

Article

A Deep Learning Framework for Signal Detection and Modulation Classification

Xiong Zha *, Hua Peng, Xin Qin, Guang Li and Sihan Yang

PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, Henan, China; Peng_hua@outlook.com (H.P.); qinxin_0920@163.com (X.Q.); GL_for_study@outlook.com (G.L.); courage32@163.com (S.Y.)

* Correspondence: mici0928@163.com

Received: 20 July 2019; Accepted: 16 September 2019; Published: 19 September 2019



Abstract: Deep learning (DL) is a powerful technique which has achieved great success in many applications. However, its usage in communication systems has not been well explored. This paper investigates algorithms for multi-signals detection and modulation classification, which are significant in many communication systems. In this work, a DL framework for multi-signals detection and modulation recognition is proposed. Compared to some existing methods, the signal modulation format, center frequency, and start-stop time can be obtained from the proposed scheme. Furthermore, two types of networks are built: (1) Single shot multibox detector (SSD) networks for signal detection and (2) multi-inputs convolutional neural networks (CNNs) for modulation recognition. Additionally, the importance of signal representation to different tasks is investigated. Experimental results demonstrate that the DL framework is capable of detecting and recognizing signals. And compared to the traditional methods and other deep network techniques, the current built DL framework can achieve better performance.

Keywords: deep learning; signal detection; modulation classification; the single shot multibox detector networks; the multi-inputs convolutional neural networks

1. Introduction

Cognitive radio (CR) [1–3] has been used to refer to radio devices that are capable of learning and adapting to their environment. Due to the increasing requirements for wireless bandwidth of radio spectrum, automatic signal detection and modulation recognition techniques are indispensable. It can help users to identify the modulation format and estimate signal parameters within operating bands, which will benefit communication reconfiguration and electromagnetic environment analysis. Besides, it is widely used in both military and civilian applications, which have attracted much attention in the past decades [4–7].

Multi-signals detection is a task to detect the existing signals in a specific wideband, which is one of the essential components of CR. The most significant difference between signal and non-signal is energy. Hence, many wideband multi-signals detection algorithms are based on energy detector (ED). Some threshold-based wideband signal detection methods, such as [8–13], reduce the probability of false alarm or missed alarm. However, these methods are sensitive to noise changes and challenging to ensure the detection accuracy of all detection scenarios. Therefore, many non-threshold-based detection algorithms have been proposed [14–17]. However, these algorithms have high computational complexity, which results in poor online detection performance.

For automatic modulation recognition, algorithms based on signal phase, frequency, and amplitude have been widely used [18]. However, these algorithms are significantly affected by noise, and the performance can be substantially degraded in low SNR condition. High-order statistical-based

algorithms [19–22], such as signal high-order cumulants and cyclic spectrum, have excellent anti-noise performance. The computational complexity of these methods is relatively low, but the selection of features relies too much on expert experience. It is difficult to obtain features that can adapt to non-ideal conditions. In particular, it is challenging to set the decision threshold when there are plenty of modulation formats to be classified.

Deep Learning (DL) techniques [23,24] have made outstanding achievements in Computer Vision [25,26] (CV) and Natural Language Processing [27,28] (NLP) for their strong self-learning ability. Recently, more and more researchers use DL techniques to solve signal processing problems. For signal detection, many DL-based methods, such as [29–31], detect signals in narrowband environment. These methods only detect the existing of signal, but can not estimate the relevant parameters. Therefore, developing a technique leverages deep learning to detect signal efficiently and effectively is still a challenging problem. For DL-based modulation classification, there has been some reported work, including [32–36]. For example, some researchers used the signal IQ waveform as data representation and learned the sample using CNNs [32–34]. Other researchers focused on developing methods to represent modulated signals in data formats for CNNs. Among these methods, constellation-based algorithms [35,36] have been widely utilized, where signal prior knowledge is fully considered.

In this study, DL techniques are fully utilized in multi-signals detection and modulation recognition. For multi-signals detection, we use the deep learning target detection network to detect the location of each signal. In our initial research, the used model is SSD networks, which is a relatively advanced target detection network. Furthermore, we use the time-frequency spectrum as the signal characteristic expression. Due to the time-frequency characteristic of the M-ary Frequency Shift Keying (MFSK) format signals, we can identify the modulation format while the signal is detected. Meanwhile, for M-ary Phase Shift Keying (MPSK), M-ary Amplitude Phase Shift Keying (MAPSK), and M-ary Quadrature Amplitude Modulation (MQAM) signal, the difference in the time-frequency spectrum is not sufficient to identify the signal modulation. Therefore, during the signal detection procession, we identify them in the same format, and only detect the signal presence or absence. Through the signal detection network, we can roughly get the signal carrier frequency and start-stop time. After that, we use a series of traditional methods to convert these signals from the wideband into the baseband. To recognize MPSK, MAPSK, and MQAM signals, a multi-inputs CNNs is designed. Moreover, we adopt the signal vector diagram and eye diagram as the network inputs, which are more robust than in-phase and quadrature (IQ) waveform data and constellation diagram.

This paper addresses the topic of DL based multi-signals detection and modulation classification. The main contributions of this paper are summarized as follows: (1) We propose a relatively complete DL framework for signal detection and modulation recognition, which is more intelligent than traditional algorithms. (2) We establish different signal representation schemes for several tasks, which facilitate the use of the built DL framework for detection and classification. (3) We propose a multi-inputs CNNs model to extract and map the features from different dimensions.

The rest of this paper is presented as follow. In Section 2, we offer a detail introduction to the signal model and the dataset generation. Section 3 shows the DL framework for signal detection and modulation recognition. Section 4 confirms our initial experiment result from different aspects. Finally, our conclusions and directions for further research are given in Section 5.

2. Communication Signal Description and Dataset Generation

In realistic communication processing, the signal may be distorted by the effect of non-linear amplifier and channel. In actual situation, the received signal in the communication system can be expressed as:

$$r(t) = \int_{\tau=0}^{\tau_0} s(n_{clk}(t-\tau))h(\tau)d\tau + n_{add}(t) \quad (1)$$

where $s(t)$ is the transmission signal, $n_{clk}(t)$ is timing deviation, $h(t)$ represents the transmitted wireless channel, $n_{add}(t)$ is additive white Gaussian noise.

In this section, we will describe different modulated signals and their sample representation for our DL framework. We will also explain the reason why we use it and the method we enhance it.

2.1. Modulation Signal Description

For any digital modulation signal, the transmission signal can be presented as

$$s(t) = \sum_n a_n e^{j(w_n t + \phi)} g(t - nT_b) \quad (2)$$

where w_n is the signal angular frequency, ϕ is the carrier initial phase, T_b is the symbol period, a_n is the symbol sequence, $g(t)$ is the shaping filter.

For MFSK signal, it can be presented as

$$a_n = 1, w_n = w_0 + \frac{2\pi}{M}i, i = 0, 1, \dots, M-1 \quad (3)$$

For MPSK signal, it can be presented as

$$a_n = e^{j2\pi i/M}, i = 0, 1, \dots, M-1, w_n = w_0 \quad (4)$$

For MQAM signal, it can be presented as

$$\left. \begin{aligned} a_n &= I_n + jQ_n \\ I_n, Q_n &= 2i - \frac{M}{4} + 1, i = 0, 1, \dots, \frac{M}{4} - 1, w_n = w_0 \end{aligned} \right\} \quad (5)$$

MAPSK constellations are robust against nonlinear channels due to their lower peak-to-average power ratio (PAPR), compared with QAM constellations. Therefore, APSK was mainly employed and optimized over nonlinear satellite channels during the last two decades. As recommended in DVB-S2 [37], it can be presented as:

$$a_n = r_k \exp \left[j \left(\frac{2\pi}{n_k} i_k + \theta_k \right) \right] \quad (6)$$

where r_k is the radius of the k th circle, n_k is the number of constellations in k th circle, i_k is the ordinal number of constellation points in the k th circle, θ_k is the initial phase of the k th circle.

2.2. Signal Time-Frequency Description

For multi-signals detection task, we use the wideband signal time-frequency spectrum as the neural network input. To prove the feasibility of this method, we theoretically prove the time-frequency visual characteristic of each modulation. Here, we use the short-time Fourier transform [38] (STFT) to analyze the signal time-frequency characteristic.

2.2.1. MFSK Signal Time-frequency Description

The STFT of MFSK signal can be expressed as

$$STFT_{s_{\text{FSK}}}(t, w) = \int_{-\infty}^{+\infty} [s_{\text{FSK}}(\tau) \gamma^*(\tau - t)] e^{-jw\tau} d\tau = \int_{-\infty}^{+\infty} \left[\sum_{k=-\infty}^{+\infty} A g(\tau - kT_b) e^{j(\omega_k \tau + \phi_k)} \gamma^*(\tau - t) \right] e^{-jw\tau} d\tau \quad (7)$$

where $\gamma(t)$ is the window function, whose duration is T . When $\gamma(t)$ is in a symbol duration, Equation (7) can be simplified as

$$STFT_{s_{\text{FSK}}}(t, w) = \int_{-T/2}^{T/2} A e^{j(\omega_k(\tau+t) + \phi_k)} e^{-jw(\tau+t)} d\tau = A T e^{-jw t} e^{j(\omega_k t + \phi_k)} \text{Sa} \left(\frac{w - \omega_k}{2} T \right), \quad (8)$$

$$kT_b + T/2 < t < (k+1)T_b - T/2, k = 0, 1, 2, \dots$$

where $Sa(w) = \sin(w)/w$. When $\gamma(t)$ spans two symbols, Equation (7) can be simplified as

$$\begin{aligned} STFT_{s_{FSK}}(t, w) &= \int_{-T/2}^d Ae^{j(w_k(\tau+t)+\phi_k)} e^{-jw(\tau+t)} d\tau + \int_d^{T/2} Ae^{j(w_{k+1}(\tau+t)+\phi_{k+1})} e^{-jw(\tau+t)} d\tau = \\ &= Ae^{j((w_k-w)t+\phi_k)} \int_{-T/2}^d e^{-j(w-w_k)\tau} d\tau + Ae^{j((w_{k+1}-w)t+\phi_{k+1})} \int_d^{T/2} e^{-j(w-w_{k+1})\tau} d\tau = \\ &= A \frac{T+2d}{2} e^{j((w_k-w)t+\phi_k)} e^{\frac{j(w-w_k)(T-2d)}{4}} Sa\left(\frac{(w-w_k)(T+2d)}{4}\right) + \\ &+ A \frac{T-2d}{2} e^{j((w_{k+1}-w)t+\phi_{k+1})} e^{\frac{j(w-w_{k+1})(T+2d)}{4}} Sa\left(\frac{(w-w_{k+1})(T-2d)}{4}\right), \end{aligned} \quad (9)$$

