Image-Based Deep Learning for Radio Signal Recognition

Authors: Zhou Xin, HE Xiaoxin, Changwen Zheng, Zhou Xin

Date: 2018-08-31T00:00:00+00:00

Abstract

This paper innovatively proposes a technical approach that utilizes image deep learning to solve radio signal recognition problems: firstly, radio signals are materialized into two-dimensional images, transforming the radio signal recognition problem into an object detection problem in the image recognition domain; furthermore, it fully leverages the advanced achievements of artificial intelligence in the field of image recognition to enhance the intelligence level of radio signal recognition and its capability in complex electromagnetic environments. Based on this approach, this paper proposes a radio signal recognition algorithm based on image deep learning—the RadioImageDet algorithm. Experimental results show that the algorithm can effectively recognize the waveform types and time/frequency coordinates of radio signals. On a dataset of 12 types with 4,740 field-collected samples, the recognition accuracy reaches 86.03%, the mAP value reaches 77.72, and the detection time on a moderately configured desktop computer is only 33 milliseconds, which fully validates the feasibility of the approach and the effectiveness of the algorithm.

Full Text

Preamble

Radio Signal Recognition Based on Image Deep Learning

ZHOU Xin※, HE Xiaoxin, ZHENG Changwen
(Science & Technology on Integrated Information System Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: This paper innovatively proposes a technical approach that leverages image deep learning to solve radio signal recognition problems. The core idea is to transform radio signals into two-dimensional images, thereby converting the radio signal recognition task into an object detection problem in the field of image recognition. This approach enables us to harness advanced achievements in artificial intelligence for image recognition, thereby enhancing the intelligence level and recognition capability of radio signal identification in complex electromagnetic environments. Based on this concept, we propose a radio signal recognition algorithm called RadioImageDet. Experimental results demonstrate that the algorithm can effectively identify waveform types and time/frequency coordinates of radio signals. On a self-collected dataset comprising 12 types of signals with 4,740 samples, the recognition accuracy reaches 86.03%, the mAP value reaches 77.72, and the detection time on a medium-configured desktop computer is only 33 milliseconds, fully validating the feasibility of the approach and the effectiveness of the algorithm.

Keywords: Radio Signal Recognition; Deep Learning; Radio Frequency Machine Learning (RFML); Convolutional Neural Network (CNN); Image Object Detection

1 Introduction

In recent years, with the development of wireless communication technologies, particularly the rise of the Internet of Things, radio waves have become a crucial carrier for connecting everything. However, the open nature of radio waves makes them vulnerable to interference and illegal exploitation, which can disrupt or even interrupt normal communication systems, posing threats to social and national security. Therefore, strengthening radio monitoring and identification, especially in key areas such as airports and border regions and at major event sites, holds significant practical importance and urgent demand.

Radio signal recognition technology was initially developed primarily for military electronic warfare and government spectrum regulation. With the increasing demand for spectrum sharing and the advancement of cognitive radio technology in recent years, it has gradually become a focus of attention in the commercial communication domain [1]. Early radio signal recognition involved using various spectrum acquisition devices to capture radio frequency signals in the air, processing them, and visualizing them as spectrograms, waterfall plots, or persistence diagrams, which were then analyzed by professionals to identify target signals based on their time-frequency characteristics [2]. This approach demands extremely high professional expertise from operators, and both efficiency and accuracy decline substantially when monitoring duration increases or when numerous radio signals are present.

Currently, the more popular approach to radio signal recognition involves having domain experts carefully select key signal features, such as maximum power spectral density [3], signal envelope kurtosis [4], instantaneous phase standard deviation [5], phase pulse count [4], cyclic spectrum [6-7], higher-order moments [8-9], and other time/frequency/spectrum domain characteristics. Traditional signal processing algorithms then compute these feature parameter values, and classification is performed based on fixed rules or machine learning methods [10]. Typically, a model requires selecting several or even dozens of these features as inputs, resulting in very high computational complexity that makes practical deployment difficult. Moreover, due to signal waveform diversity and multipath fading effects, this method lacks universality in feature selection and decision criteria.

It is well known that artificial intelligence possesses unique advantages in pattern recognition tasks and has achieved remarkable success in recent years across numerous fields including image/speech recognition, natural language processing, healthcare, and finance, attracting widespread attention from all sectors. Radio signal recognition is essentially a special form of "pattern recognition." Therefore, we believe that integrating artificial intelligence with traditional radio signal recognition technology, particularly using machine learning methods—especially deep learning algorithms—to automatically extract pattern features from radio waves can avoid manual feature extraction based on experience and improve recognition capability in complex electromagnetic environments.

