK.

#### 3.2. Neural Network Model

Feedforward neural networks that encompass convolutional operations and possess a deep structure are commonly known as conventional convolutional neural networks (CNNs). The most representative of these is the LeNet network model proposed by Yan LeCun in 1998, which laid the foundation for modern CNNs and earned him the title of "father of convolutional neural networks". Figure 6 displays the LeNet network architecture, which consists of convolutional layers, activation layers, pooling layers, fully connected layers, and ultimately utilizes the softmax function to classify the resultant image [26].



Figure 6. LeNet network structure.

Nevertheless, due to the intricate nature of the UWAC channel environment and the impact of noise, the signal may experience distortion, and the critical feature expression in the squared bispectrum plot may also be obscured by noise. Consequently, classical CNN networks exhibit substandard recognition performance under low signal-to-noise ratios. One potential solution is to augment the depth of the network to gather a greater amount of information present in the images. However, blindly pursuing network depth can lead to problems such as gradient disappearance or explosion, and bring difficulties to manual parameter tuning of the network.

Therefore, this paper uses the ResNet model proposed by He, K. et al. in 2015 as the main structure [27]. The fundamental component of the network, ResBlock, is depicted in Figure 7. It creates "shortcut connections" between the upper and lower layers, which facilitates gradient backpropagation throughout the training process and trains deeper CNN networks. From the perspective of information transmission, this structure has advantages. In conventional convolutional neural networks, information loss may occur after convolution and other operations, and as the network's depth escalates, the loss of information becomes increasingly severe. This can result in issues such as gradient disappearance or explosion, rendering the network ineffective in training. In the design of ResBlock, through the direct bridging operation from the upper layer to the lower layer, the lower layer can receive the feature representation extracted by the convolution layer as well as the complete information from the upper layer, effectively reducing the loss of information during network transmission.

In ResNet, the skip connection operation, also known as the direct bridging operation, connects the input signal directly to the output signal, enabling the preservation and transmission of information across layers. This architecture aids in resolving the problems of dwindling and skyrocketing gradients in deep neural networks. By using skip connections, the input signal can bypass certain layers and be directly propagated to subsequent layers, thereby retaining more information and gradients.

The convolution operation is a commonly used feature extraction method in deep learning. Through convolution operation, neural networks can learn to extract local features from data such as images, speech, and text. The convolution kernel performs convolution operation on the input data by the sliding window, convolving the local features within the window with the convolution kernel, and obtaining feature maps. Through multiple convolution operations, neural networks can gradually extract more abstract and advanced features.

Batch normalization is a technique used to accelerate training and improve the performance of neural networks. It normalizes the input of each mini-batch, making the mean of each feature dimension close to 0 and the variance close to 1. This mitigates the challenges of gradient vanishing and exploding, while simultaneously enhancing the steadiness and rate of convergence of the network. In addition, batch normalization also has a certain regularization effect, which can reduce the risk of overfitting.

ReLU (rectified linear unit) activation function is a commonly used non-linear activation function. It sets the input signal to 0 for values less than 0, and keeps the values

greater than 0 unchanged. The ReLU activation function has a simple computational form and good non-linear expression ability, which helps neural networks learn more complex features and express more complex non-linear relationships. Furthermore, the ReLU activation function possesses sparse activation characteristics, which aid the network in learning more sparse feature representations and enhancing the network's generalization capability.

In summary, in ResNet, the skip connection operation achieves the preservation and transmission of information across layers, convolution operation extracts features, batch normalization accelerates training time and improves network performance, and the ReLU activation function enhances non-linear relationships, thereby improving the performance and generalization ability of neural networks.

In ResBlock, the information is divided into two paths. One path directly proceeds to the subsequent layer, while the other path traverses through a convolution layer with a kernel size of  $3 \times 3$ , a stride of 1, and a padding of 1, succeeded by a batch normalization layer and a ReLU activation layer. The batch normalization layer is employed to expedite the training process and enhance the model's generalization capability, whereas the ReLU activation function is a frequently used nonlinear activation function that amplifies the network's nonlinear expression ability. Among these, the in-place setting of the ReLU activation function is set to True, which implies that it operates in place and can conserve memory usage.



