frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 04 July 2024
DOI 10.3389/fmars.2024.1334134

OPEN ACCESS

EDITED BY
Xiangrong Zhang,
Xidian University, China

REVIEWED BY
Guanchun Wang,
Xidian University, China
Xuebo Zhang,
Northwest Normal University, China
Xiaoxue Qian,
University of Texas Southwestern Medical Center, United States

*CORRESPONDENCE
Yan Wang
eappl7799@gmail.com

RECEIVED 06 November 2023
ACCEPTED 16 May 2024
PUBLISHED 04 July 2024

CITATION
Lyu Y, Cheng X and Wang Y (2024) Automatic modulation identification for underwater acoustic signals based on the space–time neural network.
Front. Mar. Sci. 11:1334134.
doi: 10.3389/fmars.2024.1334134

Automatic modulation identification for underwater acoustic signals based on the space–time neural network

Yaohui Lyu[1], Xiao Cheng[2] and Yan Wang[2*]

[1]College of Electronic Engineering, Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China, [2]School of Physics and Electronic Engineering, Taishan University, Tai'an, China

In general, CNN gives the same weight to all position information, which will limit the expression ability of the model. Distinguishing modulation types that are significantly affected by the underwater environment becomes nearly impossible. The transformer attention mechanism is used for the feature aggregation, which can adaptively adjust the weight of feature aggregation according to the relationship between the underwater acoustic signal sequence and the location information. In this paper, a novel aggregation network is designed for the task of automatic modulation identification (AMI) in underwater acoustic communication. It is feasible to integrate the advantages of both CNN and transformer into a single streamlined network, which is productive and fast for signal feature extraction. The transformer overcomes the constraints of sequential signal input, establishing parallel connections between different modulations. Its attention mechanism enhances the modulation recognition by prioritizing the key information. Within the transformer network, the proposed network is strategically incorporated to form a spatial–temporal structure. This structure contributes to improved classification results, and it can obtain more deep features of underwater acoustic signals, particularly at lower signal-to-noise ratios (SNRs). The experiment results achieve an average of 89.4% at −4 dB ≤ SNR ≤ 0 dB, which exceeds other state-of-the-art neural networks.

KEYWORDS

underwater acoustic communication, modulation identification, signal recognition, deep learning, neural network

1 Introduction

With the development of wireless communication, certain emission parameters identified at the employed transmitter have been a hot topic in the field of telecommunication. Normally, the time–frequency information of signals is derived from unknown or partially known sources. Signal classification plays a crucial role in both

military and civilian wireless communication systems, serving as an integral component of intelligent radios (Demirors et al., 2015). One of the key challenges in signal classification is to automatically identify the modulation scheme of an unknown signal, which is known as automatic modulation identification (AMI). AMI is essential for intelligent radios to be able to adaptively select the best modulation scheme for the current environment, and to detect and mitigate interference from other signals.

AMI plays an important role in the military (Eldemerdash et al., 2016). Modern electronic warfare (EW) comprises three major aspects: electronic support (ES), electronic attack (EA), and electronic protect (EP) (Poisel, 2008). The ES goal is to obtain information from radio signal emissions. The successful signal detection is determined by AMI. The modulation classification results could provide EA with the valuable support, which can extend into all modules in EW. With the crowded communication resources and the emerging number of consumers, the problem of the spectrum scarcity becomes more severe in civilian wireless settings (Miao et al., 2010). Nevertheless, the actual requirements for the largest capacity and the best quality of service face a substantial difficulty of multiple interferences in the communication process. With the advent of cognitive radio (CR), the signal classification system in the civilian sector is garnering increasing attention as it leverages the flexible capabilities of transceivers to reconstruct transmission parameters. What sets the CR transceiver apart from a traditional transceiver is its ability to perceive and adapt to the transmission source's environment (Gorcin and Arslan, 2014). Therefore, CR has been interpreted as the essential part and the most attractive research field of the signal classification system in the civilian area. In the two areas mentioned above, AMI serves as the basis for the intelligent radio implementation.

One of the most challenging communication conduits is the underwater acoustic channel. In view of the low signal attenuation, sound is the most universal transmission method in underwater communications, which is regarded as a broadband system at a very low frequency, such as a few kHz (Singer et al., 2009). With this method, the center frequency is significant in the case of the bandwidth. The multipath interference has a significant impact on acoustic propagation, and it is worth noting that sound travels at a relatively slow speed, approximately 1,500 m/s. Excessive Doppler effects are induced by the movement of underwater equipment, resulting in delay extensions of tens or even hundreds of milliseconds, which lead to signal frequency-selective fading. It is a prominent restriction for underwater wireless communication, particularly when compared to the properties of light waves and electromagnetic waves (Stojanovic and Preisig, 2009).

The modulation classification algorithm is primarily composed of both likelihood-based (LB) methods (Panagiotou et al., 2000; Abdi et al., 2004; Chavali and Da Silva, 2011; Shi and Karasawa, 2011) and feature-based (FB) methods (Boudreau et al., 2000; Dobre et al., 2012; Mihandoost and Amirani, 2016). When classifiers require knowledge of the perfect channel parameters, LB achieves the highest performance in terms of classification accuracy. LB methods mainly include two steps. First, each modulation hypothesis appraises the likelihood with the received signals. The chosen channel model originates from the probability functions that can accommodate to satisfy the low-complexity requirements or be adaptable to suit the non-cooperative environment. Then, the probabilities associated with different modulation assumptions are compared with the determined classification result.

In reality, the most critical objective is to achieve multifunctionality in non-cooperative strategies and make advancements in computational complexity. It is largely constitutive of average likelihood ratio test (ALRT), generalized likelihood ratio test (GLRT), and hybrid likelihood ratio test (HLRT). The ALRT, GLRT, and HLRT classifiers hypothetically possess perfect channel information, or there may be one or two unknown channel parameters under certain circumstances. Among these classifiers, the most complex is the likelihood function of ALRT, which involves exponential operations and multi-integral calculations. The GLRT likelihood function, while simpler in expression, may result in classification deviations. The HLRT combines the advantages of both ALRT and GLRT, striking a balance between complexity and classification performance. These methods aim to reduce the complexity of the maximum likelihood classifier, which remains a key challenge. LB offers excellent classification accuracy, grounded in decision theory. The high complexity of the LB algorithm presents an opportunity for FB classifiers. While FB demonstrates suboptimal performance, it comes with lower computational demands compared to LB. FB investigates the spectral characteristics of signals and utilizes various spectral properties as factors for modulation classification. The usual structure of FB classifiers involves the wavelet-based traits captured by the wavelet functions, the high-order statistic traits examining the types and orders of signals, and the cyclic traits based on the cyclostationary analysis.