Based on this vision, this study innovatively proposes a technical approach that uses image deep learning methods to solve radio signal recognition problems: first, radio signals are transformed into two-dimensional images, converting the radio signal recognition problem into an object detection problem in the field of image recognition; then, advanced achievements in artificial intelligence for image recognition are utilized to enhance the intelligence level and recognition capability of radio signal identification in complex electromagnetic environments. Following this approach, this paper proposes a radio signal recognition algorithm based on image deep learning—the RadioImageDet algorithm—and validates the feasibility and effectiveness of the approach through physical system experiments.

2 Related Work

As early as the 1990s, traditional machine learning algorithms such as Naive Bayes, neural networks, and decision trees were already being applied to radio modulation recognition [12-14]. However, limited by the technological development level at that time, these methods merely used machine learning algorithms to classify signals based on manually extracted features, as illustrated in Figure 1 [Figure 1: see original paper].

Figure 1 Traditional machine learning-based radio signal recognition workflow

In the past two years, this situation has gradually changed with the rapid development of computer technology and deep neural network technology. Inspired by the tremendous success of deep neural networks in image recognition, O' Shea and his team published multiple papers in 2016 investigating methods and issues related to applying deep learning technology to radio communication and identification [15-17], ushering in a new era for radio identification from an experience-driven artificial feature paradigm to a data-driven representation learning paradigm. In the same year, the U.S. Defense Advanced Research Projects Agency (DARPA) announced that it would launch the SC2

Challenge (Spectrum Collaboration Challenge) in 2017, aiming to "apply advanced machine learning capabilities to the radio domain to optimize spectrum usage strategies and alleviate electromagnetic spectrum congestion" [18]. This represents DARPA' s third challenge competition of this century, following the previous two.

In 2017, DARPA made frequent moves to promote the integration of AI (Artificial Intelligence) and radio. In March 2017, DARPA organized a "ModRec" competition at DySPAN, a renowned academic conference in the cognitive radio field, challenging participants to identify signal waveforms with more than 20 different modulation types randomly generated at the conference site, claiming it to be "the first step toward managing the electromagnetic spectrum in a more efficient manner." In August, DARPA launched the "Radio Frequency Machine Learning Systems" (RFMLS) project. As a complement to the SC2 project, RFMLS aims to "establish a radio frequency discrimination capability that can distinguish unique and special signals from a noisy, crowded spectrum" [19] for future use in both civilian and military domains.

Below, we introduce related research work from three aspects relevant to this study.

2.1 CNN-Based Modulation Recognition

In [15], O' Shea proposed an end-to-end CNN-based modulation recognition model. This model takes time-domain I/Q samples as input, processes them through two convolutional layers and two fully connected layers, and outputs the signal modulation type. The specific architecture is shown in Table 1 . The paper used GnuRadio to simulate and generate signals with 11 different modulation types as training and testing data. Through comparison with traditional machine learning methods such as Naive Bayes, K-Nearest Neighbors, decision trees, multi-layer perceptrons, and support vector machines (with the traditional algorithms using 32 features constructed from cyclic moments as input), the new algorithm was found to achieve performance comparable to the best traditional algorithms, demonstrating the effectiveness of end-to-end CNN for automatic modulation recognition.

Table 1 O' Shea Model Architecture

Building upon this, [20] employed a hierarchical recognition approach for improvement: first using a CNN network to distinguish between analog and digital modulation, then using another CNN network to identify specific modulation types based on this prior knowledge, and finally using a third CNN network to recognize the modulation order. References [21-22] improved from the network structure perspective, adding consideration of temporal features on top of the CNN network and adopting a CNN+LSTM network structure for modulation recognition. Reference [23] added some edge information (i.e., traditional cyclostationary feature parameters) as input before the fully connected layer in O' Shea' s model to assist CNN recognition. This method abandoned the premise of using only raw information as model input, combining traditional machine learning-based automatic modulation recognition with novel CNN networks. Although it improved recognition accuracy, it retained a strong expert system imprint and had high computational complexity.

2.2 Radio Individual Identification

References [15, 20-23] all belong to modulation recognition, whose output needs to be combined with other information to produce results directly useful for applications such as electronic reconnaissance and spectrum regulation. In contrast, references [24-27] represent a higher level of radio individual identification, whose results can be directly applied to these scenarios and thus better align with the "end-to-end" identification concept.