Figure 7. ResBlock.

The overall residual network structure is depicted in Figure 8. Because the bispectrum plot has central symmetry, the actual input of the network is in the lower right corner of the cropped image, which accounts for 1/4 of the original image. The image passes through a convolution layer equipped with a  $3 \times 3$  kernel size, a stride of 1, and a padding of 1, subsequently accompanied by a batch normalization layer and a ReLU activation layer. Then it enters the residual network layer, which is composed of three residual blocks, as mentioned earlier, for feature extraction. This structure is deeper than a CNN network. Finally, average pooling is performed, and the recognition result is output using the softmax function through a fully connected layer. The network used the cross-entropy loss function to measure the difference between the true values and the model's predicted values for image classification tasks.



Figure 8. ResNet network structure.

Table 2 illustrates the network parameters utilized in this paper.

Table 2. Transmission signal parameters in the simulation environment.

Structure	Layer Name	Input Channels	Output Channels	Stride	Padding
	$Conv2d(3 \times 3)$	16	16	1	1
Input	BatchNorm2d	/	16	/	/
	ReLU	inplace = True			
ResBlock $\times$ 3	$Conv2d(3 \times 3)$	16	16	1	1
	BatchNorm2d	/	16	/	/
	ReLU	inplace = True			
	$Conv2d(3 \times 3)$	-1 *	16	1	1
	BatchNorm2d	/	16	/	/
	ReLU	inplace = True			
Output	AvgPool2d	kernel_size = 3			
	Full connection_1	16,384	256	/	/
	Full connection_2	256	4	/	/

\* The parameter represented by -1 varies with the number of channels in the previous layer's output.

The neural network in this study was trained and evaluated on an Ubuntu system with a 12th Gen Intel(R) Core(TM) i9-12900KF CPU and a GeForce RTX 3090 GPU, with 64 GB of graphics memory. The environment was established using the PyTorch deep learning network platform framework. During the training process, the Adam optimizer with a learning rate of 0.001 was employed, and the batch size for both the training and testing datasets was set to 64. The network was trained for 100 epochs.

#### 4. Experiment and Verification

In this paper, a training set generated under a theoretical simulation environment was constructed, and test sets generated under a theoretical simulation environment, a Bellhop underwater acoustic simulation channel environment, and a real lake trial environment were used for testing. The empirical findings demonstrate the strength of the system in blind recognition of received signals in non-cooperative communication scenarios.

#### 4.1. Simulation Verification

In this paper, we constructed training and test sets in a theoretical simulation environment, and the parameters of the transmitted signals are shown in Table 3.

To make the simulation results closer to the real environment and mitigate the disparity between the training and actual test datasets, after the transmission signals were generated, they passed through a Rayleigh fading channel with a multipath number of 7 and added in-band colored noise with a bandwidth of 4000–8000 Hz, rather than simple Gaussian white noise. This can better simulate the real underwater acoustic channel environment, enhance the training set's capacity for generalization, and enhance the persuasiveness of the simulation results.

Signal Frequency Band:	4000–8000 Hz					
Sampling Frequency:	48,000 Hz					
Modulation Mode	Source Length/bit	Data Rate/bps	Code Width/ms	Signal Duration/s		
BPSK QPSK 2FSK 4FSK OFDM-BPSK OFDM-QPSK	40,000 80,000 500 500 28,000 56,000	4000 8000 47 47 2678 5357	$\begin{array}{c} 0.25 \\ 0.25 \\ 21.00 \\ 42.00 \\ 16.00 \\ 16.00 \end{array}$	$\begin{array}{c} 10.01 \\ 10.01 \\ 10.68 \\ 10.67 \\ 10.46 \\ 10.45 \end{array}$		

Table 3. Transmission signal parameters in a simulation environment.