Machine learning algorithms (MLAs), as part of the FB methods, are widely used for AMI. Some of the reaching classification judgments specify an underlying type for the multi-stage decision trees, where each stage trades on the distinguished signal traits. However, there are some inconveniences for the optimization of various judgment thresholds and the design of the decision tree. To strengthen the algorithms on the basis of MLA, all types of methods have been adopted to complete two principal propositions in the modulation classification. First, MLAs make the classification decision thresholds more convenient to achieve. Second, MLA can be a tool to alter the data dimension on the signal pattern, which is accomplished by the auto-generated and auto-chosen traits. There are varying traits to be found for satisfying the computational demand of the classifier. The MLA classifier, such as a support vector machine, is generally in association with signal traits to advance to a higher dimension. Moreover, MLA implements the reduced-order dimension in the signal trait space, which selects k-nearest neighbor, genetic algorithm, and linear regression.

Deep learning methods have achieved great success in computer vision, natural language processing, and speech recognition. However, in underwater environments, deep learning methods face a number of challenges, including large attenuation, noise interference, and data scarcity. To address these challenges, researchers have proposed a variety of methods. In Liu et al. (2017), the redesigned ResNet in the lightweight state has better classification results, which embraces the

shallower layer without the larger receptive field in the network structure. In Yang (2017), the network possesses an expansive structure with multiple layers, enabling it to effectively capture a broader range of signal characteristics, thereby enhancing the effects of AMC. In Lee et al. (2017), the fading communication environment is analyzed, and the AMI obstacle is properly dealt with the whole conjunction neural network. In Zhang et al. (2018), a long short-term memory (LSTM) network is combined with a CNN to create a new network with two streams. This design accommodates a wide range of distinctive signal features, contributing to improved identification performance. In Yu et al. (2019), by contrasting the conventional network structure, a remarkable improvement in AMI is produced by the structural adjustment of CNN. In Yang et al. (2016), with reference to the stochastic interference of the underwater signals, the eligible results can be carried out by the exiguous deep encoder running in the automatic mode. In Li et al. (2017), the innovative DLA is compared with the traditional statistical technique, which has the evident asset in the AMI task. In Li-Da et al. (2018), a similar network structure with the alliance of LSTM and CNN is forwarded in the underwater AMI, and gives the method versatility to a certain degree in the underwater and terrestrial communication system. In Li et al. (2020), a fusion neural network, comprising an attention-enhanced CNN and a sparse autoencoder, is introduced for the AMI of short burst underwater acoustic signals. This approach demonstrates robustness against channel conditions and noise In Wang et al. (2022), IAFNet integrating impulsive noise preprocessing, attention network, and few-shot learning is proposed for underwater acoustic modulation recognition with few labeled samples under impulsive noise, improving classification accuracy by 7% compared to other methods by effectively extracting features through denoising and task-driven attention. In Zhang et al. (2022), a recurrent and convolutional neural network is proposed for underwater acoustic modulation recognition, combining RNN and CNN for automatic feature extraction. It achieves a higher accuracy of 98.21% on Trestle and 99.38% on South China Sea datasets, which has a faster recognition time of 7.164 ms compared to conventional deep learning methods. In Yao et al. (2023), deep complex networks with proposed deep complex matched filter and deep complex channel equalizer layers are explored for underwater acoustic modulation classification, which shows improved performance over real-valued DNNs (deep neural networks) by reducing multipath fading and noise influences.

In the field of underwater acoustic communications, various deep learning networks commonly exhibit direct stacking or simple combination without considering the architecture design of task-specific optimizations. Compared to terrestrial wired communications, underwater acoustic communications face more severe multipath effects, Doppler shifts, and ocean ambient noise interference. This requires a perspective tailored for underwater acoustics to comprehensively consider the characteristics of underwater sound propagation and design more optimized, dedicated deep neural networks. More precisely, the chosen techniques should be integrated to model the multidimensional characteristics of underwater acoustic signals. Diversity-aware signal selection modules, enhanced attention mechanisms, and hierarchical feature extraction structures serve as the primary implementation methods. Meanwhile, adaptively designed regularization terms and well-configured loss functions are also necessary to adapt to the highly dynamic and stochastic underwater environment during network training. These methods can genuinely unlock robust feature extraction and modeling capabilities, enhancing the interpretability of AMI and communications. The contributions of this paper are mainly as follows:

(1) The network unit is structured with distinct branches. While the primary branch remains consistent, the auxiliary branch can assume one of three optional orientations. This enhances the classification capability of the network by facilitating the exchange of learned advanced signal features between different branches. It broadens the range of extracted signal characteristics while maintaining a lower number of parameters.

(2) The hybrid routing network structure invests a simplistic format for the complex routing logic network. After several network units are overlaid structurally, the used network can generate the multiple routing modes, which enhances the performance of the extracted signal traits and has the faster training speed.

(3) The transformer network is introduced to handle long temporal signal series, and the high-dimensional features of temporal domain signals are dynamically acquired in the multi-head attention mechanism, which enhances the recognition ability at lower SNRs.

2 Signal model

The AMI task effectively constitutes a multi-classification problem, exhibiting a strong resemblance to other conventional tasks within the field of deep learning. The received signals take on a complex representation in the temporal domain, encompassing various modulation styles. The channel can be expressed as Figure 1, and the underwater received acoustic signals can be represented as in Equation (1):

$$g(t) = h(t, \delta) \otimes s(t) + n(t) = \sum_{i=1}^{I} e_i(t) k(t - \delta_i(t)) + n(t) \quad (1)$$

where $s(t)$ is the sending signal; $h(t, \delta)$ is the channel impulse response with multipath, Doppler effect, and time delay; $n(t)$ is AWGN; $e_i(t)$ is the attenuation at the ith path; and $\otimes$ denotes the signal convolution. $\delta_i(t)$ is the ith path time delay, I is the total number of multipath signals, and a similar Doppler scaling factor β is set in all paths, $\delta_i(t) \approx \delta_i - \beta t$ (Li et al., 2008). The sending signals can be analog (e.g., single-sideband modulation) or digital (e.g., phase-shift keying).

3 The proposed method

3.1 Signal preprocessing

The input signal count is determined by a constant value in the standard DLA, but this approach may not yield optimal results in AMI.

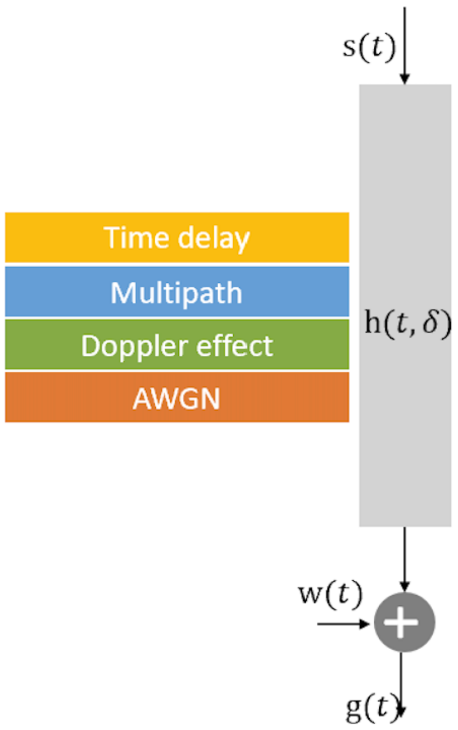

FIGURE 1
The underwater acoustic channel model.