Reference [24] uses the minor eigenvalues of signal matrices to distinguish signals and performs radio signal recognition through matrix eigenvalue solving methods. Reference [25] utilizes nonlinear spurious emission characteristics of physical waveforms and RF devices to differentiate signal sources and employs phase space reconstruction (RPS) for analysis and identification. Reference [26] uses the similarity of signal power spectral density (PSD) to distinguish signals and adopts a neuro-fuzzy similarity measurement method for analysis and identification. All these references belong to traditional expert system-based radio identification methods. In [27], Schmidt proposed a CNN-based radio communication standard recognition model with a structure similar to O' Shea' s model but taking frequency-domain Fast Fourier Transform (FFT) results as input and directly outputting radio communication standard names. The paper involved three communication standards: 802.11, 802.15.1, and 802.15.4. Reference [28] also used a similar CNN network but took Choi-Williams time-frequency distribution (CWD) transformation results as model input to identify eight common radar waveforms including Linear Frequency Modulation (LFM), Frank sequences, and Costas codes. This study also belongs to radio individual identification, with input being spectrogram waterfall images containing time/frequency information and output being signal waveform names and their time/frequency coordinates.

2.3 Deep Learning-Based Image Object Detection

The task of image object detection is to find all objects of interest in an image and determine their positions and sizes, representing one of the core problems in computer vision. Before 2013, traditional object detection methods represented by DPM (Deformable Part Model) held a dominant position. In 2014, R. Girshick proposed the R-CNN (Region proposal CNN) algorithm framework [29], achieving a major breakthrough in object detection accuracy and initiating a research boom in deep learning-based object detection. Subsequently, a series of research results emerged, including Fast R-CNN [30], Faster R-CNN [31], FPN [32], YOLOv2 [33], and SSD [34], showing significant improvements in detection speed and accuracy. Currently, deep learning-based image object detection is mainly divided into two categories: one is region proposal-based object detection algorithms represented by R-CNN and Faster R-CNN, and the other is regression-based end-to-end object detection algorithms represented by YOLOv2 and SSD. Since the former requires dividing the detection process into two steps—region proposal and detection classification—while the latter directly implements object localization and recognition in a single convolutional network, simplifying the network structure and achieving real-time detection speed while ensuring accuracy, this study will use YOLOv2 as the foundation for designing the radio signal recognition model.

3 RadioImageDet Algorithm Model Design

This paper proposes a radio signal recognition algorithm based on image deep learning—the RadioImageDet algorithm. This algorithm is based on an end-to-end convolutional neural network that can simultaneously identify multiple signals' waveform types in complex electromagnetic environments with time/frequency aliasing and locate their start/end times and frequency ranges. The algorithm structure is shown in Figure 2 [Figure 2: see original paper].

The RadioImageDet algorithm consists of two parts: a preprocessing module and an end-to-end object detection module. The preprocessing module takes raw I/Q sampling data from an observation window as input, transforms time-domain data into more recognizable frequency-domain representation through Discrete Fourier Transform (DFT), and then converts continuous DFT results into a two-dimensional spectrogram waterfall image through data image processing. The end-to-end object detection module takes the spectrogram waterfall image obtained from preprocessing as input, extracts concrete signal features through CNN feature extraction layers, and then transforms these abstract features into meaningful output results such as signal waveform types and time/frequency coordinates through the object detection layer.

Figure 2 RadioImageDet algorithm structure diagram

3.1 Preprocessing Module

The RadioImageDet algorithm first performs data preprocessing. The purpose of preprocessing is to transform raw I/Q sampling data into a representation form that is more conducive to automatic feature extraction and recognition. To fully leverage the advantages of convolutional neural networks in feature extraction and reduce computational complexity, we desire the preprocessing process to be lossless and computationally inexpensive. According to discrete signal processing theory and modern communication system principles, raw I/Q sampling data is the fundamental input for backend processing in digital communication systems. Any complex digital signal processing algorithm is completed through mathematical transformation based on this data, so it naturally contains lossless information. This is also why O' Shea et al. directly selected I/Q sampling data as neural network input [15, 20]. Mathematically, Discrete Fourier Transform (DFT) is a lossless transformation of finite-length I/Q sampling data, as shown in equation (1), and DFT transformation results, as a frequency-domain representation of signals, are more conducive to radio signal recognition based on electromagnetic radiation [35].

Therefore, this paper first performs DFT transformation on raw I/Q sampling data to obtain the signal' s frequency-domain expression, followed by subsequent processing—consistent with the approach in [26, 27].

The result of a single DFT operation is a one-dimensional complex sequence that can be visualized as "a line," representing the spectral density at a certain time point. However, radio signal waveforms always change over time, and observing features at only one "time point" makes accurate signal recognition difficult. Therefore, stacking and combining observation results from multiple consecutive time points into "a surface" can simultaneously express time-domain and frequency-domain information features, helping to improve signal recognition rates. In this two-dimensional plane, the horizontal axis represents frequency, the vertical axis represents time, and element values are the DFT transformation results.