As shown in Figure 9, the signal passed through a channel with a multipath number of 7 and added in-band colored noise, and the SNR (SNR) range varied from -10 dB to 20 dB, gradually increasing in 2 dB steps. The regularity of  $\tilde{C}_{42}(x(n))$ ,  $C_{42}$ ,  $\tilde{C}_{63}(x(n))$  and  $C_{63}$ , estimated as SNR changes, can be used to set a classification threshold of 0.7, which can distinguish OFDM from BPSK, QPSK, 2FSK, and 4FSK signals when the SNR is 0 dB. Comparing (a) and (b), and (c) and (d), it can be found that even after adding noise,  $C_{42}$  and  $C_{63}$  still have a larger feature spacing than  $\tilde{C}_{42}(x(n))$  and  $\tilde{C}_{63}(x(n))$ . Comparing (b) and (d), it can be seen that the discrimination effect of  $C_{63}$  is better than that of  $C_{42}$ .



**Figure 9.** Higher-order cumulants in a theoretical simulation environment with noise and multipath effects. (a)  $\tilde{C}_{42}(x(n))$  trend with SNR changes; (b)  $C_{42}$  trend with SNR changes; (c)  $\tilde{C}_{63}(x(n))$  trend with SNR changes; (d)  $C_{63}$  trend with SNR changes.

As shown in Figure 10, the squared spectrogram of the signal was plotted after adding multipath effects and in-band colored noise with SNRs of -10 dB, 0 dB, 10 dB, and 20 dB. The constructed training and test sets consisted of 12,800 and 640 images, respectively, which were input into ResNet for training and testing. The network's testing recognition outcomes were exhibited in the shape of a confusion matrix, as illustrated in Figure 11. It is discernible that ResNet achieved 100% recognition accuracy for BPSK, QPSK, 2FSK, and 4FSK signals at an SNR of 0 dB.



**Figure 10.** Squared bispectrogram generated at an SNR of 10dB. (**a**) BPSK; (**b**) QPSK; (**c**) 2FSK; (**d**) 4FSK.

As shown in Figure 12, the curves of the recognition accuracies of ResNet and LeNet networks with SNR changes are presented. It is evident that ResNet outperforms LeNet under low SNR conditions. LeNet achieves a 100% recognition rate at an SNR of 2 dB, while ResNet achieves a 100% recognition rate at an SNR of 0 dB. To summarize, it can be observed that ResNet achieves 100% intra-class recognition accuracy at an SNR of 0 dB for BPSK, QPSK, 2FSK, and 4FSK signal classifications. The intra-class recognition accuracy is around 90% for SNR > -3 dB and reaches 80% for SNR > -8 dB.



**Figure 11.** The testing confusion matrix at different SNRs under theoretical simulation conditions. (a) SNR = -10 dB; (b) SNR = 0 dB; (c) SNR = 10 dB; (d) SNR = 20 dB.



Figure 12. Network recognition accuracy curve under theoretical simulation conditions.

The training and testing processes of the network with an SNR of 20 dB are presented in Figure 13. It is apparent from Figure 13a that the ResNet achieves a testing accuracy that tends to be 100% from the beginning, while the testing accuracy of LeNet fluctuates significantly. Moreover, as shown in Figure 13b,c, the testing loss and training loss demonstrate that ResNet can converge rapidly and achieve optimal performance compared to LeNet.



**Figure 13.** The training and testing process of the network at an SNR of 20 dB. (**a**) Test accuracy curve; (**b**) test loss curve; (**c**) train loss curve.

This paper further demonstrates the effectiveness of the proposed method by comparing it with the methods proposed in references [28,29]. Reference [28] used the original bispectrum of the signal as the dataset for the deep learning neural network, directly estimating the bispectrum of the signal. Reference [29] used the smooth pseud Wigner-Ville distribution (SPWVD) to plot the time–frequency diagram of the signal as the dataset for the neural network. Figure 14 shows the performance of the three methods in the inter-class recognition of PSK, FSK, and OFDM signals in a computer simulation environment. Both the original bispectrum and SPWVD methods also exhibit good performance in recognizing PSK, FSK, and OFDM signals, with an inter-class recognition accuracy of over 90% when SNR > 0 dB. The proposed squared bispectrum method achieves 100% inter-class recognition accuracy when SNR  $\geq -4$  dB. Figure 15 shows the performance of the three methods in intra-class recognition of BPSK, QPSK, 2FSK, and 4FSK signals, highlighting the superiority of the proposed method. The average recognition accuracy of the original bispectrum method is 78.91%, while the average recognition accuracy of the SPWVD time–frequency method is 67.42%.