To capture the underlying signal characteristics and improve AMI accuracy, the signal constellation pixels can be grouped in various ways. The proposed network takes variable-sized pixel groups as input. This approach enables the extraction of diverse signal modulation traits, thereby significantly enhancing classification accuracy.

The various numbers of the signal pixel groups are shown as in Equation (2):

$$\mathbb{R}_{\phi'}(k) = c_h(\Phi - \sum_{m=1}^{M-1} \phi(m)) \tag{2}$$

where M is the grouped total number, m is the group number. $\phi(\cdot)$ is the signal pixel sequence corresponding to the group number, and Φ is the total number of signals. The signal pixel $c(\cdot)$ is the current retrieval group, and h is the signal pixel number obtained. $\mathrm{R}\phi'(\cdot)$ is served as the input sequence of the modulation signals in the used network, and ϕ' is the current pixel sequence. The proposed network produces the classification results using the various input pixel groups.

3.2 Proposed network structure

The aggregated multipath network achieves a similar effect to the low-density scattered network in extracting diverse signal features, but it does not increase the number of model parameters. Moreover, it avoids the intricate structure of the low-density scattered network, which employs numerous small convolutions and pooling operations within its layer structure. The high complexity, apart from affecting model efficiency and training speed, can also diminish the network's ability to extract features from weak signals in varying underwater conditions, potentially resulting in reduced recognition performance.

To address this limitation, a new network, called the aggregation multipath network, is proposed in Figure 2. This approach significantly improves the network's capacity to capture in-depth contextual information from signals and provides an effective solution for creating a compact neural network model capable of handling diverse input signal data. The aggregation multipath network solves these problems by keeping the unchanged structure in routes. The key to realizing the ideal model capacity and efficiency is to maintain a large number of routes with the same width. In this way, there are neither dense convolutions nor too many *Add* operations.

The aggregation multipath network is composed of the following main components. The preprocessed underwater acoustic signal (preprocessed signals) is passed to the input module, which contains the 2D convolution and MaxPooling layer, with the 3 × 3 convolution kernel. To complete these operations, these attributes are connected into the next main network structure unit (cyan dashed border), which is constructed by superimposing a multipath unit (*MulPU*), an attribute rearrangement unit (*AttRU*), and a panoramic feature reification unit (*PanFRU*), and transmitted to a deep and wide network for further learning. The multipath unit and PanFRU can be iterated multiple times, effectively extracting distinctive features from weak underwater acoustic signals. When the above stage is completed, the global average pooling (GAP) is intended to reduce the attribute map size to 1×1, and finally the fully connected layer (dense) outputs the modulation prediction value.

This architecture is carefully structured to ensure that the features utilized after the AttRU are not only valid but also relevant and representative of the underlying signal characteristics. By iteratively refining and focusing on key attributes, the network minimizes the risk of incorporating irrelevant or misleading features. Moreover, the GAP and dense layers at the end of the network serve to further validate and consolidate these features before the final modulation classification.

Regarding potential performance degradation across different modulation classes, the network is designed to be robust and adaptive to various signal types. The iterative feature extraction process, coupled with the network's ability to handle multiple signal paths and attributes, ensures that the system remains effective across a range of modulation classes. This design approach helps mitigate the risk of significant performance drops for certain modulation types, thereby maintaining consistent and reliable identification accuracy across different scenarios.

At the beginning of each unit, the proposed network is divided into different routes. The corresponding formula is as follows in Equation (3):

$$\mathbb{M} = \sum_{p=1}^{P} [m_0 + \sum_{v=1}^{V} \psi(m_v) + m_P] \tag{3}$$

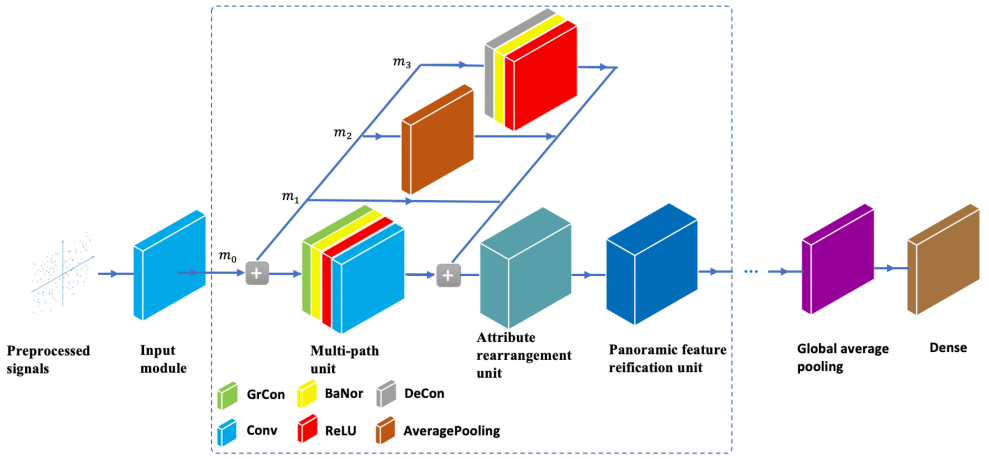

FIGURE 2
The aggregation multipath network structure.

where p represents the superimposed units, p = 1,2,⋯,P, and P represents the maximum number of superimposed units. m_0 represents the main route, $\psi(m_{(\cdot)})$ represents a selection function to the supplementary routes, v represents the alternative mode of different supplementary routes, $v = 1,\cdots,V$, and V represents the total number of supplementary routes. m_P represents the panoramic feature reification unit in Figure 3. The ith layer can choose any optional supplementary route required from 1 to V, and $\mathbb{M}$ represents the final network structure.

The main route m_0 consists of four basic modules: group convolution (*GrCon*) with a 1 × 1 convolution kernel; the batch normalization (*BaNor*) module, the ReLU activation function (*ReLU*) module, and 2D convolution (*Con*) with a 2 × 2 convolution kernel. The three routes that can be selected on the supplementary path correspond to m_1, m_2, and m_3, respectively. The supplementary route m_1 is a directly connected link. The supplementary route m_2 includes average pooling (AveragePooling). The supplementary route m_3 consists of three basic modules: Depthwise convolution (*DeCon*) with a 2 × 2 convolution kernel, *BaNor*, and *ReLU*. *DeCon* is a special convolution that operates on each input attribute map independently, which reduces the computational amount by removing unnecessary data.