$(k+1)T_b \quad -T/2 < t < (k+1)T_b + T/2, d = (k+1)T_b - t, k = 0, 1, 2, \dots$

where w_{k+1} is the carrier angular frequency of the $k+1$ -th symbol. If $w_{k+1} = w_k$, it indicates that the carrier angular frequency does not jump, so Equation (8) is same as Equation (9). We take the modulus square of Equation (8). The result can be expressed as

$$\begin{aligned} SPEC_{s_{FSK}}(t, w) &= |STFT_{s_{FSK}}(t, w)|^2 = A^2 T^2 Sa^2\left(\frac{(w-w_k)T}{2}\right), \\ &kT_b + T/2 < t < (k+1)T_b - T/2, k = 0, 1, 2, \dots \end{aligned} \quad (10)$$

And for Equation (9), it can be expressed as:

$$\begin{aligned} SPEC_{s_{FSK}}(t, w_k) &\approx \frac{A^2(T+2d)^2}{4} \leq A^2 T^2, -T/2 < d < T/2, \\ &(k+1)T_b - T/2 < t < (k+1)T_b + T/2, k = 0, 1, 2, \dots \end{aligned} \quad (11)$$

Obviously, the value of $SPEC_{s_{FSK}}(t, w_k)$ will increase as the increase of jumping time d . The energy decreases gradually as $\gamma(t)$ slips away from the symbol. So when $d = T/2$, the window is completely within one symbol, and the maximum value is obtained.

$$\begin{aligned} SPEC_{s_{FSK}}(t, w_k)_{\max} &= A^2 T^2, \\ &(k+1)T_b - T/2 < t < (k+1)T_b + T/2, k = 0, 1, 2, \dots \end{aligned} \quad (12)$$

When $d = -T/2$, the window completely spans to next symbol, and the minimum value is obtained

$$\begin{aligned} SPEC_{s_{FSK}}(t, w_k)_{\min} &= 0, \\ &(k+1)T_b - T/2 < t < (k+1)T_b + T/2, k = 0, 1, 2, \dots \end{aligned} \quad (13)$$

From our analysis, we can easily get the characteristics of FSK modulation: (1) There will be sharp brightness changes in the time-frequency image at the frequency change moment. (2) The signal modulation number M and frequency spacing are important parameters for the MFSK time-frequency characteristics, which determine the value of w_k .

2.2.2. Amplitude-Phase Modulation Signal Time-frequency Description

For MPSK, MAPSK, and MQAM signal, since they all belong to amplitude-phase modulation, the derivation processing of the signal time-frequency characteristics is the same as MPSK. Hence, we specify the time-frequency characteristics of the MPSK signal, and the STFT can be expressed as:

$$\begin{aligned} STFT_{s_{PSK}}(t, w) &= \int_{-\infty}^{+\infty} [s_{PSK}(\tau)\gamma^*(\tau-t)] e^{-jw\tau} d\tau = \\ &= \int_{-\infty}^{+\infty} \left[\sum_{k=-\infty}^{+\infty} Ag(\tau - kT_b) e^{j(w_c\tau + \phi_c + \phi_k)} \gamma^*(\tau-t) \right] e^{-jw\tau} d\tau \end{aligned} \quad (14)$$

As the derivation of MFSK signal time-frequency characteristics, when $\gamma(t)$ is in a symbol duration, the Equation (14) can be simplified as:

$$\begin{aligned} STFT_{s_{PSK}}(t, w) &= \int_{-T/2}^{T/2} Ae^{j(w_c(\tau+t)+\phi_c+\phi_k)} e^{-jw(\tau+t)} d\tau = ATe^{-jw_t} e^{j(w_c t + \phi_c + \phi_k)} Sa\left(\frac{(w-w_c)T}{2}\right), \\ &kT_b + T/2 < t < (k+1)T_b - T/2, k = 0, 1, 2, \dots \end{aligned} \quad (15)$$

When $\gamma(t)$ spans two symbols, the Equation (14) can be simplified as:

$$\begin{aligned}
 STFT_{s_{PSK}}(t, w) &= \int_{-T/2}^d Ae^{j(w_c(\tau+t)+\phi_c+\phi_k)} e^{-jw(\tau+t)} d\tau + \int_d^{T/2} Ae^{j(w_c(\tau+t)+\phi_c+\phi_{k+1})} e^{-jw(\tau+t)} d\tau = \\
 &= Ae^{j((w_c-w)t+\phi_c+\phi_k)} \int_{-T/2}^d e^{-j(w-w_c)\tau} d\tau + Ae^{j((w_c-w)t+\phi_c+\phi_{k+1})} \int_d^{T/2} e^{-j(w-w_c)\tau} d\tau = \\
 &= A \frac{T+2d}{2} e^{j((w_c-w)t+\phi_c+\phi_k)} e^{\frac{j(w-w_c)(T-2d)}{4}} Sa\left(\frac{(w-w_c)(T+2d)}{4}\right) + \\
 &\quad + A \frac{T-2d}{2} e^{j((w_c-w)t+\phi_c+\phi_{k+1})} e^{\frac{j(w-w_c)(T+2d)}{4}} Sa\left(\frac{(w-w_c)(T-2d)}{4}\right), \\
 &\quad (k+1)T_b - T/2 < t < (k+1)T_b + T/2, d = (k+1)T_b - t, k = 0, 1, 2, \dots
 \end{aligned}
 \tag{16}$$

where ϕ_{k+1} is the phase of the $k+1$ -th symbol. And if $\phi_{k+1} = \phi_k$, Equation (15) is equal to Equation (16). We take the modulus square of (15), and the result can be expressed as:

$$\begin{aligned}
 SPEC_{s_{PSK}}(t, w) &= |STFT_{s_{PSK}}(t, w)|^2 = A^2 T^2 Sa^2\left(\frac{w-w_c}{2} T\right), \\
 &\quad kT_b + T/2 < t < (k+1)T_b - T/2, k = 0, 1, 2, \dots
 \end{aligned}
 \tag{17}$$

And for (16), it can be expressed as:

$$\begin{aligned}
 SPEC_{s_{PSK}}(t, w_c) &= \frac{A^2 T^2}{2} (1 + \cos(\phi_k - \phi_{k+1})) + 2A^2 d^2 (1 - \cos(\phi_k - \phi_{k+1})) \leq A^2 T^2, \\
 &\quad (k+1)T_b - T/2 < t < (k+1)T_b + T/2, d = (k+1)T_b - t, k = 0, 1, 2, \dots
 \end{aligned}
 \tag{18}$$

We take the partial derivative for Equation (18):

$$\begin{aligned}
 \frac{\partial SPEC_{s_{PSK}}(t, w_c)}{\partial d} &= 4A^2 d (1 - \cos(\phi_k - \phi_{k+1})), -T/2 < d < T/2, \\
 &\quad (k+1)T_b - T/2 < t < (k+1)T_b + T/2, k = 0, 1, 2, \dots
 \end{aligned}
 \tag{19}$$

From Equation (19), we can easily learn that $SPEC_{s_{PSK}}(t, w_c)$ get the minimum value when $\phi_{k+1} = \phi_k$ or $d = 0$. But the minimum value is much greater than 0, which is greatly different for the MFSK signal.

$$\begin{aligned}
 SPEC_{s_{PSK}}(t, w_c)_{\min} &= \frac{A^2 T^2}{2} (1 + \cos(\phi_k - \phi_{k+1})) \gg 0, \\
 &\quad (k+1)T_b - T/2 < t < (k+1)T_b + T/2, k = 0, 1, 2, \dots
 \end{aligned}
 \tag{20}$$

Hence, for MPSK signal, there is only one wide frequency band in the time-frequency diagram, and the brightness fluctuation appears in a small range, which is different from MFSK. And from derivation processing, we can know that the MPSK time-frequency characteristics are less affected by M, so it is hard to distinguish PSK signals with different M. Figure 1 presents different modulation signals in the wideband.

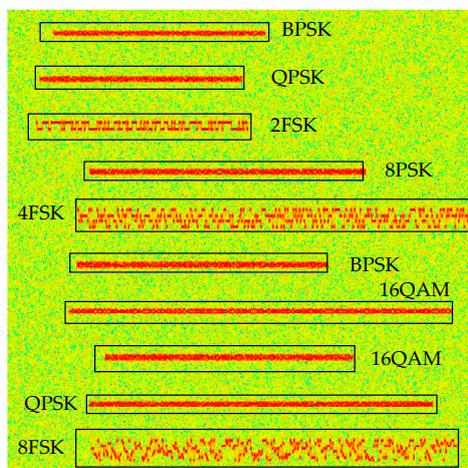


Figure 1. Different modulation signals in the wideband.

2.3. Signal Eye Diagram and Vector Diagram Description

The function of the eye diagram is to observe the baseband signal waveform by an oscilloscope. Through the eye-diagram, we can adjust the receiver filter to improve system performance. Besides, due to the characteristics of the modulated signal itself, different modulation modes have apparent visual differences in the eye diagram. As shown in Figure 2, because of the different modulation scales, there are different eye numbers in each eye diagram. For OQPSK, since the two orthogonal signals stagger for half a symbol period, the eye-opening position is always staggered, while other modulated signals always appear at the same time.