Figure 14. Inter-class recognition comparison curves under theoretical simulation conditions.



Figure 15. Intra-class recognition comparison curves under theoretical simulation conditions.

#### 4.2. Bellhop Dataset Simulation Validation

In this work, a Bellhop simulation model was used to generate underwater acoustic channels. The simulated source depth was set to 100 m, the receiver depth was set to 50 m, and the lateral communication distance between the transmitter and receiver was 1 km. The sound ray diagram of the simulated water is shown in Figure 16, and the generated channel impulse response is shown in Figure 17. It can be seen that the channel received by the receiver comprises 7 multipaths.



Figure 16. Sound ray diagram between the transmitter and receiver in the Bellhop.



Figure 17. Impulse response of the UWAC channel in the Bellhop simulation.

As shown in Figure 18, the signal passed through a Bellhop channel and added colored noise within the bandwidth with a step size of 2 dB, ranging from 10 dB to 20 dB. The trends of estimated  $\tilde{C}_{63}(x(n))$  and  $C_{63}$  with the change of SNR were obtained. The results show that the classification threshold set to 0.7 can distinguish OFDM from BPSK, QPSK, 2FSK, 4FSK and other modulated signals at an SNR of 4 dB in the Bellhop simulated underwater acoustic channel.



**Figure 18.** Higher-order cumulants in the Bellhop-simulated channel environment. (**a**) The trend of  $\tilde{C}_{63}(x(n))$  with the change of SNR. (**b**) The trend of  $C_{63}$  with the change of SNR.

Using the training set of 12,800 images generated in a theoretical simulation environment, the test set of 640 images generated by Bellhop was tested. The squared spectrogram generated by the Bellhop channel is shown in Figure 19. The testing recognition outcomes are presented in Figure 20. It is evident that ResNet can precisely discern BPSK, QPSK, 2FSK, and 4FSK signals with 100% accuracy at an SNR of 2 dB. The effectiveness is basically the same as that in the theoretical simulation environment, indicating that the quality of the training set generated by the simulation meets the simulation requirements of complex underwater acoustic channel environments.



Figure 19. Squared spectrogram generated at an SNR of 10 dB. (a) BPSK; (b) QPSK; (c) 2FSK; (d) 4FSK.



**Figure 20.** The testing confusion matrix at different SNRs under Bellhop simulation conditions. (a) SNR = -10 dB. (b) SNR = 0 dB. (c) SNR = 2 dB. (d) SNR = 20 dB.

As shown in Figure 21, the test set generated under the Bellhop environment is more affected by the simulated underwater acoustic channel noise than the test set generated in a theoretical simulation environment. In the range of -10 dB to 0 dB, the recognition accuracies of ResNet and LeNet networks decreased, but overall, ResNet has better signal modulation-type recognition performance than LeNet. ResNet can achieve 100% recognition accuracy at an SNR of 2 dB, while LeNet still fluctuates in recognition accuracy in environments with SNRs above 6 dB and cannot reach 100% accuracy. This indicates that ResNet is more robust than LeNet in terms of changes in SNRs.



Figure 21. Network recognition accuracy curve under the Bellhop-simulated channel conditions.

Figure 22 shows the performances of the three methods in inter-class recognition of PSK, FSK, and OFDM signals in a Bellhop-simulated underwater acoustic channel environment. Figure 23 demonstrates the performances of the three methods in intra-class recognition of BPSK, QPSK, 2FSK, and 4FSK signals. Further validation of the results in the computer simulation environment shows that both the original bispectrum and SPWVD time–frequency methods achieve inter-class recognition accuracy of over 90% for PSK, FSK, and OFDM signals at 2 dB, slightly outperforming the proposed squared bispectrum method. In intra-class recognition, the average recognition accuracy of the original bispectrum method is 75.92%, while the average recognition accuracy of the SPWVD time–frequency method is 65.15%, highlighting the superiority of the proposed method in this paper.



Figure 22. Inter-class recognition comparison curves under Bellhop simulation conditions.



Figure 23. Intra-class recognition comparison curves under Bellhop simulation conditions.