If element values are mapped according to image grayscale values, a visual picture can be drawn—this picture is the spectrogram waterfall image, as specifically shown in Figure 3 [Figure 3: see original paper]. Therefore, this paper adds a data image processing step after DFT transformation, combining continuous multiple DFT transformation results into the spectrogram waterfall format to increase the time-domain information content of individual samples and facilitate subsequent convolutional neural network processing and signal recognition.

Figure 3 Spectrogram waterfall synthesis schematic diagram

Theoretically, the size of a spectrogram waterfall image is M×N×2, where 2 indicates the image has two channels (I channel and Q channel), N represents the number of points in the discrete Fourier transform, and M represents the number of DFT result groups combined in one spectrogram waterfall image. The M value reflects the time span of a sample—the larger the M value, the longer the represented time span and the more complex the computation. For computational convenience, this paper makes the following three simplifications:

1. To simplify image conversion operations, the I and Q channels are directly merged into one channel according to equation (2), with the merged physical meaning being power spectral density (PSD);
2. To utilize the Fast Fourier Transform (FFT) algorithm to accelerate computation, the N value must be a power of 2, and this paper takes N=512;
3. To match certain CNN-based image object detection models, M is set equal to N to form a square image.

Therefore, the simplified spectrogram waterfall image size is N×N, where N is a power of 2.

3.2 End-to-End Object Detection Module

After preprocessing, abstract radio signals are transformed into concrete two-dimensional images. The next step is to extract features from these images and identify the signals and their locations. The nature of this task is identical to object detection tasks in computer vision, with the only difference being that these images are spectrogram waterfall images synthesized from DFT-transformed radio signals rather than natural images. Due to the special nature of the input images, there may be more suitable neural network structures for recognition and object detection, but this paper does not delve deeply into this aspect for now. Instead, it constructs a radio signal recognition network model based on the currently popular YOLOv2 model [33] with simple parameter modifications, called RadioYOLO.

As shown in Figure 4 [Figure 4: see original paper], the RadioYOLO model structure mainly references the Inceptionv1 network, using only 3×3 and 1×1 convolutional kernels to compress model parameters as much as possible, and doubling the number of feature channels after each pooling operation to preserve feature completeness as much as possible. Simultaneously, drawing on Inceptionv2' s ideas, batch normalization is applied to the output of each convolutional layer, increasing model robustness and training speed while replacing dropout to prevent overfitting. Borrowing from ResNet' s skip connection concept, the output of conv13 is reshaped and directly concatenated with conv20 in the channel dimension before being output to the next layer, enabling the output layer to retain higher-resolution features and reducing the probability of "gradient vanishing." The leakyReLU nonlinear activation function is used to reduce the probability of neuron death during training.

Compared with the YOLOv2 model, RadioYOLO' s main modifications include:

1. **Model input image size**: Changed to 512×512 according to the FFT transformation size;
2. **Normalized anchor box sizes**: Selected 5 anchors (width, height) based on common characteristics of radio signals:
    a) (0.015, 0.9): matches narrowband continuous signals
    b) (0.5, 0.9): matches wideband continuous signals
    c) (0.025, 0.05): matches narrowband transient signals
    d) (0.469, 0.01): matches wideband transient signals
    e) (0.469, 0.1): matches wideband transient signals
3. **Model output tensor size**: Derived a new output size of 16×16×85 based on input size and dataset characteristics.

Table 2 RadioYOLO Model Architecture (each convolutional layer includes batch normalization and leakyReLU activation; size values omit channel numbers)

As shown in Table 2, RadioYOLO contains 23 convolutional layers and 5 pooling layers, with input being a 512×512×1 spectrogram waterfall image and output being a 16×16×K tensor, where K is calculated as shown in equation (3).

Here, anchor_num represents the number of anchors, and class_num represents the number of signal types to be recognized. In the verification experiments of this paper, anchor_num=5 and class_num=12, so K=85.

The 5 in parentheses indicates that besides signal types, there are 5-dimensional feature parameter prediction values [tx, ty, tw, th, to], representing the top-left coordinates (bx, by), border size (bw, bh), and confidence (conf) of signal bounding boxes. To improve model training stability, the predicted values tx, etc., are not real values but normalized intermediate values. Their relationship with real values satisfies equation (4). Therefore, after obtaining the predicted feature tensor through the neural network model, the final real prediction values still need to be calculated through equation (4) to complete object detection.