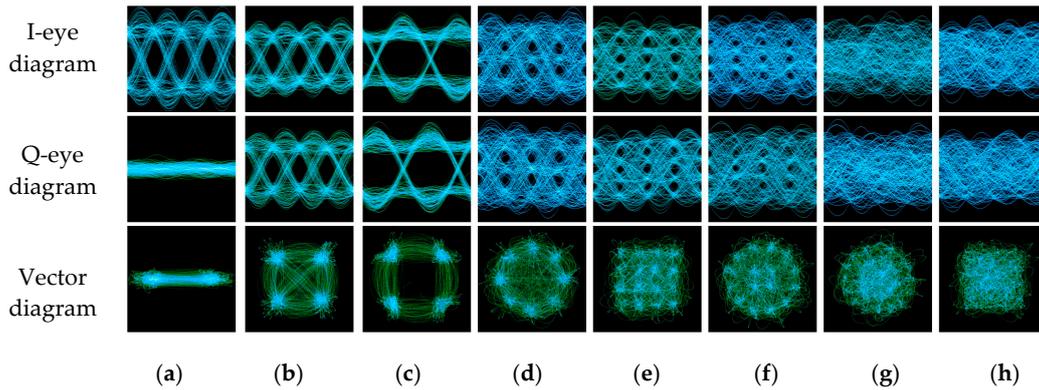


Figure 2. The eye diagram and vector diagram of different modulation signals in 15dB (a) BPSK; (b) QPSK; (c) OQPSK; (d) 8PSK; (e) 16QAM; (f) 16APSK; (g) 32APSK; (h) 64QAM.

By reconstructing the signal IQ waveforms in the corresponding time, the signal vector diagram shows the symbol trajectory. From its formation mechanism, it is similar to the signal constellation diagram. However, unlike the constellation diagram, the vector diagram can reflect the signal phase information. For example, it can easily distinguish QPSK from OQPSK with the same initial phase, because there is no 180° phase shift in OQPSK, while it exists in QPSK. Meanwhile, compared with the constellation diagram, the vector diagram is more convenient to obtain and requires less prior information.

2.4. The Generation Processing of the Dataset

Figure 3 presents the processing of our dataset construction and annotation. To make samples more diverse, we set sampling phase offset, frequency offset, phase offset, and amplitude attenuation in sample generation processing.

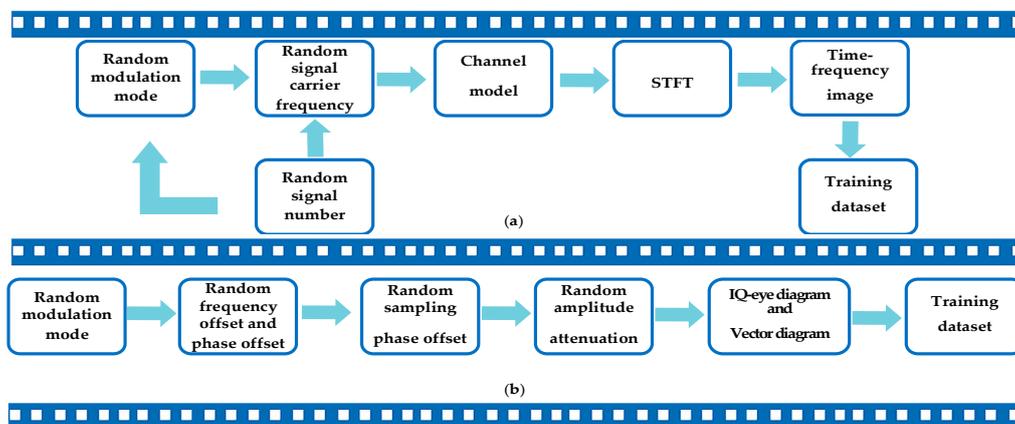


Figure 3. The generation processing of the dataset. (a) dataset for signal detection; (b) dataset for modulation recognition.

For signal detection, we need to determine the reconnaissance frequency range and set the signal number in the wideband at first. We set different frequency offset for each signal, and ensure that the signals do not overlap in the frequency domain. Then we perform the STFT on the wideband. Not only we record the modulation format of each signal, but also record the start-stop time, carrier frequency, and bandwidth. Then we convert them into the coordinates on time-frequency image, which are the label information for the network.

For modulation recognition, traditional eye diagram and vector diagram are binary images, which do not consider the signal aggregation degree at a particular location. Hence, we consider the signal aggregation degree and enhance the traditional eye diagram and vector diagram. Figure 4 presents the enhancement processing of the dataset. In our initial research, since CNNs are insensitive to edge information, the signal amplitude is quantified between $[-1.05, 105]$ by 128 after normalizing the amplitude. Furthermore, the parameter settings are obtained by experiments. For example, we choose 800 symbols and 4 symbols as a waveform group to generate the eye diagram and the vector diagram, and related experiments will also be described in detail in subsequent chapters. Moreover, we perform the following operations on the images to make the image details more prominent, where \mathbf{Im}_0 is the original image, \mathbf{Im}_1 , \mathbf{Im}_2 , \mathbf{Im}_3 are the channels of the enhanced image and α, β are scaling factors.

$$\begin{aligned}\mathbf{Im}_1 &= \text{unit8}\left(\frac{\mathbf{Im}_0 - \min(\mathbf{Im}_0)}{\max(\mathbf{Im}_0) - \min(\mathbf{Im}_0)} \times 255\right), \\ \mathbf{Im}_2 &= \text{unit8}(\alpha \times \log_2(\mathbf{Im}_1 + 1)), \\ \mathbf{Im}_3 &= \text{unit8}(\exp(\mathbf{Im}_1 / \beta))\end{aligned}\quad (21)$$

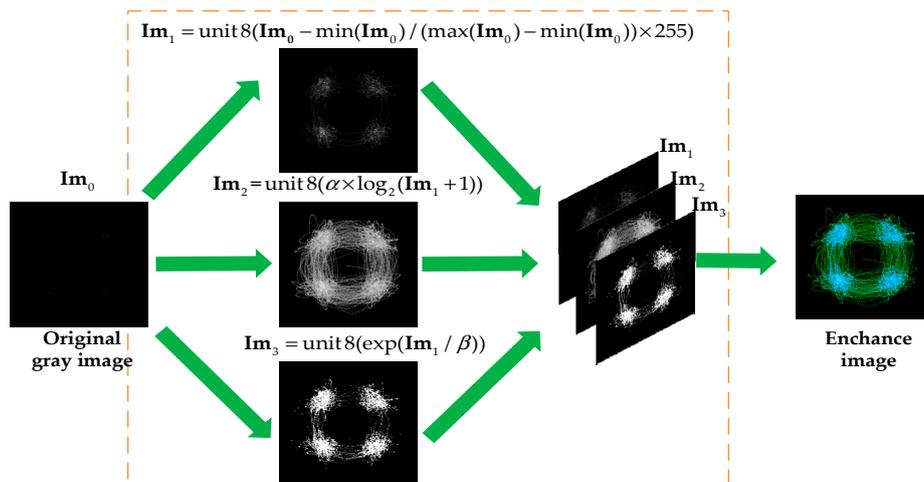


Figure 4. The enhancement processing of the dataset.

3. Deep Learning Framework for Signal Detection and Modulation Recognition

DL networks aim at learning different hierarchies of features from data. As one of the branches of DL techniques, the CNNs perform well in the field of image recognition. It performs feature learning via non-linear transformations implemented as a series of nested layers. Each layer consists of several kernels that perform a convolution over the input. Generally, the kernels are usually multidimensional arrays which can be updated by some algorithms [39]. Our DL framework achieves multi-signals detection and modulation recognition. We use different deep neural networks for different tasks. For signal detection, we use SSD networks. For modulation recognition, we design a multi-inputs CNNs.

3.1. SSD Networks for Signal Detection

We use SSD networks to achieve multi-signals detection. For DL target detection techniques, the existing algorithms are mainly divided into two kinds: algorithms based on region recommendation

(two-stage methods) and algorithms based on regression (one-stage methods). Regression-based algorithms include YOLO series algorithms [40–43] and SSD series algorithms [43,44], while region recommendation-based algorithms include RCNN [45], Fast RCNN [46], and Faster RCNN [47]. In our research, since the regression-based algorithms are faster than region recommendation-based algorithms, we use SSD networks as our signal detection model. SSD networks can generate a series of fixed-size borders and the possibility of the containing target in each border. Finally, the final detection and recognition results are calculated by the non-maximum suppression algorithm [48]. The structure of SSD networks is shown in Figure 5, and it can be divided into four parts.

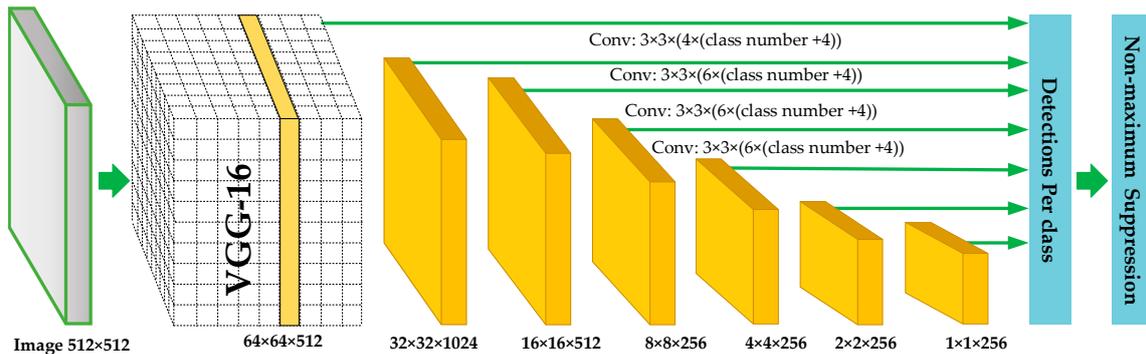


Figure 5. The network model for signal modulation recognition.

Part 1: The networks for feature extraction. The initial part of SSD networks is the first layers of VGG16 network, which is used as the primary network to extract the deep features of the whole input image. Behind the primary network, the model structure is the pyramid networks, which contains a series of simple convolution layers to make feature maps smaller and smaller. With the pyramid network structure, we can get several feature maps with different scales.

Part 2: The design of the default box. In this part, we will design several feature default box for different scales of feature maps. Each feature map at the top of the VGG16 networks is associated with a set of feature default box. As shown in Figure 6, there are dotted borders at each position of 4×4 and 8×8 feature maps. These fixed-size borders are default boxes, and their scale parameters are designed by the different feature maps scales. For example, assuming that we need M feature maps to predict, the scale parameters of the default box are as follows:

$$S_k = S_{\min} + \left(\frac{S_{\max} - S_{\min}}{m - 1} \right) \bullet (k - 1), k \in [1, m] \tag{22}$$

where S_{\min} is the bottom scale, and S_{\max} is top scale. The length-width ratio of feature default can be expressed as: $a_r \in \{1, 2, 3, 1/2, 1/3\}$. So the feature default box length is $W_k^a = S_k \sqrt{a_r}$ and the width is $H_k^a = S_k / \sqrt{a_r}$.