#### 4.3. Lake-Test Dataset Verification

In September 2018, a field experiment was conducted at Danjiangkou Reservoir, Danjiangkou City, Hubei Province, China, to verify the recognition performance of the modulation recognition system proposed in this paper under practical underwater channel conditions. The experiment was conducted on the lake and the communication distance was 1 km. The arrangement of the transmitting and receiving apparatus is illustrated in Figure 24. The sound source at the emitting end was directly connected to the transducer via a long cable using a laptop and power amplifier to generate sound. The receiving end used a mid-frequency underwater acoustic communication integrated machine developed by our research team, as shown in Figure 25.



Figure 24. Layout of the transmission and reception devices.



Figure 25. Self-developed hydrophone.

The parameters of the transmitting signal during the lake trial are shown in Table 4.

Information Source:	526 Character English Short Sentence					
Signal Frequency Band:	4000–8000 Hz					
Synchronization Head	LFM Signal with a Width of 0.1 s					
Modulation Mode	Source Length/bit	Data Rate/bps	Code Width/ms	Signal Duration/s		
BPSK QPSK 2FSK 4FSK OFDM-BPSK OFDM-OPSK	20,000 80,000 500 80,000 80,000 80,000	4000 8000 10 10 3200 6400	$\begin{array}{c} 0.25 \\ 0.25 \\ 100.00 \\ 200.00 \\ 21.00 \\ 21.00 \end{array}$	$5.40 \\ 10.40 \\ 50.40 \\ 50.40 \\ 25.25 \\ 12.85$		

Table 4. Modulation parameters of the signal transmitted during the lake trial.

Generate wav file, 48 kHz sampling.

In this study, operations such as searching for synchronization headers and cropping were performed on the received data. The signal sampling frequency was 48,000 Hz. The length of the BPSK signal was 5 s, the length of the QPSK signal was 10 s, the length of the 2FSK signal was 10 s, the length of the 4FSK signal was 50 s, the length of the OFDM-BPSK signal was 25 s, and the length of the OFDM-QPSK signal was 11 s. A square bispectrogram was generated every 0.3 s, resulting in a total of 370 test set images.

As depicted in Figure 26, the improved sixth-order cumulative quantity  $C_{63}$  threshold was set to 0.7, which was found to be feasible based on the measured data collected during the lake trial. The system was able to effectively distinguish between OFDM and other modulation types.



**Figure 26.** Advanced cumulative quantity  $C_{63}$  of the lake trial data.

As shown in Figure 27, the square bispectrum diagram generated from the offline processing of the lake trial data exhibits little difference compared to the simulated environment, indicating that the network still maintains its effectiveness in identifying the lake trial data.



**Figure 27.** Square bispectrum diagram generated from the lake trial data. (a) BPSK; (b) QPSK; (c) 2FSK; (d) 4FSK.

As illustrated in Figure 28, the recognition outcomes of ResNet and LeNet on the lake trial data demonstrate that ResNet has effectively accomplished modulation identification of BPSK, QPSK, 2FSK, and 4FSK signals with a recognition accuracy of 98.44%, whereas LeNet achieved a recognition accuracy of 90.00%. Therefore, ResNet outperforms LeNet in terms of performance.



**Figure 28.** The testing confusion matrix for the recognition of the lake trial received signals. (**a**) ResNet; (**b**) LeNet.

#### 5. Conclusions

This paper proposes a modulation recognition system for UWAC signals using highorder cumulants and deep learning networks. By utilizing the difference in high-order cumulant values  $C_{63}$  of signals with different modulation types, the OFDM signal is distinguished from other modulation types. Subsequently, by leveraging the distinctions in the square bispectrum diagrams of signals with varying modulation types, a deep learning network is employed for the visual recognition of distinct images, effectively recognizing the modulation type of UWAC signals.