AttRU is the attribute exchange operation between the different routes in Figure 4. The boxes serve as representations of signal attributes. The process of box reconstruction is a specialized technique employed to overcome the inherent limitations of diverse pathways and to harness the full potential of the abundant signal attributes. This method entails the rotation and exchange of a specific proportion of attributes among different pathways, culminating in the provision of results to the subsequent unit for enhanced learning. The general convolution operates comprehensively across all input attribute maps, a technique referred to as full-attribute convolution, emphasizing an attribute-dense connection where convolution is applied to all attributes. Notably, information within distinct routes may exhibit similarities within the same box. Without attribute exchange, the learned attributes are inherently limited. Conversely, when attributes are exchanged between different routes, information learned can also be exchanged. This exchange augments the information within each box, enabling the extraction of more features. This approach ultimately fosters the acquisition of attributes from all other boxes within each route, contributing to more favorable outcomes.

The 2D attribute matrix corresponding to each route vector is $\mathbf{x_1}$, $\mathbf{x_2}$,⋯,$\mathbf{x_\theta}$, and the selected attribute range percentage is λ. The attributes involved in the exchange in Equation (4) are:

$$t_{\lambda\theta} = \mathcal{J}_\lambda(\mathrm{x}_\theta) \tag{4}$$

where $\mathcal{J}_\lambda(\cdot)$ represents selecting λ percentage of the route features for rotation exchange. The symbol $t_{\lambda\theta}$ represents the vector representation after alteration based on the λ ratio, equivalent to the

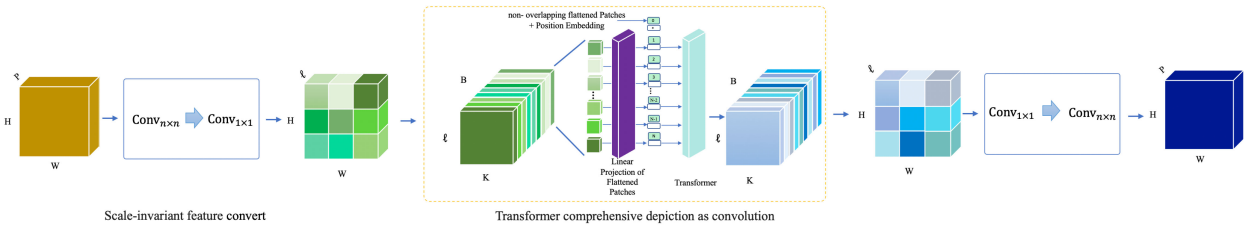

FIGURE 3
Panoramic feature reification unit.

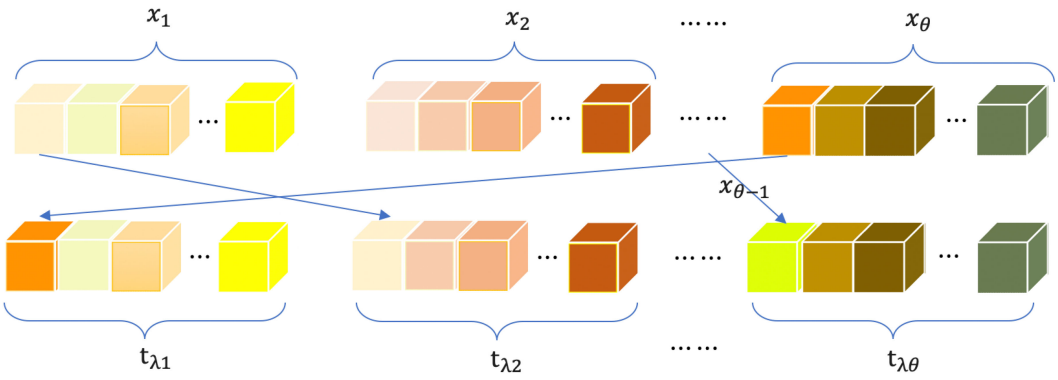

FIGURE 4
Attribute rearrangement operation (λ = 10%).

output of $\mathcal{T}_\lambda(\cdot)$. After the first network unit learns, the proportionally selected initial matrix in Equation (5) is:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_\theta \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1\theta} \\ x_{21} & x_{22} & \cdots & x_{2\theta} \\ \vdots & & \ddots & \vdots \\ x_{\theta 1} & x_{\theta 2} & \cdots & x_{\theta\theta} \end{bmatrix} \tag{5}$$

the rotating and changing operation starts at 1 and ends at θ. At λ = 10%, the corresponding matrix transformation becomes in Equation (6):

$$\begin{bmatrix} t_{\lambda 1} \\ t_{\lambda 2} \\ \vdots \\ t_{\lambda \theta} \end{bmatrix} = \begin{bmatrix} x_{\theta 1} & x_{12} & \cdots & x_{1\theta} \\ x_{11} & x_{22} & \cdots & x_{2\theta} \\ \vdots & & \ddots & \vdots \\ x_{(\theta-1)1} & x_{\theta 2} & \cdots & x_{\theta\theta} \end{bmatrix} \tag{6}$$

Attribute exchange is a key technique used in the aggregation multipath network to overcome the limitations of the sparse fragmented network and achieve better classification results. It involves rotating and changing a certain proportion of attributes between different routes, which allows the network to learn more diverse features.

PanFRU consists of a scale-invariant feature convert, transformer comprehensive depiction as convolution (*TransCDC*), and the reverse process, shown in Figure 3. The upper layer input $A \in A^{H \times W \times P}$, and H, W, and P represent the height, width, and passages of the input tensor, respectively. PanFRU utilizes a convolutional layer with an $n \times n$ kernel size, followed by another convolutional layer with a 1×1 kernel to generate a feature map A_D of size $H \times W \times \ell$. The $n \times n$ convolution captures local spatial information in the input. The 1×1 convolution then projects each spatial location to an ℓ-dimensional space, where ℓ is twice than the number of input passages P. This allows the 1×1 convolution to learn new representations by taking linear combinations of the input high-level abstract underwater signal data.

TransCDC unfolds the feature map A_E of size $K \times B \times \ell$ into linear projection of flattened patches. At position 0, non-overlapping flattened patches and position embeddings are added, just like a standard vision transformer. Relative positional information pertains to the specific distribution of signals and serves as a discriminative factor in identifying modulation categories within the underwater acoustic signal modulation constellation. The original underwater acoustic data do not contain the relative position information of modulations, and it leads to the same effect in the different position vector. Each distinct position vector corresponds to the positional information embedded within the input underwater acoustic signal sequence, and these vectors are subsequently fed into the transformer network. $K = w * h$ is the flattened patch size, and $w = W/n$ and $h = H/n$ are patch height and width. $B = (H * W)/K$ is the number of patches.