4.1 Data Collection

In previous radio recognition research literature, radio data typically came from simulation systems such as MATLAB simulation software, GnuRadio simulation software, and signal generators. In laboratory environments, single or mixed noisy signals were generated through manual code control [15-17, 20-23, 27-28]. This approach reduces the difficulty of data acquisition and preprocessing, facilitating data annotation using prior information to complete algorithm verification. However, due to complex factors such as channel fading, clock jitter, unknown radiation sources, and receiver RF characteristics, there are often significant differences between simulation data and real electromagnetic environments, affecting the authenticity and practicality of algorithms.

Therefore, this study developed a high-speed spectrum acquisition system using the GnuRadio+USRP software radio platform to collect real-world spectrum data, as shown in Figure 5 [Figure 5: see original paper]. USRP devices are open-source software radio modules designed and produced by National Instruments (NI), adopting the AD9361 RF integrated chip solution with operating frequency bands covering 70MHz to 6GHz. GnuRadio is open-source software that uses ordinary computers to implement modulation/demodulation, encoding/decoding, and other digital signal processing functions. When used with USRP, it can quickly build wireless communication principle verification systems. Through optimized data storage algorithms, this system can support minute-level lossless storage of signals with up to 50MHz instantaneous bandwidth and can automatically sweep and store data on preset frequency sets.

Figure 5 High-speed spectrum acquisition system

All training and testing data in this paper come from field spectrum sampling using this platform. The sampling rate is 40 Msps, with single-frequency sampling duration of approximately 5 seconds. Sampling locations include seven sites in three areas: Beijing Zhongguancun, Tianjin Binhai New Area, and Tianjin Bohai Bay. The signals of interest are 12 types of common 2G/3G/4G cellular downlink signals in mainland China. Within frequency bands containing signals of interest, one sampling frequency point was established for every 10 MHz offset, totaling 34 frequency points, with a total dataset size of approximately 650 GB.

Table 3 shows the 2G/3G/4G cellular downlink frequency allocation table for the three major operators in Beijing Zhongguancun area. This table was compiled based on relevant standard documents and calibrated through actual measurement observations using instruments. Combined with measured results from Beijing, Tianjin, and other locations, it was found that this table is not universally applicable within mainland China. Operators may independently adjust frequency resource allocation under certain conditions based on the number of network-accessed users and traffic characteristics in different regions, making signal types at the same frequency point potentially different across regions. While this has no impact on the core algorithm of this study, it does create some difficulties for data annotation work.

Table 3 Beijing area three major operators 2G/3G/4G cellular downlink frequency allocation table

4.2 Data Preprocessing and Annotation

The raw data format is time-domain I/Q sampling, which must first be converted into spectrogram waterfall images before it can be recognized and processed by the RadioYOLO model. This study used Python language and relevant third-party libraries to write a series of scripts following the preprocessing procedure described in Section 3.1, transforming the collected 650 GB of time-domain I/Q sampling data into 4,740 grayscale images of 512×512 pixels, including 3,792 training images and 948 testing images.

Subsequently, training and testing samples need to be annotated. Data annotation is inherently a tedious manual task, but based on the characteristic that signals to be recognized in this dataset have fixed operating frequency bands and persistently exist, we designed a semi-automatic data annotation method that significantly reduces repetitive work. This method first manually annotates 2 spectrogram images at each site and frequency point, then automatically annotates the remaining spectrogram images at that site and frequency point using software, with annotation coordinates randomly jittering within the range of the two manually annotated coordinates. An annotation example is shown in Figure 6 [Figure 6: see original paper]. Within a 40 MHz range centered at 2140 MHz, there exists 1 Tele4G_DL_FDD downlink signal (partially visible), 2 Uni3G_DL_WCDMA downlink signals, and 1 Uni4G_DL_FDD downlink signal. By annotating visual bounding boxes, the software automatically generates text annotation information for the image in the background according to a specific format.

It is worth noting that although data collection and annotation in this study are frequency-based, and for authorized communication systems such as 2G/3G/4G cellular systems, signal recognition could almost be completed using frequency information alone, this is not universally applicable and cannot complete the verification of core technology. Therefore, during model training and detection, we deliberately avoided center frequency information and only processed digital baseband signals within the receiver passband. This means that even if a signal is tuned to an unknown unauthorized frequency band, it will not affect the detection results of this algorithm.