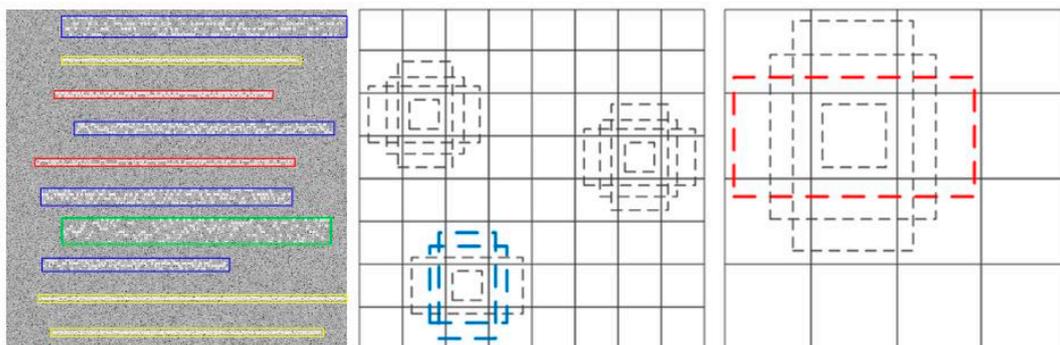


Figure 6. The Design of the default box.

Part 3: Detection and Recognition. In this part, we can predict the target category and location. We add a set of convolution kernels behind several different scales feature maps. Using these convolution kernels, we can get a fixed set of test results. For an $m \times n \times p$ feature maps, a small convolution kernel with $3 \times 3 \times p$ size is used as the fundamental prediction element. Finally, the classification probability of each feature default box and the offsets are obtained.

Part 4: Non-maximum suppression. In the last part, we use non-maximum suppression to select the best prediction results. For the feature default boxes that are matched by each real target border, we calculate their intersection-parallel ratios. The expression is shown as follow

$$IoU = (A \cap B) / (A \cup B) \quad (23)$$

where A and B are two borders. We will select the feature default box whose *IoU* are greater than 0.5 as best results, and then obtains the highest confidence degree feature default box by non-maximum suppression.

In offline train stage, the whole objective optimal function of the SSD networks includes two parts: confidence loss and location loss. The expression is shown as follow

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (24)$$

where x is used to indicate whether the feature default box is a target or not. N is the number of the feature default boxes that are matched to real target borders. Parameter α is used to adjust the ratio between L_{conf} and L_{loc} , default $\alpha = 1$. L_{conf} is softmax loss function. L_{loc} is used to measure the performance of the boundary box prediction, and in our initial research, we use the typical *smooth_{L1}* function to calculate

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0), \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p c_i^p} \quad (25)$$

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (26)$$

where *Pos* and *Neg* represent all positive and negative borders, respectively. \hat{c}_i^p represents the confidence degree for p th feature default matching i th target. l_i^m represents the prediction bias of the i -th feature default box. (cx, cy) is the box center coordinates and (w, h) is the box width and height. \hat{g}_j^m represents the deviation between the real target border g_j^m and the default box d_i^m . \hat{g}_j^m is calculated as follow: $\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^w$, $\hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^h$, $\hat{g}_j^w = \log(g_j^w / d_i^w)$, $\hat{g}_j^h = \log(g_j^h / d_i^h)$.

3.2. Multi-Inputs CNNs for Modulation Recognition

For signal modulation recognition task, the modulation set is {BPSK, QPSK, OQPSK, 8PSK, 16QAM, 16APSK, 32APSK, 64QAM}, because they are all belonging to amplitude-phase modulation and we cannot distinguish each other from time-frequency characteristic in SSD network. Hence, we use the eye diagram and vector diagram as the model inputs. The multi-inputs CNNs model is shown in Figure 7. The initial size of the samples is 128×128 , and we use softmax as the output layer's active function and relu as all other network layers' active function.

The signal features extraction can be divided into three stages. On the first stage, we use 7×7 convolution kernels to convolute IQ eye diagram and vector diagram, respectively. To ensure the dynamic range unification of the feature maps, we apply the batch normalization (BN) [49] on first layer network outputs. We perform the max pooling operation on the BN feature maps. Then we connect the feature maps from IQ eye diagram inputs.

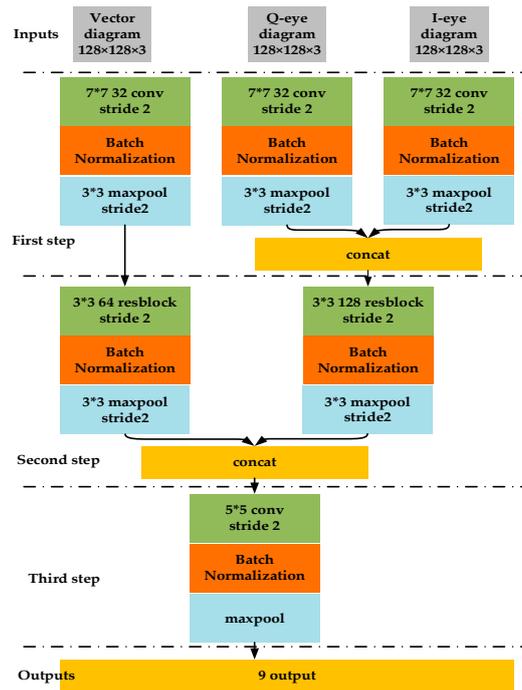


Figure 7. The network model for signal modulation recognition.

On the second signal feature extraction stage, we adopt the residual network structure to avoid the degradation caused by the network over-depth. The basic structure of ResNet [50] is shown in Figure 8. After the second feature extraction stage, each input feature maps are connected. After the batch normalization in the third stage, we directly process the feature maps by global maximum sampling to reduce the network parameters.

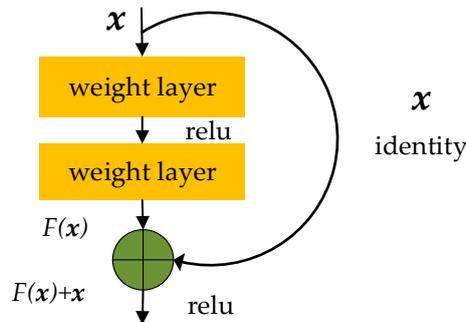


Figure 8. The basic structure of ResNet.

In the processing of network optimization, we adopt Adam algorithm [51] to solve the network parameters optimal solution. The categorical cross-entropy error is chosen as the loss function, which is represented as:

$$J_1(w, b; x_1, x_2, x_3, y) = - \sum_i^N (y_i)^T \log(f_1(x_{1,i}, x_{2,i}, x_{3,i}; w, b)) + \lambda_1 \sum_i \|w\|^2 \tag{27}$$

3.3. The Description for Deep Learning Framework

According to above introduction of the signal detection network and the modulation recognition network, we describe the use of our DL framework. Figure 9 presents the system model. The steps of the model used are as follows:

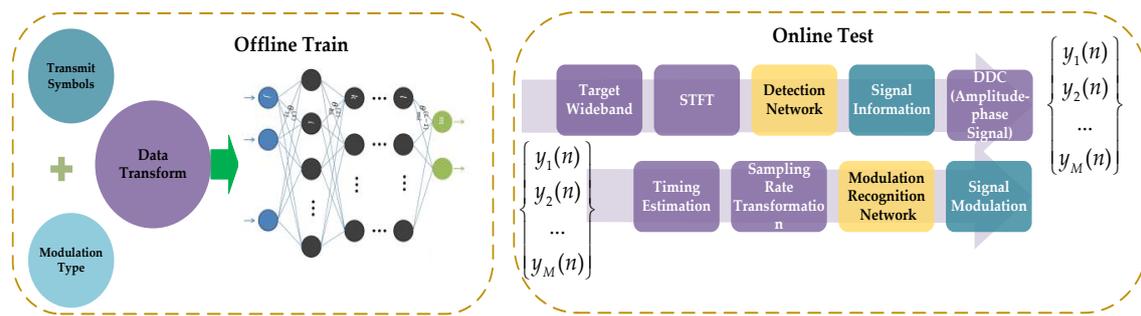


Figure 9. System model.

Step 1: We construct the signal detection network and modulation recognition network and train each network with their appropriate samples.

Step 2: In the online testing phase, we perform STFT for the wideband signals, and use the trained SSD networks to detect the signals in the time-frequency spectrum. In this step, we can obtain the center frequency and start-stop time of each signal. And for MFSK signal, we can get its modulation format.

Step 3: For the amplitude-phase modulation signal, we can get the central frequency and start-stop time in step 2. With this knowledge, we filter the target signal and use the envelope spectrum to estimate the signal symbol rate. Then we down convert the signal and perform the matched filter by using the estimated symbol rate.

Step 4: If the timing deviation exists in the target signal, it is necessary to extract the sample value at the optimum sampling position for signal eye diagram and vector diagram. We use the non-data-aided timing estimation algorithm in [52]. The specific expression is as follows, where L_0 is the length of the signal symbols, T is the sampling period and N is the oversampling number:

$$\hat{\tau} = \arg \left\{ \sum_{k=0}^{NL_0-1} \left| s\left(\frac{kT}{N}\right) \right|^2 e^{-j2\pi k/N} \right\} \quad (28)$$

Step 5: We alter the target signal sampling rate, and obtain the baseband signal with a maximum delay 32 sampling period. Moreover, we generate the eye-diagram and vector diagram with the processed signal.

Step 6: We use the trained modulation recognition network to identify the signal by its eye diagram and vector diagram. Finally, we complete the signal detection and modulation recognition.

4. Results

In Sections 2 and 3, we have discussed the methods which convert complex signal samples into images without noticeable information loss and introduced the structures of our DL framework. Table 1 shows the time complexity of our DL framework on the different process, in which N is the number of signals in the wideband range. It can be seen that our framework has low time complexity due to the evolution of GPUs, which is acceptable for many practical communication systems.

Table 1. The time complexity of the DL framework (ms).