For each patch j, transformers are applied to $A_E(j)$ to encode inter-patch relationships, $\mathbf{1} \leqslant \mathbf{j} \leqslant \mathbf{J}$, producing a global feature map A_F of size $K \times \ell \times B$ in Equation (7):

$$A_F(j) = \mathfrak{O}(A_E(j)) \tag{7}$$

TransCDC retains both patch order and signal constellation pixel order within each patch. In contrast to the standard vision transformer, *TransCDC* does not suffer from the loss of spatial ordering. A_F can be folded back to spatial dimensions $H \times W \times \ell$, which is projected to P dimensions via 1×1 convolution and concatenated with A. These concatenated features are fused by another $n \times n$ convolution. $\mathfrak{O}(\cdot)$ is the standard vision transformer operation. Since $A_E(j)$ encodes local $n \times n$ spatial information and $A_F(j)$ encodes global relationships across all K patches for each location, *TransCDC* allows each signal constellation pixel to incorporate global context from the entire input. It is difficult to distinguish the modulation types in the spatial dimension. The discrimination ability of transformer can be effectively improved by the position information. The attention mechanism of the transformer can remember the key distinguishing information like the human visual attention mechanism. The model is improved to alleviate the signal fading, which enhances the modulation recognition ability.

After the learned features are processed by the AttRU and PanFRU modules, they undergo a crucial transformation via GAP. GAP serves to distill the features into a more manageable and representative form. This is mathematically represented as follows in Equation (8):

$$Y = GAP(X) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{ij} \tag{8}$$

Here, **X** represents the feature matrix post-AttRU and PanFRU processing, and **Y** denotes the final output feature vector. H and W are the height and width of the feature map, respectively. The GAP operation averages the features spatially, reducing each channel to a single scalar value. This simplified representation is crucial for the final classification or recognition task, enabling the network to output concise and effective recognition results.

4 Experiment

In the underwater acoustic wireless channel (Wang et al., 2022), the generated signals are more approximate to the realistic situation of disturbances. The dataset involves 10 types of modulation signals, namely, BPSK, QPSK, 8PSK, 4PAM, 16QAM, 64QAM, FM, DSB, CPFSK, and 4FSK. The signal is transmitted at a carrier frequency of 10 kHz with a symbol transmission rate of 1,000 symbols/s. The channel is modeled as a Rayleigh fading channel with 20 cosines to represent frequency selective fading. The receiver has a standard offset drift process in the sample rate and a maximum deviation of 15 Hz in the random mode. The raised cosine pulse-shaping filter used at the transmitter has a 0.25 roll-off factor. The additive noise is added in the communication process, which is Gaussian white, zero mean, and bandlimited noise. The random seed number generated is 0x1999 in the noise source. The deviation of the maximum sample rate is set to 25 Hz, and the drift process standard offset is 0.1 Hz per sample in the sample rate. The cosine number is set to 10 in the frequency selective fading simulation. A total of 2,000,000 modulation data are included in the dataset, and SNRs are in the range of −20 dB to +18 dB. The dataset is divided into a training set, a validation set, and a testing set (60%, 20%, and 20%, respectively). There is a complex floating point I/Q value in each signal data, and the pixel groups are 32, 64, 128, and 256 in Figures 5, 6.

The training setting of the used network is that the batch size is chosen as 128, and the optimizer selects the stochastic gradient descent (SGD) with momentum = 0.85, decay = $5\times e^{-4}$, and learning rate = 0.01. To elevate the extension ability of the trained network, the early stopping technique is applied with five patience epochs.

Figure 5 visually represents the convergence performance of the proposed network throughout both the training and validation phases. In Figure 5A, the training signal dataset is transmitted through the network in use, resulting in training loss values computed using the categorical cross-entropy loss function. In Figure 5B, the validation signal dataset is employed to assess the trained network, yielding validation loss values

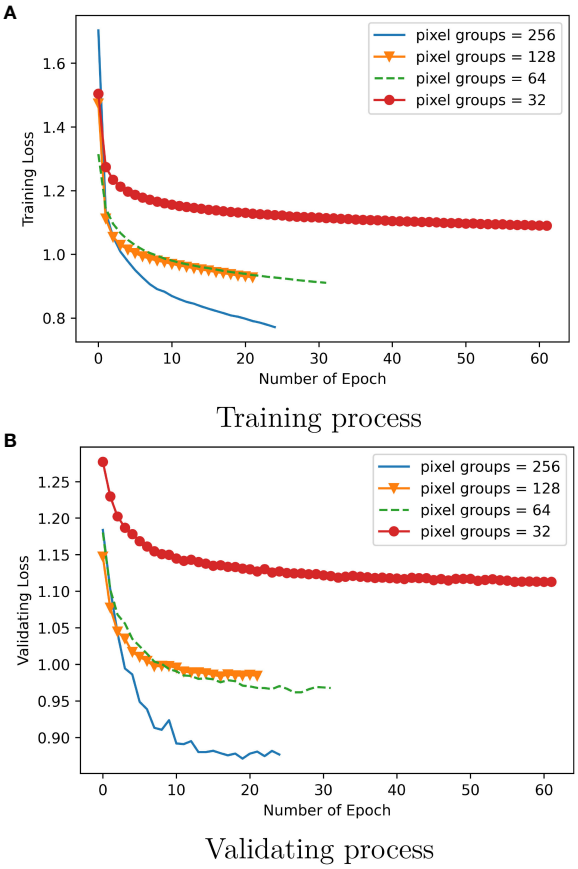

FIGURE 5
(A, B) Training and validating process in the hybrid route network.

computed using the same loss function as in the training phase. In the horizontal axis, "Epoch" signifies the number of complete cycles in which the entire training or validation dataset is processed by the network and returns once. At the beginning, the four kinds of pixel groups show a rapid convergence in the iterative procedure of the top five. At pixel groups = 32, the training loss has a longer epoch number, which is higher than the other three by an average of more than 0.25. There is a similar convergence performance for the pixel groups of 64 and 128. At pixel groups = 256, the training loss gets the best results. As the pixel groups increase, the epoch number declines. The training process shows that the used network can work effectively to learn the signal data's traits. Comparing the training and validating process, there is an approximate convergence tendency. The validating process has a smooth course like the training process, which can productively fulfill the validation of the trained network. In the validating process, the epoch number also reduces as the pixel groups increase, which is similar to the training process. It shows that the various pixel groups can be effectively handled by the used method.

Figure 6 shows the modulation classification performance in various pixel groups with different routing forms. The