The simulation results show that with the improved sixth-order cumulant and a multipath number of 7, the system can distinguish between OFDM and BPSK, QPSK, 2FSK, and 4FSK signals under 0 dB conditions with a recognition accuracy of 100%. With the ResNet neural network recognizing the improved bispectrum diagram, the system can distinguish between BPSK, QPSK, 2FSK, and 4FSK modulation signals under 0 dB conditions, with a recognition accuracy of 100% in the simulation environment. In the UWAC simulation channel generated by BELLHOP, the system achieved recognition accuracy of 100% at an SNR of 2 dB, verifying the feasibility of the simulation results. In the real lake trial environment at Danjiangkou, the recognition accuracy of the lake trial data was 98.44%, proving that the proposed system, which combines the high-order cumulant theory with artificial intelligence, can achieve blind modulation recognition of common UWAC signals in non-cooperative scenarios under low SNR and underwater multipath conditions.

Author Contributions: Conceptualization, R.Z., C.H. and L.J.; methodology, R.Z., C.H. and C.Z.; software, R.Z., C.Z. and C.L.; validation, R.Z., C.H., L.J., C.Z., C.L. and J.L.; formal analysis, R.Z. and J.L.; investigation, R.Z. and C.H.; resources, C.H. and C.Z.; data curation, R.Z. and J.L.; writing—original draft preparation, R.Z.; writing—review and editing, R.Z., C.H. and L.J.; visualization, R.Z.; supervision, J.L. and C.L.; project administration, C.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (62071383) and the Fundamental Research Funds for Central Universities (23GH02027).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The findings presented in this article were made possible through the diligent guidance of the authors' supervisor and the invaluable assistance of classmates, as well as the care and help of the teachers in the research group. With their collective help, the authors were able to obtain the lake test data. The unwavering support of the family and girlfriend of one of the authors was also instrumental. The authors extend sincere gratitude to all of them.

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

#### Abbreviations

The following abbreviations are used in this manuscript:

2FSK	2-Frequency Shift Keying
4FSK	4-Frequency Shift Keying
AMC	Automatic Modulation Classification
BPSK	Binary Phase Shift Keying
CNN	Convolutional Neural Networks
CV	Computer Vision
DL	Deep Learning
FB	feature-based
LB	likelihood-based
ML	Machine Learning
OFDM	Orthogonal Frequency Division Multiplexing
QPSK	Quadrature Phase Shift Keying
ResNet	Residual Network
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
UWAC	Underwater Acoustic Communication

#### References

- 1. Peurifoy, J.; Shen, Y.; Jing, L.; Yang, Y.; Cano-Renteria, F.; DeLacy, B.G.; Joannopoulos, J.D.; Tegmark, M.; Soljačić, M. Nanophotonic particle simulation and inverse design using artificial neural networks. *Sci. Adv.* **2018**, *4*, eaar4206. [CrossRef] [PubMed]
- 2. Sheverdin, A.; Monticone, F.; Valagiannopoulos, C. Photonic inverse design with neural networks: The case of invisibility in the visible. *Phys. Rev. Appl.* **2020**, *14*, 024054. [CrossRef]
- 3. Lin, X.; Eldemerdash, Y.A.; Dobre, O.A.; Zhang, S.; Li, C. Modulation classification using received signal's amplitude distribution for coherent receivers. *IEEE Photonics Technol. Lett.* **2017**, *29*, 1872–1875. [CrossRef]
- Si, H.; Hu, X.; Shi, Z.; Li, S. Non-cooperative signal modulation recognition algorithm based on joint feature parameter extraction. J. Commun. 2020, 41, 172–185. (In Chinese)
- 5. Xie, L.; Wan, Q. Cyclic feature-based modulation recognition using compressive sensing. *IEEE Wirel. Commun. Lett.* 2017, 6, 402–405. [CrossRef]
- Jajoo, G.; Kumar, Y.; Yadav, S.K. Blind signal PSK/QAM recognition using clustering analysis of constellation signature in flat fading channel. *IEEE Commun. Lett.* 2019, 23, 1853–1856. [CrossRef]