Simple for Signal Detection	Signal Detection	Down Conversion	Simple for Modulation Recognition	Modulation Recognition
28.5	50.5	$10.3 \times N$	20.8	10.6

We also have conducted several experiments to demonstrate the performances of the DL framework for joint signal detection and modulation recognition in wireless communication systems. Our experiments can be divided into two parts: (1) performances on multi-signals detection and (2) performances on modulation recognition. The rest of this section is organized as follows

Multi-signals detection: First, we show some results of our detection network, and explain the reasons for these results. Then, we evaluate the model performances from three aspects: the modulation format, carrier frequency, and start-stop time. We also compare our network with other detection networks.

Signal modulation recognition: We evaluate our model recognition performances on each modulated signal. We also discuss the network performances when the frequency offset exists. Meanwhile, we compare our method with traditional methods and other DL based methods. Finally, we discuss the influence of symbol and eye numbers and compare the performance between signal-input networks and multi-inputs networks.

4.1. Performance on Signal Detection

For multi-signal detection, we need to know each signal carrier frequency, start-stop time and modulation format. Figure 10 shows some detection results from our model. From Figure 10a,b, it is indicated that our model is beneficial for multi-signals detection, and it can accurately estimate the relevant information about each signal. Moreover, our model has very a promising application prospect in engineering because it has a good visualization effect.

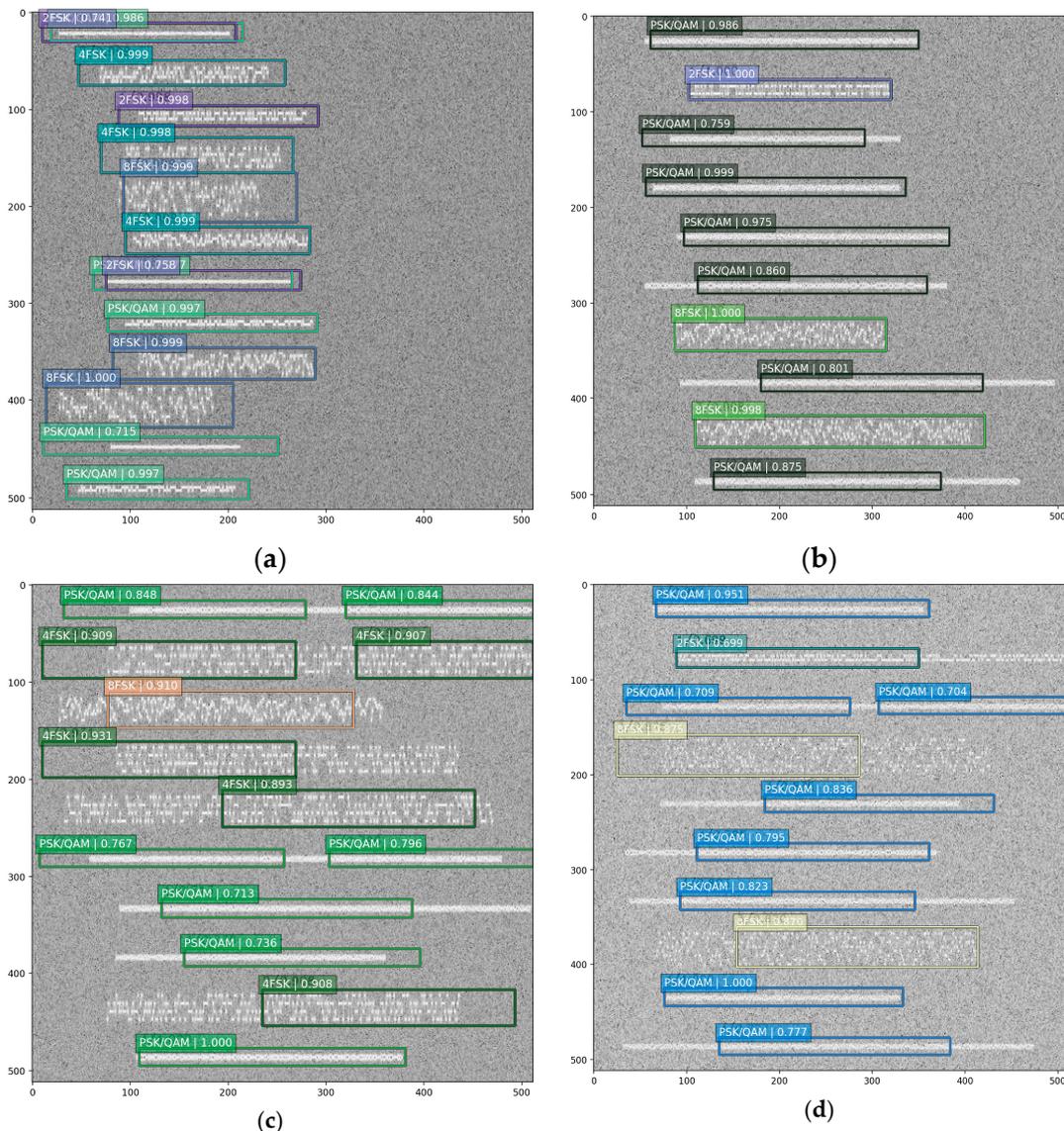


Figure 10. Cont.

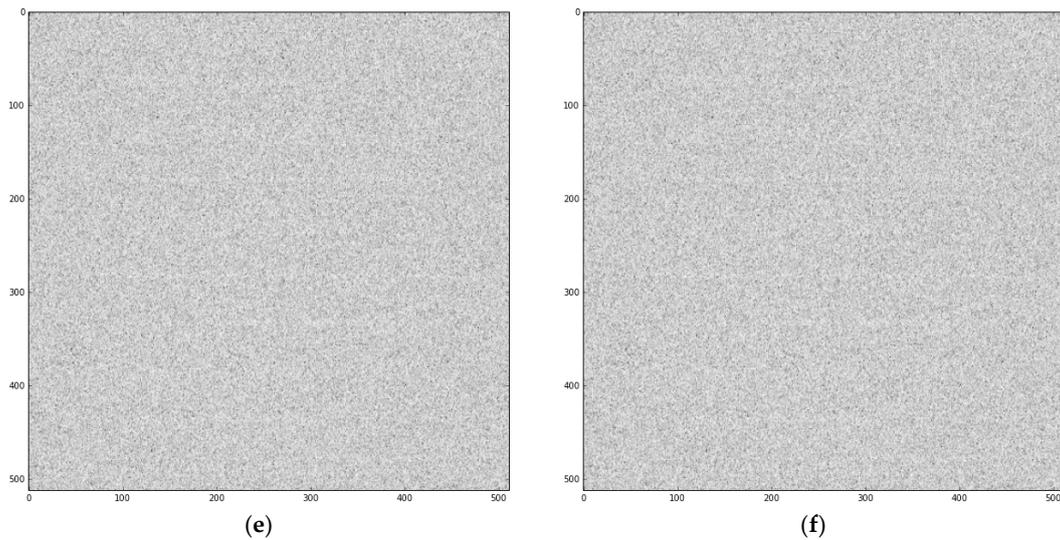


Figure 10. The SSD networks detection results. (a) ideal result 1; (b) ideal result 2; (c) not perfect result 1; (d) not perfect result 2; (e) no signal result 1; (f) no signal result 2.

To some extent, our model is not perfect yet, and there are still some aspects that need to be improved. From Figure 10c,d, we can learn that once the signal length is large, the estimation of the signal start-stop time is not precise, while the estimation of the carrier frequency is precise. The cause of this phenomenon may be that the time-frequency spectrum has large deformation and extreme length-width ratio, while the natural image is not. Therefore, we need to further optimize the default box in the SSD networks. Figure 10e,f show the network performance when there is no signal exists. It can be observed that the model does not produce a false alarm, which is useful in engineering.

Figure 11 shows our model detection precision versus different SNRs. We choose the mean Average Precision (mAP) as the performance index of the model. To calculate the mAP, we need to calculate precision and recall. For calculating precision and recall, we need to identify True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Recall is defined as the proportion of all positive examples ranked above a given rank. Precision is the proportion of all examples above that rank which are from positive. The Average Precision (AP) summarizes the shape of the precision/recall curve. Hence, the mAP is the mean of all the AP values across all classes as measured above. They can be calculated as follows

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \quad (29)$$

$$AP = \int_0^1 P(r)dr \quad (30)$$

$$mAP = \frac{\sum_{\text{num_classes}} AP_i}{\text{num_classes}} \quad (31)$$

It can be deduced that with the increase of the SNR, the mAP value of the SSD network is increasing. When the SNR is 5 dB, the mAP value can reach 90% in *IoU* is 0.5. Different *IoU* thresholds can lead to different results. Although the increase of the threshold can obtain more reliable signal carrier frequency and start-stop time, it sacrifices the precision of signal detection. Besides, we can adopt some traditional methods to further estimate these signal parameters. Finally, we choose 0.5 as the threshold of *IoU*.

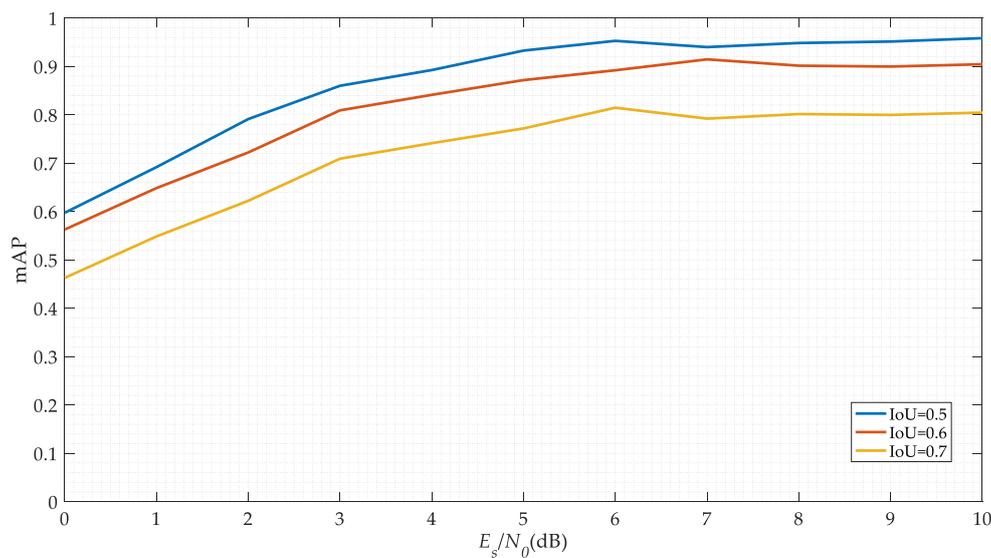


Figure 11. The SSD networks performances.