superposition of three units in the network structure is shown. (*X*, *X*) represents the auxiliary selection corresponding to the intermediate overlay units. Other forms of (*X*, *X*, *X*) have similar meanings, and *X* is the different choice of m_1, m_2, or m_3. λ is the percentage of the selected packet range as a track for the exchange between auxiliary branches. In the 12 different hybrid routing forms, 10 modulation types with pixel groups of 32, 64, 128, and 256 can be effectively identified. When pixel groups = 32, there are similar modulation classification results in the different routing forms at the low SNRs between −20 dB and −16 dB. When SNRs > −16 dB, (l_1, l_2, l_3), λ = 100% is better at approximately 1.2%, 3.2% than (l_1, l_2, l_3), λ = 60%, (l_1, l_2, l_3), λ = 20% from −16 dB to 0 dB, which is more effective by 4.8%, 8.2%, and 13.6% than λ = 100%, λ = 60%, and λ = 20% of other routing forms on average. There is a similar trend at pixel groups = 64. Further adding routing branches did not improve the classification accuracy. It is due to the fact that the hybrid routing network can better extract the numerous signal traits by the full exchange of tracks, and the advantageous classification effect can be achieved under the (l_1, l_2, l_3), λ = 100% form. As the pixel groups increase to 128, almost the same classification results are achieved with different hybrid routing networks during SNRs< −16 dB. With the increase of SNRs, the classification ability differs from the selection mode of auxiliary branches. When −16 dB< SNRs< 5 dB, the pixel groups of 128 in the form of (l_1, l_2, l_3), λ = 100% is increased by approximately 6.9% and 10.5% in the form of (l_1, l_2, l_3), λ = 60% and (l_1, l_2, l_3), λ = 20%, which has a distinct advantage over other routing forms. When the pixel groups are 256, the presence of different routing forms results in sheep herd performance when the SNR is less than a specified dB level. When λ is 100%, (l_1, l_2, l_3) has the best effects at SNRs > −15 dB, and slightly improves to 1.3%, 0.7%, and 1.1% compared to the other three routing forms of (l_1, l_2), (l_1, l_3), and (l_2, l_3), which have a mean increase of 2.5% compared to the other two trait exchange percentages in the different routing forms. When the sufficient pixel groups of signal data are provided to the used network, more signal high-dimensional traits are extracted, which helps to promote the ability to identify the modulation types. It explains that the hybrid routing network is an efficient method for AMI with various pixel groups.

Figure 7 shows the classification of various modulation styles at different SNRs. At SNR ≤ −14 dB, the classification rate of SSB is higher. This occurs because other modulation styles are often misidentified as SSB modulation. There is a marked classification improvement of 64QAM from −8 dB to −6 dB. The main reason is that the constellations of underwater acoustic signals between 16QAM and 64QAM have distinguishable high-dimensional features, which can be discovered by the proposed network. As SNR increases, the proposed method can clearly differentiate between the two modulations, and other modulations have been correctly classified for each of the categories. The classification rate of 10 modulation styles has exceeded 85% at SNR = −2 dB. The proposed method can overcome the influence of the underwater acoustic signal interferences, which achieves high recognition accuracies at lower SNRs.

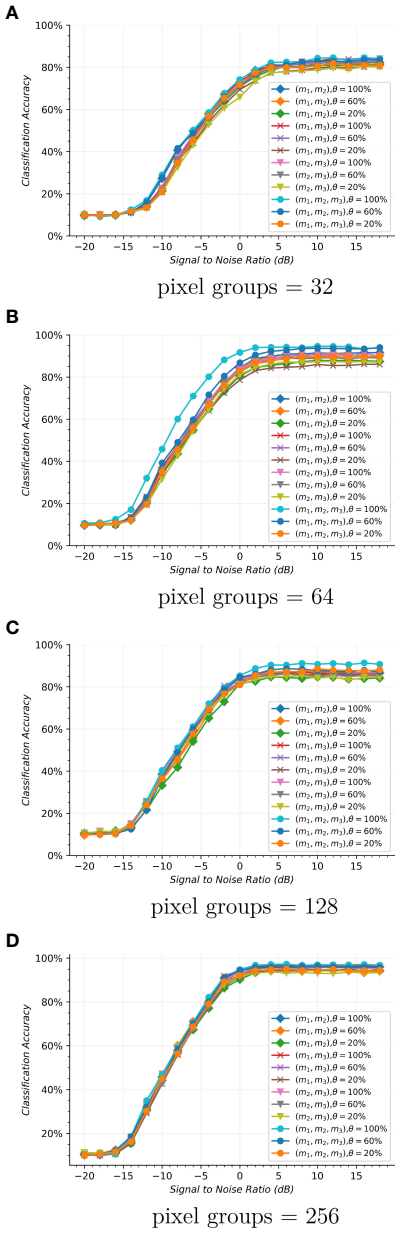

FIGURE 6
(A–D) Modulation classification accuracy with the various pixel groups in the different route forms.

In Figure 8, the efficacy of recognizing various modulation styles is displayed at typical SNRs of 256 pixel groups. Notably, 4FSK and 8FSK exhibit substantial recognition accuracy at an SNR of −4 dB. Similarly, SSB and FM are identified as prevalent analog modulation styles. In underwater acoustic settings, characterized by poor communication quality due to low SNR, analog signal waveforms suffer from significant distortion. The distortion often leads to misinterpretations resembling those seen in underperforming systems. Despite these challenges, the proposed method successfully differentiates between analog modulations that are

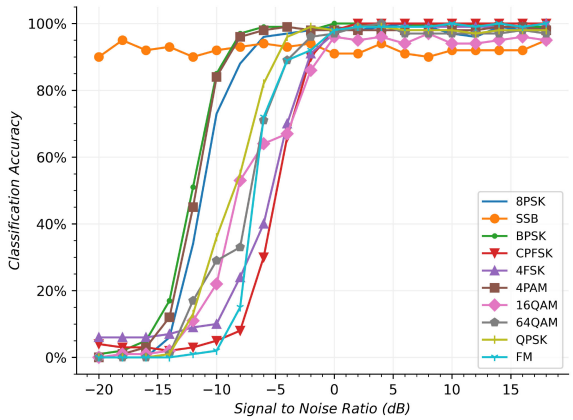

FIGURE 7
Each modulation classification results in different SNRs.

typically prone to confusion. Additionally, the proposed method demonstrates precise recognition of 16QAM and 32QAM. It is common for the signal constellations to exhibit a degree of similarity, which could lead to inferior performance. At the specified SNR levels, the proposed method distinguishes effectively between 16QAM and 32QAM. This proficiency extends to other modulation styles as well, yielding enhanced recognition capabilities. The proposed method progressively assimilates more signal traits, culminating in optimal recognition outcomes.

The proposed method is compared with ablation methods and other latest methods, including Proposed Method (without PanFRU), which is the proposed method without PanFRU; Proposed Method (without MAU), which is the proposed method without MAU (MulPU and AttRU); CLDNN (West and O'shea, 2017), which integrates convolutional and LSTM layers for spatial and temporal feature capture; LSTM (Chen et al., 2020), a recurrent network adept at learning long-term dependencies in time-series data; Transformer network (Dosovitskiy et al., 2020), known for its attention mechanisms and global dependency handling; ResNet (O'shea et al., 2018), employing skip connections for training deeper networks; Squeeze-and-Excitation Network (Wei et al., 2020), emphasizing informative features through channel-wise relationships; HybridCRNN (Zhang et al., 2022), a fusion of convolutional and recurrent layers for local and temporal feature extraction; and RanForest (random forest) (Fang et al., 2022), an ensemble method using multiple decision trees for improved classification, especially in non-linear contexts. These methods represent a broad spectrum of modern approaches in AMI, each with distinct advantages in processing and analyzing complex signals. As shown in Figure 9, the proposed method outperforms ablation methods and other network methods at all SNRs, except at an SNR of −20dB. The classification rate of all network methods is very low, and various modulation styles cannot be recognized. When SNRs are greater than −18 dB, the proposed method demonstrates a recognition advantage over the ablation methods, namely, the Proposed Method without PanFRU and the Proposed Method without MAU. This clearly indicates that the Proposed Method employing PanFRU and MAU has a superior capability in extracting hidden classification information from underwater acoustic signals. Compared with other network methods, the proposed method has always maintained the advantage of the classification rate at SNRs ≥ −15 dB. It is due to the fact that the proposed method mines more deep representations of underwater acoustic signals, and its attention mechanism and multipath structure plays a crucial role for a good recognition effect.