Figure 6 Data annotation example (2140 MHz frequency point)

5 Results and Analysis

To verify algorithm performance, we implemented the RadioYOLO model and the entire RadioImageDet algorithm using Python under the TensorFlow framework, and conducted model training and testing. The dataset comes from preprocessed real spectrum data, containing 12 types of signals, with 3,792 training samples and 948 testing samples. Specific data characteristics are detailed in Chapter 3. During model training, we used the first 80% of the training set for training and the remaining 20% for model validation during training. The model training optimizer used the adam algorithm, with epoch limit set to 50 and batch_size set to 8. From actual operation results, under the condition of using NVIDIA GTX1060T graphics card acceleration, the training speed is approximately 600 milliseconds per batch, with each epoch taking about 230 seconds. The testing speed for a single sample is approximately 33 milliseconds, with a detection frame rate of 30 frames per second, basically meeting real-time application requirements.

The recognition result output is shown in Figure 7 [Figure 7: see original paper]. The model software automatically draws bounding boxes on the original image based on recognition results, indicating signal types and recognition confidence. As can be seen from the figure, some signals have very obvious features on spectrogram waterfall images that can be easily judged by human eyes—this is the basic premise and motivation for this research. From the detection results, signals that can be easily identified by human eyes can also be correctly identified by the algorithm model in most cases, though there are some exceptions, as shown in Figure 7-(g). For signals not easily recognizable by human eyes, the algorithm model has a higher error probability, as shown in Figure 7-(b), but can also correctly identify them in some cases, as shown in Figure 7-(e). These phenomena intuitively prove the feasibility and effectiveness of using image deep learning methods for radio signal recognition!

Figure 7 Model test result samples

For more rigorous evaluation of algorithm performance, we need more quantitative metrics. As shown in Figure 8 [Figure 8: see original paper], for binary classification problems, the entire sample set can be divided into four quadrants based on real and predicted values, shown as gray areas in the figure: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

Through simple extension, the binary confusion matrix can also be transformed into a multi-class confusion matrix. Based on the confusion matrix, various classifier performance evaluation metrics can be constructed.

Figure 8 Confusion matrix for binary classification problems

Accuracy is the most common evaluation metric, defined as accuracy = (TP+TN)/(P+N), representing the proportion of correctly classified samples among the total sample count. Generally speaking, higher accuracy indicates a better classifier. Figure 9 [Figure 9: see original paper] shows the confusion matrix of our model' s test results, where the vertical axis represents true labels and the horizontal axis represents predicted labels. Within the 12×12 coordinate grid, box colors represent values—darker colors indicate larger values. Generally, darker colors on the confusion matrix diagonal are better, while lighter colors at other coordinates are better. As can be seen, the model performs well for most signal recognitions but tends to misidentify Uni4G_DL_FDD1800 and Uni4G_DL_FDD2100 as Tele4G_DL_FDD2100 and Tele4G_DL_FDD2100, respectively. Based on the confusion matrix and definition formula, the model' s recognition accuracy is calculated to be 86.04%.

Figure 9 Test results—confusion matrix for 12-class signal recognition

Although accuracy is an intuitive evaluation metric, it sometimes cannot fully reflect algorithm quality, especially when sample data distribution is imbalanced. Therefore, we introduced the more powerful PR curve. The PR curve, or Precision-Recall curve, has Precision as the vertical coordinate and Recall as the horizontal coordinate. Precision, also called positive predictive value, is defined as precision = TP/(TP+FP), representing the proportion of actual positive examples among those classified as positive. Recall, also called sensitivity, is defined as recall = TP/(TP+FN), representing the proportion of correctly classified positive examples among all positive examples. The PR curve can intuitively demonstrate classification quality and is widely used in classification and retrieval fields. Generally, curves closer to coordinate (1,1) indicate better classifier performance.

However, the PR curve is ultimately a vector and not convenient for numerical comparison, so we typically use Average Precision (AP), which is the area under the PR curve, for classification effect evaluation. For a classifier that can often identify multiple targets simultaneously, while a single PR curve and its AP value can only represent the classification effect for one target type, we can average all AP values to obtain the mAP value. Currently, mAP has become one of the core performance metrics for detection algorithms.

Figure 10 [Figure 10: see original paper] shows the PR curve for 12-class signal recognition test results, and Table 5 shows the corresponding AP values. From Figure 10, although Uni4G_DL_FDD1800 and Uni4G_DL_FDD2100 have high recognition precision (both above 90%), their recall rates are relatively low (approximately 30% and 60%, respectively), resulting in poor overall AP metrics of only 26.02 and 58.50. In other words, if the classifier determines a signal to be Uni4G_DL_FDD1800 or Uni4G_DL_FDD2100, it is likely correct, but many detection misses will occur. Tele4G_DL_FDD2100 shows the opposite situation, with recall reaching 100% but precision less than 60%, meaning the classifier can detect all Tele4G_DL_FDD2100 signals but will also produce many false positives. Mobi3G_TDSCDMA exhibits characteristics of both, with moderate recall and precision, resulting in average overall recognition performance. Apart from these four signal types, the other eight signals show good recognition performance, with some signals' AP values even exceeding 95. The average AP value (mAP) for all signal recognitions reaches 77.72, which is very close to the current best metric of 78.8 in the natural image object detection field [31].