Once we detected the signals, we need to evaluate the precision of the estimated parameters. We use the normalized offset of the estimated and the actual parameters as the criterion of measurement. They can be presented as follows:

$$\Delta f = \frac{|f_{pre} - f_{real}|}{R} \quad (32)$$

$$\Delta t_{start} = \frac{|t_{pre_start} - t_{true_start}|}{T}, \Delta t_{stop} = \frac{|t_{pre_stop} - t_{true_stop}|}{T} \quad (33)$$

where f_{pre} is the predicted value of the carrier frequency, f_{real} is the actual value of the carrier frequency, R is the symbol rate, t_{pre_start} and t_{pre_stop} are the predict values of the start and the stop time, t_{true_start} and t_{true_stop} are the actual values of the start and the stop time, and T is the signal duration. Table 2 shows the carrier frequency and the start-stop time precision when the signal is detected. It can be seen that the precision of the carrier frequency is higher than start and stop time. These phenomena are consistent with Figure 10c,d. And in future research, we need to combine the prior information of the signal to design the default boxes and the networks.

Table 2. The offset in the estimation of the various parameters.

Offset	Carrier Frequency	Start Time	Stop Time
0 dB	2.0%	35.2%	38.6%
5 dB	1.3%	22.3%	21.4%
10 dB	0.5%	9.6%	8.7%

We also compare our model performances with the RCNN networks and the Fast RCNN. From Figure 12a, we can see that the mAP of the Fast RCNN and the RCNN is higher than the SSD networks, but the improvement is not significant. And from Figure 12b, we can infer that the SSD networks has considerable advantages in processing speed compared with the RCNN and Fast RCNN. Our model processing speed can reach 0.05 s for each time-frequency spectrum, and such a computational complexity is acceptable for many practical communications systems.

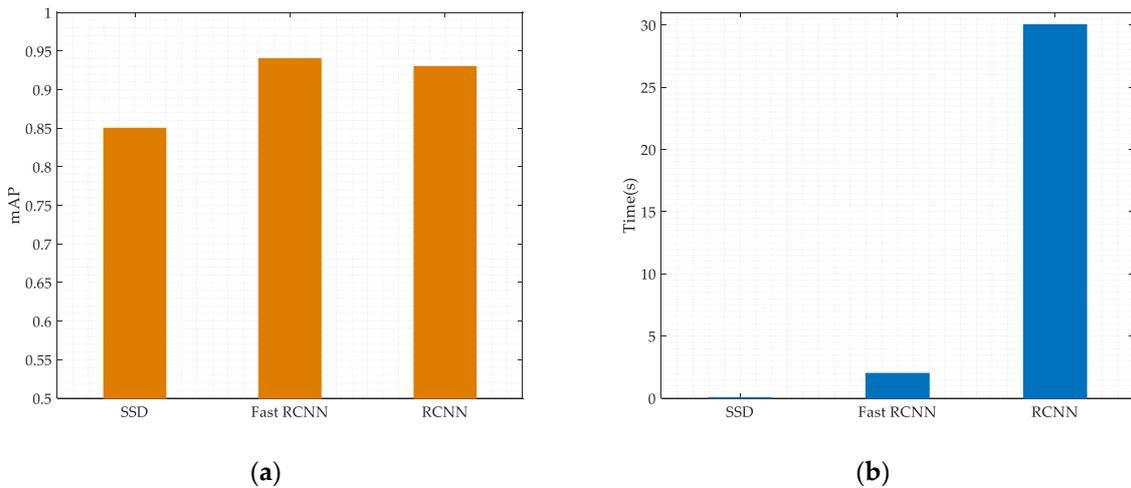


Figure 12. Different networks performances. (a) performances for mAP; (b) performances for time.

4.2. Performance on Modulation Recognition

For signal modulation recognition, we set a series of experiments to test the network performances. Figure 13a shows the recognition performances of each modulated signal under the different SNR. It can be seen that the algorithm can still achieve better performance when the SNR is very low. Because its modulation complexity, the performance of 64QAM signal is worse than other signals, but it still can achieve 94% accuracy at 7 dB. For BPSK and OQPSK signals, they have distinct visual characteristics from other modulated signals in the eye diagram and the vector diagram, which recognition accuracy can reach 100% even in 0 dB. And it is also obvious that the recognition performance of circular modulation signals {8PSK, 16APSK, 32APSK} is better than QAM modulation signal. To understand the results better, the confusion matrices in different SNR levels are presented in Figure 13b–d. It can be seen that the network shows excellent performance in discriminating BPSK, QPSK, OQPSK, 8PSK, and 16APSK. Moreover, in our experiments, it can be seen that 16QAM is more likely confused with 64QAM, while 16APSK is more likely confused with 64QAM.

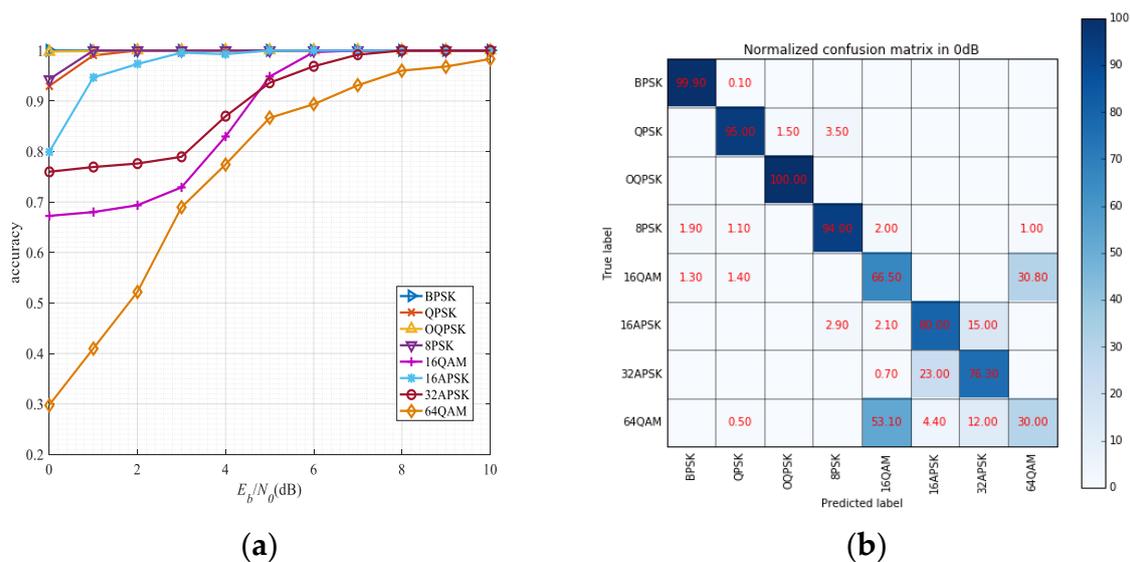


Figure 13. Cont.

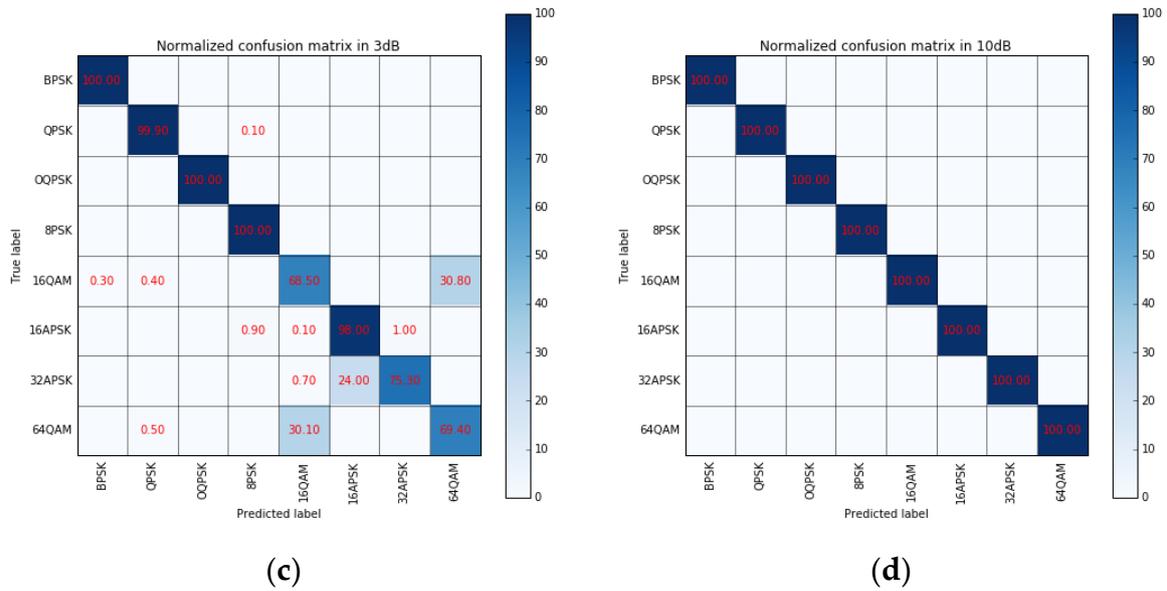


Figure 13. The performance of each modulation (a) classification accuracy for each modulation versus SNR; (b) normalized confusion matrix in 0 dB; (c) normalized confusion matrix in 3 dB; (d) normalized confusion matrix in 10 dB

For accuracy comparison, we consider four different modulation classification algorithms.

- (1) Cumulant: A traditional signal processing algorithm using the fourth-order cumulant C40 as the classification statistics [19].
- (2) SVM-7: An ML-based algorithm using the SVM with seven features, including three fourth-order cumulants C40, C41, and C42 and four sixth-order cumulants C60, C61, C62, and C63 [20].
- (3) CNNs for IQ waveform: A DL-based algorithm using the CNNs with the signal IQ waveform [4].
- (4) CNNs for constellation: An DL-based algorithm using the CNNs with the signal constellation [35].

Figure 14 presents the average classification accuracy of five algorithms versus SNR. The average accuracy is obtained by averaging the classification performance of eight modulation categories. The performance results of our algorithm outperform all other algorithms.