The Proposed Method's epoch time cost is compared with various other networks in Table 1. These results were obtained by a CPU i7, GPU 3090, Ubuntu 22.04, and TensorFlow version 2.10. The term "epoch time" denotes the duration required for each epoch of training. Among the methods evaluated, the Proposed Method emerges as the most time-efficient. A comparative analysis of the Proposed Method with its ablation variants, including both the Proposed Method (without PanFRU) and Proposed Method (without MAU), further elucidates its efficiency. In comparison, the Transformer, LSTM, ResNet, CLDNN, SENet, and HybridCRNN

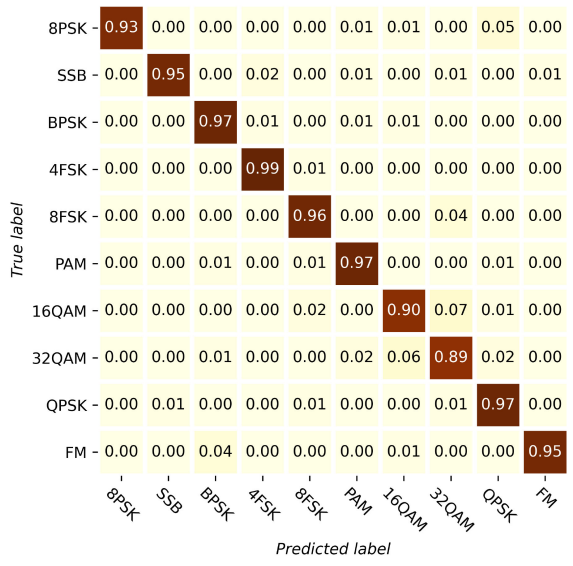

FIGURE 8
Modulation classification results at SNR = 0 dB.

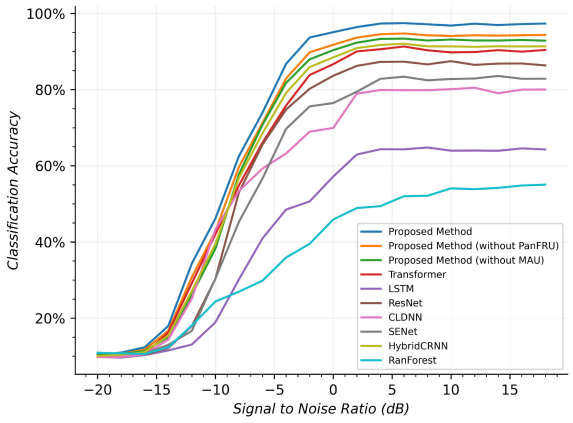

FIGURE 9
Modulation classification results in different networks.

exhibit epoch times, which are approximately 1.6, 2.0, 4.8, 1.5, 3.2, and 3.1 times longer than that of the Proposed Method, respectively. It is important to note that the epoch time for the RanForest method is not applicable in this context. A critical aspect contributing to the efficiency of the Proposed Method is its CNN structure coupled with an optimized transformer architecture. This configuration facilitates parallel processing, which is notably more efficient than the sequential processing required by the Transformer and LSTM architectures, as these necessitate the preservation of intermediate states. The streamlined and potent network design of the Proposed Method surpasses the performance of more intricate architectures such as ResNet, CLDNN, SENet, and HybridCRNN.

5 Conclusion

The paper analyzed the modulation identification of the hybrid network in the underwater acoustic environment. The complexity of underwater communication makes it difficult to achieve a high identification accuracy. The proposed network can obtain the effective signal traits of modulation signals with various signal pixel groups. The proposed network, featuring multiple routing forms and optional auxiliary exchange branches, enhances the extraction of numerous signal characteristics, thereby significantly improving identification performance. Remarkably, this network not only maintains a compact parameter size but also shortens the training duration. Enhancing the efficiency of underwater non-cooperative communication under constrained conditions holds considerable practical significance. The proposed network method can also be extended to the other signal classification scenes for underwater communication. In the future, we will study the hybrid routing method for the low SNR modulation identification in an underwater acoustic environment.

Building upon the insights from current research on modulation identification in hybrid networks within underwater acoustic environments, it is recognized that addressing the challenges posed by non-Gaussian noise types, such as alpha noise and Middleton Class A and Class B noise, especially prevalent in shallow water conditions, is a significant consideration for future studies. Upcoming research endeavors plan to specifically target the impact of non-Gaussian noise on modulation identification in underwater acoustic communication. Recognizing that such noise types can substantially affect the performance of communication systems, particularly in shallow water environments, there is an aim to develop and integrate methodologies in the network that can effectively handle and adapt to these complex noise scenarios. This will include expanding the scope of the algorithmic framework to better accommodate the distinct characteristics of non-Gaussian noise, ensuring that the approach remains robust and effective under a wider range of environmental conditions. Thus, the exploration and incorporation of strategies to address non-Gaussian noise types will be a key focus in future work, enhancing the applicability and reliability of the network in diverse underwater communication settings.

TABLE 1 The time cost of different networks.

Network method	Epoch time (s)
Proposed method	38
Proposed method (without PanFRU)	36
Proposed method (without MAU)	35
Transformer	62
LSTM	75
ResNet	184
CLDNN	58
SENet	122
HybridCRNN	117
RanForest	–

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

YL: Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing. XC: Data curation, Validation, Visualization, Writing – review & editing. YW: Funding acquisition, Resources, Software, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by Scientific Research Startup Foundation of Taishan University (No. Y-01-2020016), Shandong Provincial Natural Science Foundation (No. ZR2022MF347).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abdi, A., Dobre, O. A., Choudhry, R., Bar-Ness, Y., and Su, W. (2004). "Modulation classification in fading channels using antenna arrays," in *IEEE MILCOM 2004. Military Communications Conference*, Monterey, CA, USA 1, 211–217. doi: 10.1109/MILCOM.2004.1493271

Boudreau, D., Dubuc, C., Patenaude, F., Dufour, M., Lodge, J., and Inkol, R. (2000). "A fast automatic modulation recognition algorithm and its implementation in a spectrum monitoring application," in *MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No.00CH37155)*, Los Angeles, CA, USA. 2, 732–736. doi: 10.1109/MILCOM.2000.904026