Table 5 Test results—AP values

6 Conclusion

As an important research field in wireless communications, radio signal recognition technology finds wide application in military electronic warfare, government spectrum regulation, and commercial spectrum sharing access. Considering its essence as a "pattern recognition" task and the unique advantages of artificial intelligence in pattern recognition tasks, this paper proposes a radio signal recognition algorithm based on image deep learning—the RadioImageDet algorithm. This algorithm first uses traditional digital signal processing methods to transform I/Q sampling data streams into a two-dimensional spectrogram waterfall image, converting the radio signal recognition problem into an object detection problem in the field of image recognition. It then leverages advanced achievements in artificial intelligence for image recognition, constructing an end-to-end deep neural network model for radio signal recognition called RadioYOLO based on the YOLOv2 model, and uses this model to identify signal waveform types and time/frequency coordinates. This approach of converting radio signal recognition into an image object detection problem is the first of its kind proposed domestically and internationally, opening a new avenue for future research on intelligent radio signal recognition technology in complex electromagnetic environments.

To test the algorithm' s real performance, this study developed a high-speed spectrum acquisition system based on the GnuRadio and USRP platform and collected spectrum data in Beijing, Tianjin, and other locations to build a radio dataset. Experimental results demonstrate that the RadioImageDet algorithm can effectively recognize concretized radio signals with recognition performance approximating human visual capability. On a self-collected dataset comprising 12 types of signals with 4,740 samples, the recognition accuracy reaches 86.03%, the mAP value reaches 77.72, and the detection time on a medium-configured desktop computer is only 33 milliseconds, fully validating the feasibility of the approach and the effectiveness of the algorithm.

The RadioImageDet algorithm represents only a preliminary first attempt at the technical approach of radio signal recognition based on image deep learning, and some unsatisfactory aspects can be observed from the experimental results. In the future, we will further investigate other signal concretization methods and neural network models more suitable for radio signal characteristics.

References

[1] Octavia A. Dobre. Signal Identification for Emerging Intelligent Radios: Classical Problems and New Challenges [J]. IEEE Instrumentation & Measurement Magazine, 2015, 18(2): 11-18. [2] 姚亚峰, 黄载禄. 通信信号调制识别技术 [J]. 通信技术, 2003, 6(1): 41-50. [3] E. E. Azzouz, A. K. Nandi. Automatic identification of digital modulation types [J]. Signal Processing, 1995, 47(1): 55-69. [4] J Lopatka, M Pedzisz. Automatic Modulation Classification Using Statistical Moments and a Fuzzy Classifier [C]. International Conference on Signal Processing Proceedings, 2000, 3(1): 1500-1506. [5] A. K. Nandi, E. E. Azzouz. Algorithms for Automatic Modulation Recognition of Communication Signals [J]. IEEE Transactions on Communications, 1998, 46(4): 431-436. [6] J. Reichert. Automatic Classification of Communication Signals Using Higher Order Statistics [C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992, 5:221-224. [7] C. M. Spooner, A. N. Mody, J Chuang. Modulation Recognition Using Second- and Higher-Order Cyclostationarity [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1: 1-3. [8] S. S. Soliman, S. Z. Hsue. Signal Classification Using Statistical Moments [J]. IEEE Transactions on Communications, 1992, 40(5): 908-916. [9] Y. Yamashita, H. Ochiai. A Classification of OFDM Signals With or Without DFT Precoding Based on High-Order Moment [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1: 1-2. [10] 陈健, 阔永红, 李建东，等. 通信信号自动识别方法 [J]. 电路与系统学报, 2005, 10(5): 102-109. [11] D. L. Parnas. The Real Risks of Artificial Intelligence [J]. Communications of the ACM, 2017, 60(10):27-31. [12] M. L. D. Wong, A. K. Nandi. Automatic Digital Modulation Recognition Using Spectral and Statistical Features with Multi-layer Perceptron [C]. Signal Processing and its Applications, 2001, 2: 390-393. [13] G. Arulampalam, V. Ramakonar, et al. Classification of Digital Modulated Schemes Using Neural Networks [C]. International Symposium on Signal Processing and its Applications, 1999, 2: 649-652. [14] K. Triantafyllakis, M. Surligas, G. Vardakis. Phasma: An Automatic Modulation Classification System Based on Random Forest [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1: 1-3. [15] T. O' Shea, J. Corgan, T. C. Clancy. Convolutional Radio Modulation Recognition Networks [C]. International Conference on Engineering Applications of Neural Networks, 2016, 1:213-226. [16] T. O' Shea, J. Corgan, T. C. Clancy. Unsupervised Representation Learning of Structured Radio Communication Signals [C]. First International Workshop on Sensing, Processing and Learning for Intelligent Machines, 2016, 1:1-5. [17] T. O' Shea, N. West, M. Vondal, et al. Semi-supervised Radio Signal Identification [C]. International Conference on Advanced Communication Technology, 2017, 1:33-38. [18] J. L. Ziegler, R. T. Arn, W. Chambers. Modulation Recognition with GNU