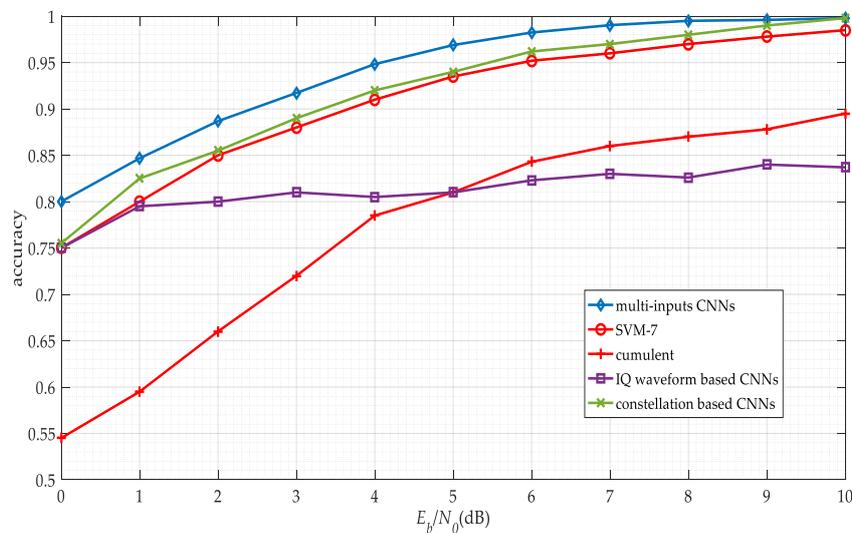


Figure 14. Different methods performance versus SNR.

Considering the error of the carrier frequency estimation by SSD networks and FFT in practice, we research the network recognition performance in different frequency offsets. We set a series of frequency offset for signals, and the result is shown in Figure 15. It can be seen that the recognition accuracy of the signals with a frequency offset is lower than those without frequency offset. When the signals have a large frequency offset, the network is no longer suitable. We also collect some signals from a real satellite communication system, and the real-time wireless channel is performed in the received signals. And then, we use a signal playback device, a DSP card, and PCs to simulate signal reception process. From Figure 15a, we can obtain that the recognition accuracy on real data is lower on simulated data at same SNR level. It may be due to the training data, which not consider the actual channel environment clearly. But the recognition accuracy can still reach 90% when the SNR is 4 dB. And for further research, we will make full use of the real signal to make our model more robust.

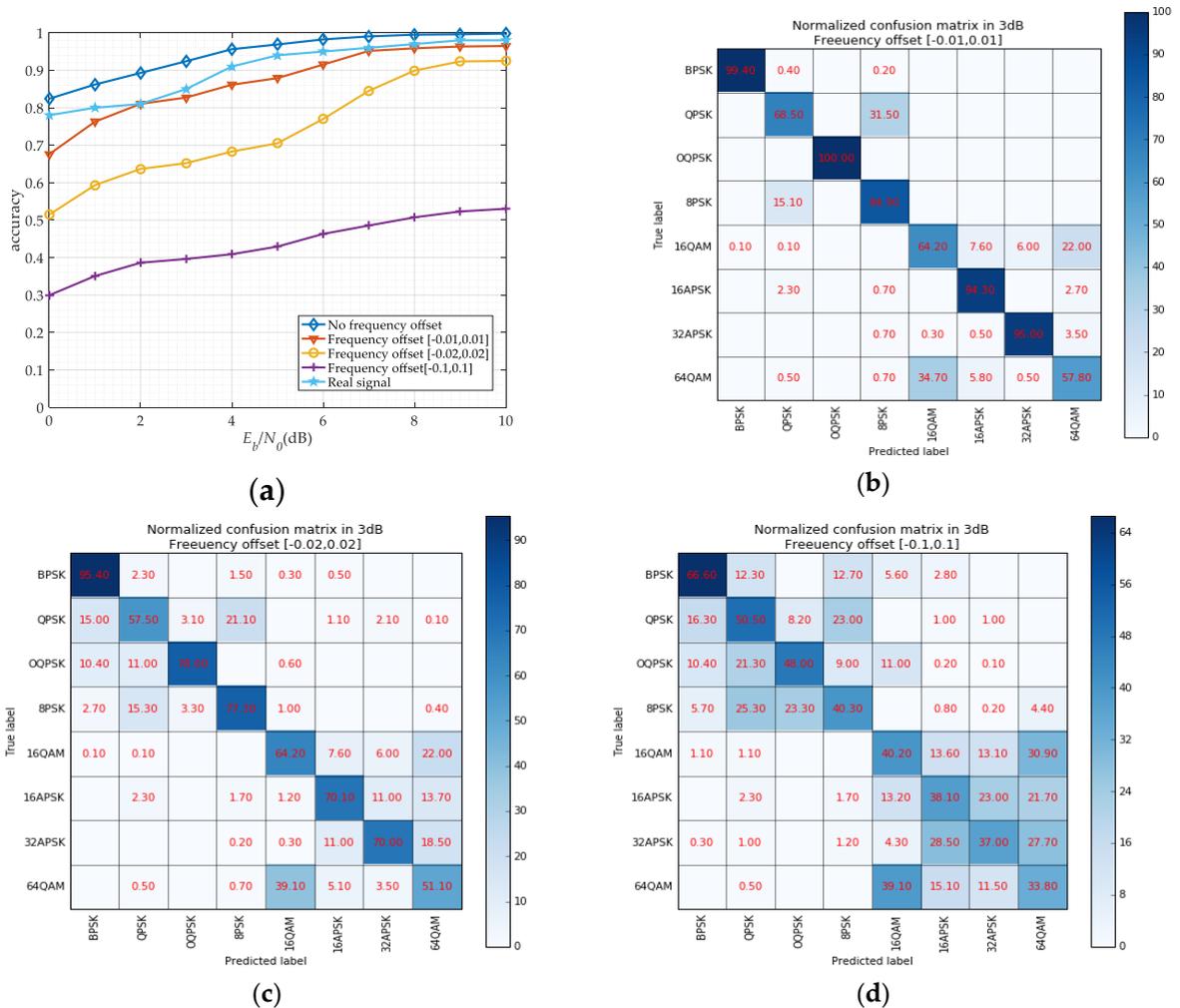


Figure 15. The network performance on different frequency offsets range. (a) classification accuracy for different frequency offset versus SNR; (b) normalized confusion matrix in 3 dB when the frequency offset is [-0.01, 0.01]; (c) normalized confusion matrix in 3dB when the frequency offset is [-0.02, 0.02]; (d) normalized confusion matrix in 10 dB when the frequency offset is [-0.1, 0.1].

We also consider the influence of the symbol numbers and the eye number in eye diagram on the network performance. We obtain the best parameter settings of samples by grid search. The symbol number is set as 200, 400, 800, and 1000, respectively, while the eye number is set as 2, 3, 4, and 5. The results are shown in Figure 16. It can be seen that these parameters do affect network performance. With the increase of symbol number and eye number, the overall accuracy of the model

is gradually increasing. But we also can see that when the symbol number is 1000 and the eye number is 5, the improvement of performance is not obvious. Therefore, we finally choose 800 symbols and 4 eye numbers to generate the eye diagram and the vector diagram.

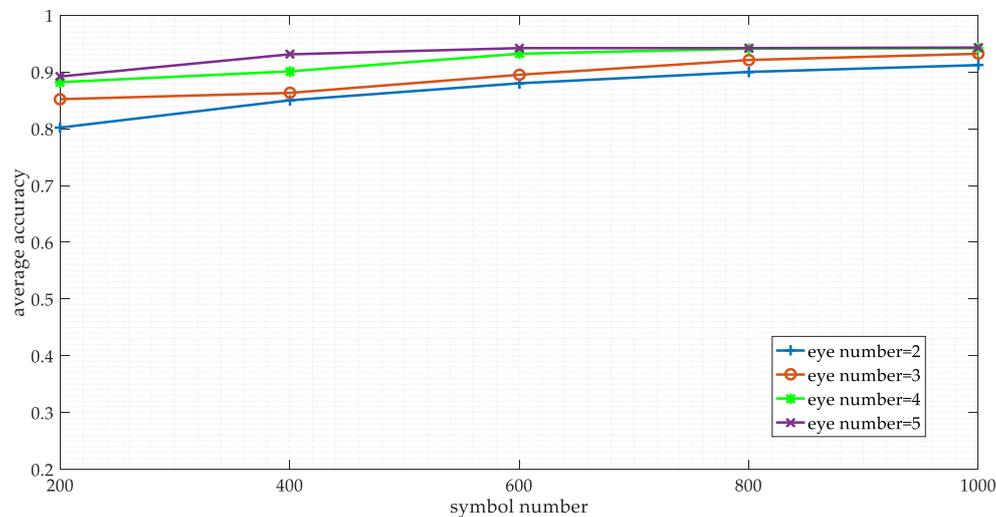


Figure 16. The network performance on the different sample parameters.

Finally, we compare the performance of the single input network with the multi-inputs network in this work. The results are shown in Figure 17. The modulation recognition algorithm based on a single eye diagram has poor performance. The performance of the I-eye diagram is lower than that of the Q-eye diagram, which may be due to the setting of the initial phase in the same modulation format. And the performance of the vector diagram based method is also inferior to our method, since it does not make full use of the signal waveform information.

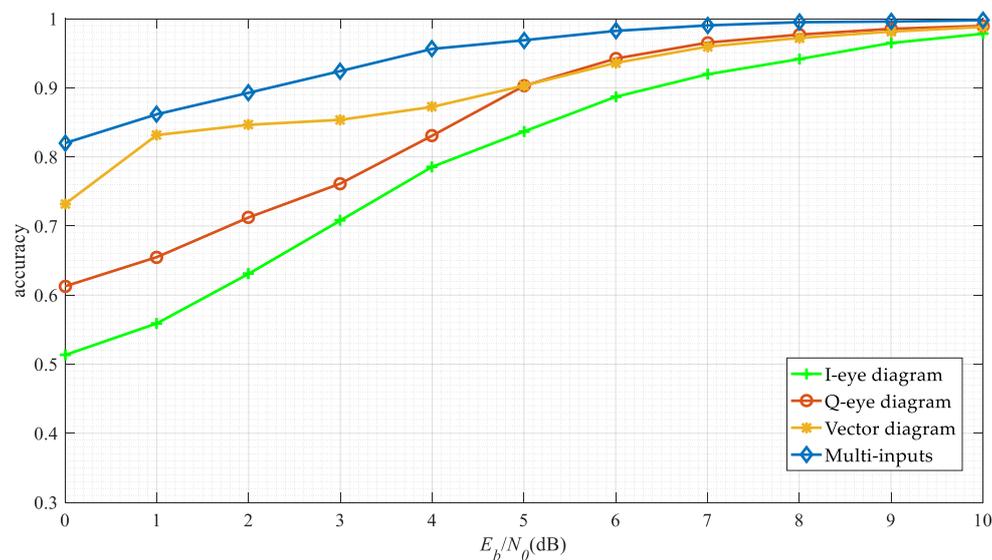


Figure 17. The network performance on the different input model.