Chavali, V. G., and Da Silva, C. R. (2011). Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-gaussian channels. *IEEE Trans. Commun.* 59, 2051–2056. doi: 10.1109/TCOMM.2011.051711.100184

Chen, Y., Shao, W., Liu, J., Yu, L., and Qian, Z. (2020). Automatic modulation classification scheme based on lstm with random erasing and attention mechanism. *IEEE Access* 8, 154290–154300. doi: 10.1109/Access.6287639

Demirors, E., Sklivanitis, G., Melodia, T., Batalama, S. N., and Pados, D. A. (2015). Software-defined underwater acoustic networks: Toward a high-rate real-time reconfigurable modem. *IEEE Commun. Magazine* 53, 64–71. doi: 10.1109/MCOM.2015.7321973

Dobre, O. A., Oner, M., Rajan, S., and Inkol, R. (2012). Cyclostationarity-based robust algorithms for QAM signal identification. *IEEE Commun. Lett.* 16, 12–15. doi: 10.1109/LCOMM.2011.112311.112006

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Eldemerdash, Y. A., Dobre, O. A., and Öner, M. (2016). Signal identification for multiple-antenna wireless systems: Achievements and challenges. *IEEE Commun. Surveys Tutorials* 18, 1524–1551. doi: 10.1109/COMST.2016.2519148

Fang, T., Wang, Q., Zhang, L., and Liu, S. (2022). Modulation mode recognition method of non-cooperative underwater acoustic communication signal based on spectral peak feature extraction and random forest. *Remote Sens.* 14, 1603. doi: 10.3390/rs14071603

Gorcin, A., and Arslan, H. (2014). Signal identification for adaptive spectrum hyperspace access in wireless communications systems. *IEEE Commun. Magazine* 52, 134–145. doi: 10.1109/MCOM.2014.6917415

Lee, J., Kim, J., Kim, B., Yoon, D., and Choi, J. (2017). Robust automatic modulation classification technique for fading channels via deep neural network. *Entropy* 19, 454. doi: 10.3390/e19090454

Li, Y., Wang, B., Shao, G., and Shao, S. (2020). Automatic modulation classification for short burst underwater acoustic communication signals based on hybrid neural networks. *IEEE Access* 8, 227793–227809. doi: 10.1109/Access.6287639

Li, C., Zhou, Q., Han, X., Yin, J., and Shao, M. (2017). Underwater non-cooperative communication signal recognition with deep learning. *J. Acoustical Soc. America* 142, 2732–2732. doi: 10.1121/1.5014979

Li, B., Zhou, S., Stojanovic, M., Freitag, L., and Willett, P. (2008). Multicarrier communication over underwater acoustic channels with nonuniform doppler shifts. *IEEE J. Oceanic Eng.* 33, 198–209. doi: 10.1109/JOE.2008.920471

Li-Da, D., Shi-Lian, W., and Wei, Z. (2018). "Modulation classification of underwater acoustic communication signals based on deep learning," in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, Kobe, Japan, 1–4. doi: 10.1109/OCEANSKOBE.2018.8559101

Liu, X., Yang, D., and Gamal, A. E. (2017). Deep neural network architectures for modulation classification. *arXiv preprint arXiv:1207.0580*.

Miao, G., Himayat, N., and Li, G. Y. (2010). Energy-efficient link adaptation in frequency-selective channels. *IEEE Trans. Commun.* 58, 545–554. doi: 10.1109/TCOMM.26

Mihandoost, S., and Amirani, M. C. (2016). "Automatic modulation classification using combination of wavelet transform and GARCH model," in *2016 8th International Symposium on Telecommunications (IST)*, Tehran, Iran, 484–488. doi: 10.1109/ISTEL.2016.7881868

O'shea, T. J., Roy, T., and Clancy, T. C. (2018). Over-the-air deep learning based radio signal classification. *IEEE J. Selected Topics Signal Process.* 12, 168–179. doi: 10.1109/JSTSP.2018.2797022

Panagiotou, P., Anastasopoulos, A., and Polydoros, A. (2000). "Likelihood ratio tests for modulation classification," *MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No.00CH37155)*, Los Angeles, CA, USA, 2, 670–674. doi: 10.1109/MILCOM.2000.904013

Poisel, R. A. (2008). *Introduction to communication electronic warfare systems* (Norwood, MA, USA: Artech House, Inc).

Shi, Q., and Karasawa, Y. (2011). Noncoherent maximum likelihood classification of quadrature amplitude modulation constellations: Simplification, analysis, and extension. *IEEE Trans. Wireless Commun.* 10, 1312–1322. doi: 10.1109/TWC.2011.030311.101490

Singer, A. C., Nelson, J. K., and Kozat, S. S. (2009). Signal processing for underwater acoustic communications. *IEEE Commun. Magazine* 47, 90–96. doi: 10.1109/MCOM.2009.4752683

Stojanovic, M., and Preisig, J. (2009). Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Commun. Magazine* 47, 84–89. doi: 10.1109/MCOM.2009.4752682

Wang, H., Wang, B., and Li, Y. (2022). Iafnet: Few-shot learning for modulation recognition in underwater impulsive noise. *IEEE Commun. Lett.* 26, 1047–1051. doi: 10.1109/LCOMM.2022.3151790

Wei, S., Qu, Q., Wu, Y., Wang, M., and Shi, J. (2020). Pri modulation recognition based on squeeze-and-excitation networks. *IEEE Commun. Lett.* 24, 1047–1051. doi: 10.1109/COML.4234

West, N. E., and O'shea, T. (2017). "Deep architectures for modulation recognition," *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)* Baltimore, MD, USA, 2017, 1–6. doi: 10.1109/DySPAN.2017.7920754

Yang, G. Q. (2017). Modulation classification based on extensible neural networks. *Math. Problems Eng.* 2017, 1–10. doi: 10.1155/2017/6416019

Yang, H., Shen, S., Xiong, J., and Zhang, X. (2016). "Modulation recognition of underwater acoustic communication signals based on denoting & deep sparse autoencoder," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Institute of Noise Control Engineering)*, Vol. 253. 5506–5511.

Yao, X., Yang, H., and Sheng, M. (2023). Automatic modulation classification for underwater acoustic communication signals based on deep complex networks. *Entropy* 25, 318. doi: 10.3390/e25020318

Yu, X., Li, D., Wang, Z., Guo, Q., and Wei, X. (2019). A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals. *Wireless Networks* 25, 3735–3746. doi: 10.1007/s11276-018-1667-6

Zhang, D., Ding, W., Zhang, B., Xie, C., Li, H., Liu, C., et al. (2018). Automatic modulation classification based on deep learning for unmanned aerial vehicles. *Sensors* 18, 924. doi: 10.3390/s18030924

Zhang, W., Yang, X., Leng, C., Wang, J., and Mao, S. (2022). Modulation recognition of underwater acoustic signals using deep hybrid neural networks. in *IEEE Transactions on Wireless Communications* 21(8), 5977–5988. doi: 10.1109/TWC.2022.3144608