radio, Keras and HackRF [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1:1-3. [19] 于金华, 周羽丰, 陈柱文. Recent Trends of DARPA—RFMLS [Online]. http://m.sohu.com/a/192607899_466078. 2017. [20] K. Karra, S. Kuzdeba, J. Petersen. Modulation Recognition Using Hierarchical Deep Neural Networks [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1: 1-3. [21] N. West, T. O' Shea. Deep Architectures for Modulation Recognition [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1:1-6. [22] N. West, K. Harwell, B. McCall. DFT Signal Detection and Channelization with a Deep Neural Network Modulation Classifier [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1: 1-3. [23] K. Arumugam, I. Kadampot, M. Tahmasbi, et al. Modulation Recognition Using Side Information and Hybrid Learning [C]. IEEE International Symposium on Dynamic Spectrum Access Networks, 2017, 1: 1-2. [24] L. Huang, J. Yu, Z. Shen, et al. Radio Individual Identification Via Stable Communication Signals Based On Subordinate Component Analysis [C]. IET International Radar Conference, 2016, 1:1-4. [25] Y. Yuan, Z. Huang, F. Wang, et al. Radio Specific Emitter Identification Based on Nonlinear Characteristics of Signal [C]. IEEE International Black Sea Conference on Communications and Networking, 2015, 1:77-81. [26] K. Ahmad, G. Shresta, U. Meier, et al. Neuro-Fuzzy Signal Classifier (NFSC) for Standard Wireless Technologies [C]. International Symposium on Wireless Communication Systems, 2010, 1:616-620. [27] M. Schmidt, D. Block, U. Meier. Wireless Interference Identification with Convolutional Neural Networks [J]. IEEE 15th International Conference on Industrial Informatics, 2017, 1:180-185. [28] Zhang M, Diao M, Guo L. Convolutional Neural Networks for Automatic Cognitive Radio Waveform Recognition [J]. IEEE Access, 2017, 5(1):11074-11082. [29] R. Girshick, J. Donahue, T. Darrell and J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 1: 580-587. [30] R. Girshick. Fast R-CNN [C]. IEEE International Conference on Computer Vision, 2015, 1:1440-1448. [31] S. Ren, K. He, R. Girshick and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149. [32] T. Lin, P. Dollar, R. Girshick, K. He, et al. Feature Pyramid Networks for Object Detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017, 1:936-944. [33] J. Redmon and A. Farhadi. YOLO9000: Better, Faster, Stronger [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017, 1: [34] W. Liu, D. Anguelov, D. Erhan, et al. SSD: Single Shot Multibox Detector [C]. European Conference on Computer Vision, 2016, 1:21-37. [35] B. Danev and S. Capkun. Transient-based Identification of Wireless Sensor Nodes [C]. International Conference on Information Processing in Sensor Networks, 2009, 1:25-36. [36] C. Szegedy, W. Liu, Y. Jia, et al. Going Deeper with Convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015, 1:1- [37] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016, 1:770-778.

Author Biographies:

ZHOU Xin (1986-), male, from Xiangcheng, Henan, Ph.D., senior engineer and master's supervisor at Institute of Software, Chinese Academy of Sciences. Main research interests include cognitive wireless networks, wireless channel access, and RF machine learning.

HE Xiaoxin (1966-), female, M.S., senior engineer at Institute of Software, Chinese Academy of Sciences. Main research interests include cognitive radio and information systems.

ZHENG Changwen (1969-), male, Ph.D., researcher and doctoral supervisor at Institute of Software, Chinese Academy of Sciences. Main research interests include intelligent information processing and space system simulation.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.