5. Conclusions and Discussions

In our research, we have demonstrated our initial efforts to establish a DL framework for multi-signals detection and modulation classification problem. In our method, the time-frequency spectrums are exploited for multi-signals detection task, while the eye-diagrams and vector diagrams

are exploited for the modulation classification task. The simulation results prove that DL technologies have the ability to solve the problems in the communication field and have higher performance than other methods.

However, in the future, we will do more rigorous analysis and more comprehensive experiments. Besides, for practical use, we will collect the samples generated from the real channels, and then retrain or fine-tune the model for better performance.

Author Contributions: Conceptualization, X.Z. and H.P.; methodology, X.Z.; software, X.Z.; validation, X.Z., H.P. and X.Q.; formal analysis, G.L.; investigation, X.Z.; resources, G.L.; data curation, X.Z.; writing—original draft preparation, X.Q.; writing—review and editing, G.L.; visualization, S.Y.; supervision, X.Z.; project administration, H.P.; funding acquisition, H.P.

Funding: This research was funded by the National Natural Science Foundation of China (No. 61401511) and the National Natural Science Foundation of China (No. U1736107).

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

References

1. Mitola, J.; Maguire, G.Q. Cognitive radio: Making software radios more personal. *IEEE Pers. Commun.* **1999**, *6*, 13–18. [[CrossRef](#)]
2. Axell, E.; Leus, G.; Larsson, E.G.; Poor, H.V. Spectrum Sensing for Cognitive Radio: State-of-the-Art and Recent Advances. *IEEE Signal. Process. Mag.* **2012**, *29*, 101–116. [[CrossRef](#)]
3. Haykin, S. Cognitive radio: Brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **2005**, *2*, 201–220. [[CrossRef](#)]
4. O'Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional Radio Modulation Recognition Networks. In Proceedings of the International Conference on Engineering Applications of Neural Networks, Aberdeen, UK, 2–5 September 2016; pp. 213–226.
5. 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](#)]
6. Wu, H.C.; Saquib, M.; Yun, Z. Novel Automatic Modulation Classification Using Cumulant Features for Communications via Multipath Channels. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 3098–3105.
7. Fu, J.; Zhao, C.; Li, B.; Peng, X. Deep learning based digital signal modulation recognition. In Proceedings of the 2016 IEEE International Conference on Electronic Information and Communication Technology (ICEICT), Hohhot, China, 20–22 August 2016; pp. 955–964.
8. Salt, J.E.; Nguyen, H.H. Performance prediction for energy detection of unknown signals. *IEEE Trans. Veh. Technol.* **2008**, *57*, 3900–3904. [[CrossRef](#)]
9. Tadaion, A.A.; Derakhtian, M.; Gazor, S.; Nayebi, M.M.; Aref, M.R. Signal activity detection of phase-shift keying signals. *IEEE Trans. Commun.* **2006**, *54*, 1439–1445. [[CrossRef](#)]
10. Lehtomaki, J.J.; Vartiainen, J.; Juntti, M.; Saarnisaari, H. Analysis of the LAD methods. *IEEE Signal. Process. Lett.* **2008**, *15*, 237–240. [[CrossRef](#)]
11. Lehtomaki, J.J.; Vartiainen, J.; Juntti, M.; Saarnisaari, H. CFAR outlier detection with forward methods. *IEEE Trans. Signal. Process.* **2007**, *55*, 4702–4706. [[CrossRef](#)]
12. Macleod, M.D. Nonlinear recursive smoothing filters and their use for noise floor estimation. *Electron. Lett.* **1992**, *28*, 1952–1953. [[CrossRef](#)]
13. Salembier, P.; Liesegang, S.; Lopez-Martinez, C. Ship Detection in SAR Images Based on Maxtree Representation and Graph Signal Processing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2709–2724. [[CrossRef](#)]
14. Bao, D.; De Vito, L.; Rapuano, S. A histogram-based segmentation method for wideband spectrum sensing in cognitive radios. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 1900–1908. [[CrossRef](#)]
15. Bao, D.; De Vito, L.; Rapuano, S. Spectrum segmentation for wideband sensing of radio signals. In Proceedings of the 2011 IEEE International Workshop on Measurements and Networking, Anacapri, Italy, 10–11 October 2011; pp. 47–52.
16. Koley, S.; Mirza, V.; Islam, S. Gradient-Based Real-Time Spectrum Sensing at Low SNR. *IEEE Commun. Lett.* **2015**, *19*, 391–394. [[CrossRef](#)]

17. Mallat, S.; Zhong, S. Characterization of Signals from Multiscale Edges. *IEEE Trans. Pattern. Anal. Mach. Intell.* **1992**, *14*, 710–732. [[CrossRef](#)]
18. Nandi, A.K.; Azzouz, E.E. Algorithms for automatic modulation recognition of communication signals. *IEEE Trans. Commun.* **1998**, *4*, 431–436. [[CrossRef](#)]
19. Xie, L.J.; Wan, Q.; Swami, C.A.; Sadler, B.M. Hierarchical digital modulation classification using cumulants. *IEEE Trans. Commun.* **2000**, *48*, 416–429.
20. Aslam, M.W.; Zhu, Z.; Nandi, A.K. Automatic modulation classification using combination of genetic programming and KNN. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 2742–2750.
21. Xie, L.; Wan, Q. Cyclic Feature-Based Modulation Recognition Using Compressive Sensing. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 402–405. [[CrossRef](#)]
22. Shermeh, A.E.; Ghazalian, R. Recognition of communication signal types using genetic algorithm and support vector machines based on the higher order statistics. *Digit. Signal. Process.* **2010**, *20*, 1748–1757. [[CrossRef](#)]
23. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
24. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
25. Dolan, R.; DeSouza, G. GPU-based simulation of cellular neural networks for image processing. In Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009.
26. Jalil, B.; Leone, G.R.; Martinelli, M.; Moroni, D.; Pascali, M.A.; Berton, A. Fault Detection in Power Equipment via an Unmanned Aerial System Using Multi Modal Data. *Sensors* **2019**, *19*, 3014. [[CrossRef](#)] [[PubMed](#)]
27. Montavon, G.; Samek, W.; Müller, K.R. Methods for interpreting and understanding deep neural networks. *Digit. Signal. Process.* **2018**, *73*, 1–15. [[CrossRef](#)]
28. Reitter, D.; Moore, J.D. Alignment and task success in spoken dialogue. *J. Mem. Lang.* **2014**, *76*, 29–46. [[CrossRef](#)]
29. Yuan, Y.; Sun, Z.; Wei, Z.; Jia, K. DeepMorse: A Deep Convolutional Learning Method for Blind Morse Signal Detection in Wideband Wireless Spectrum. *IEEE Access* **2019**, *7*, 80577–80587. [[CrossRef](#)]
30. Ke, D.; Huang, Z.; Wang, X.; Li, X. Blind Detection Techniques for Non-Cooperative Communication Signals Based on Deep Learning. *IEEE Access* **2019**, *7*, 89218–89225. [[CrossRef](#)]
31. Mendis, G.J.; Wei, J.; Madanayake, A. Deep Learning based Radio-Signal Identification with Hardware Design. *IEEE Trans. Aerosp. Electron. Syst.* **2019**. [[CrossRef](#)]
32. Ali, A.; Yangyu, F. Unsupervised feature learning and automatic modulation classification using deep learning model. *Phys. Commun.* **2017**, *25*, 75–84. [[CrossRef](#)]
33. Meng, F.; Chen, P.; Wu, L.; Wang, X. Automatic Modulation Classification: A Deep Learning Enabled Approach. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10760–10772. [[CrossRef](#)]
34. Zheng, S.; Qi, P.; Chen, S.; Yang, X. Fusion Methods for CNN-Based Automatic Modulation Classification. *IEEE Access.* **2019**, *7*, 66496–66504. [[CrossRef](#)]
35. Peng, S.L.; Jiang, H.Y.; Wang, H.X. Modulation Classification Based on Signal Constellation Diagrams and Deep Learning. *IEEE Trans. Neural. Netw. Learn. Syst.* **2019**, *3*, 718–727. [[CrossRef](#)] [[PubMed](#)]
36. Tang, B.; Tu, Y.; Zhang, Z.; Lin, Y. Digital Signal Modulation Classification with Data Augmentation Using Generative Adversarial Nets in Cognitive Radio Networks. *IEEE Access.* **2018**, *6*, 15713–15722. [[CrossRef](#)]
37. Morello, A.; Mignone, V. DVB-S2: The second generation standard for satellite broad-band services. *Proc. IEEE* **2006**, *94*, 210–227. [[CrossRef](#)]
38. Griffin, D.; Lim, J. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust.* **1984**, *2*, 236–243. [[CrossRef](#)]
39. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *11*, 2278–2324. [[CrossRef](#)]
40. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
41. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. *arXiv* **2016**, arXiv:1612.08242.
42. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
43. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the 14th European Conference ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.

44. Fu, C.-Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. DSSD: Deconvolutional Single Shot Detector. *arXiv* **2017**, arXiv:1701.06659.
45. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
46. Girshick, R. Fast R-CNN. In Proceedings of the IEEE 2015 International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
47. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
48. Neubeck, A.; Van Gool, L. Efficient non-maximum suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 850–855.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
50. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
51. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
52. Morelli, M.; D'Andrea, A.N.; Mengali, U. Feedforward ML-based timing estimation with PSK signals. *IEEE Commun. Lett.* **1997**, *1*, 80–82. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).