- 7. Wu, Z.; Zhou, S.; Yin, Z.; Ma, B.; Yang, Z. Robust automatic modulation classification under varying noise conditions. *IEEE Access* 2017, *5*, 19733–19741. [CrossRef]
- 8. Wang, Y.; Gui, G.; Ohtsuki, T.; Adachi, F. Multi-task learning for generalized automatic modulation classification under non-Gaussian noise with varying SNR conditions. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 3587–3596. [CrossRef]
- 9. Wang, Y.; Liu, M.; Yang, J.; Gui, G. Data-driven deep learning for automatic modulation recognition in cognitive radios. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4074–4077. [CrossRef]
- 10. Zheng, S.; Qi, P.; Chen, S.; Yang, X. Fusion methods for CNN-based automatic modulation classification. *IEEE Access* 2019, 7, 66496–66504. [CrossRef]
- Nandi, A.K.; Azzouz, E.E. Algorithms for automatic modulation recognition of communication signals. *IEEE Trans. Commun.* 1998, 46, 431–436. [CrossRef]
- O'Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. In Proceedings of the Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, 2–5 September 2016; Proceedings 17; Springer: Berlin/Heidelberg, Germany, 2016; pp. 213–226.
- O'Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-air deep learning based radio signal classification. *IEEE J. Sel. Top. Signal Process.* 2018, 12, 168–179. [CrossRef]
- Jeong, S.; Lee, U.; Kim, S.C. Spectrogram-based automatic modulation recognition using convolutional neural network. In Proceedings of the 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN), Prague, Czech Republic, 3–6 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 843–845.
- 15. Zhang, Z.; Wang, C.; Gan, C.; Sun, S.; Wang, M. Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD. *IEEE Trans. Signal Inf. Process. Netw.* **2019**, *5*, 469–478. [CrossRef]
- Cheng, E.; Yan, J.; Sun, H.; Qi, J. Research on MPSK modulation classification of underwater acoustic communication signals. In Proceedings of the 2016 IEEE/OES China Ocean Acoustics (COA), Harbin, China, 9–11 January 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–5.
- Xu, Z.; Wu, X.; Gao, D.; Su, W. Blind Modulation Recognition of UWA Signals with Semi-Supervised Learning. In Proceedings of the 2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xi'an, China, 25–27 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
- Li, Y.; Wang, B.; Shao, G.; Shao, S.; Pei, X. Blind detection of underwater acoustic communication signals based on deep learning. *IEEE Access* 2020, *8*, 204114–204131. [CrossRef]
- 19. Li, Y.; Wang, B.; Shao, G.; Shao, S. Automatic modulation classification for short burst underwater acoustic communication signals based on hybrid neural networks. *IEEE Access* 2020, *8*, 227793–227809. [CrossRef]
- Wang, H.; Wang, B.; Li, Y. IAFNet: Few-shot learning for modulation recognition in underwater impulsive noise. *IEEE Commun. Lett.* 2022, 26, 1047–1051. [CrossRef]
- Zhang, W.; Yang, X.; Leng, C.; Wang, J.; Mao, S. Modulation recognition of underwater acoustic signals using deep hybrid neural networks. *IEEE Trans. Wirel. Commun.* 2022, 21, 5977–5988. [CrossRef]
- 22. Zhang, X.D. Modern signal processing. In Modern Signal Processing; De Gruyter: Berlin, Germany, 2022.
- 23. Yang, J.; Liu, H.; Yuan, B.X.; Sun, G.; Yuan, Y. Modulation Recognition for Communication Signals: Principles and Algorithms; Posts and Telecom Press: Beijing, China, 2022.
- 24. Zhao, X.; Guo, C.; Li, J. Mixed recognition algorithm for signal modulation schemes by high-order cumulants and cyclic spectrum. *J. Electron. Inf. Technol.* **2016**, *38*, 674–680.
- 25. Tan Xiaoheng, C.Y. Automatic digital modulation recognition based on combined feature parameters. *Syst. Eng. Electron.* **2011**, 33, 2733–2736.
- LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* 1998, 86, 2278–2324. [CrossRef]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- 28. Fang, C.; Xue, Z. Signal classification method based on complete bispectrum and convolutional neural network. *Appl. Res. Comput.* **2018**, *35*, 3766–3769. (In Chinese)
- 29. Liu, L.; Wu, K.; Chen, J.; Li, M. Modulation Recognition Method of Underwater Acoustic Communication Signal Based on Transfer Learning. *J. Ocean Technol.* **2022**, *41*, 1–11. (In Chinese)

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