

Article

Deep Muti-Modal Generic Representation Auxiliary Learning Networks for End-to-End Radar Emitter Classification

Zhigang Zhu , Zhijian Yi , Shiyao Li and Lin Li * 

School of Electronic Engineering, Xidian University, Xi'an 710071, China

* Correspondence: lilin@xidian.edu.cn

Abstract: Radar data mining is the key module for signal analysis, where patterns hidden inside of signals are gradually available in the learning process and its superiority is significant for enhancing the security of the radar emitter classification (REC) system. Owing to the disadvantage that radio frequency fingerprinting (RFF) caused by the imperfection of emitter's hardware is difficult to forge, current deep-learning REC methods based on deep-learning techniques, e.g., convolutional neural network (CNN) and long short term memory (LSTM) are difficult to capture the stable RFF features. In this paper, an online and non-cooperative multi-modal generic representation auxiliary learning REC model, namely muti-modal generic representation auxiliary learning networks (MGRALN), is put forward. Multi-modal means that multi-domain transformations are unified to a generic representation. After this, the representation is employed to facilitate mining the implicit information inside of the signals and to perform the better model robustness, which is achieved by using the available generic generation to guide the network training and learning. Online means the learning process of REC is only once and the REC is end-to-end. Non-cooperative denotes no demodulation techniques are used before the REC task. Experimental results on the measured civil aviation radar data demonstrate that the proposed method enables one to achieve superior performance.

Keywords: signal classification; convolutional neural network (CNN); radar emitter classification (REC); signal processing



Citation: Zhu, Z.; Yi, Z.; Li, S.; Li, L. Deep Muti-Modal Generic Representation Auxiliary Learning Networks for End-to-End Radar Emitter Classification. *Aerospace* **2022**, *9*, 732. <https://doi.org/10.3390/aerospace9110732>

Academic Editor: Yan (Rockee) Zhang

Received: 28 September 2022

Accepted: 15 November 2022

Published: 20 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Radar emitter classification (REC), also referred as specific emitter identification (SEI), is the process of extracting radio frequency fingerprints (RFF) of device authentication, and recognizing emitter individuals based on emitter-specific RFFs, which has not only become increasingly important with some military fields, e.g., air reconnaissance, battlefield surveillance, guidance and command operations, but with a broad application prospect in the new hi-tech sector of cognitive radio, self-organized networking, air navigation and traffic control, et cetera [1–3]. With the rapid development of radar technology, radar signals have the characteristics of large quantities, increasing types and density, resulting in the complicated electromagnetic space (noises, multi-path, interferences, etc.). How to mine stable RFF representation involved inside the radar signals has become the key factor to enhance the security of electronic warfare systems. As shown in Figure 1, the REC system mainly consists of several subsystems, i.e., RF system, data collection, data preprocessing, feature extraction, emitters identification, radar types management system and database. Essentially, RFF extraction is a signal recognition task [4–9], which is a cutting-edge scientific problem for signal analysis. RFF representation is constructed by calculating the numerical characters from the observations;. The main pipeline is to extract unique RFF device calibration, and to recognize the radar emitters' individuals [10].

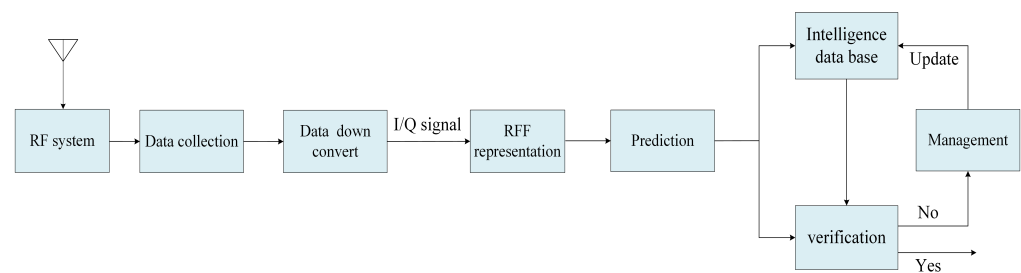


Figure 1. The architecture of a typical REC system, where radar signals are obtained by RF system, sampled by data collection and are processed by the backend RF representation as well as classification, which provides intelligence support and online update for database.

To ensure that the presentation is comprehensive, this paper makes a detailed introduction of signal recognition methods based on the practical applications of the REC and automatic modulation classification. One type of signal representation is the likelihood-based (LB) methods [11]. By comparing the likelihood functions, the LB methods employ the statistical features and judgment threshold to achieve the signal recognition, which has perfect theoretical basis, but has higher requirements on modeling skills and prior knowledge [12,13]. Thus, feature-based methods have become the mainstream practice. The commonly-used features mainly include the instantaneous phase [14]; or time-frequency transform, e.g. wavelet [15], ambiguity function representative-slice [16], three-dimensional distribution [17], compressed sensing mask feature [18], bispectrum [19], bio-inspired algorithm [20], short-time Fourier transform (STFT) [21], signal constellation diagrams [22,23], high-order cumulants [24–27], etc. Meanwhile, the backend classifiers mainly include: K-nearest neighbors (KNN) [26,28,29], binary tree classifier [30], random forest [31], Gaussian naive Bayes [32], extreme learning machine (ELM) [33] and support vector machines (SVM) [34], CNN [11,16,35], LSTM [5], and so forth.

To the best of our knowledge, current RF representation and relevant recognition approaches have moved from manually-designed features to data-driven features; from temporal information mining to transform-domain encoding; from hand-crafted classifiers to automatic deep-learning models; and from multi-step processing pipeline to the end-to-end preprocessing. It has been demonstrated that convolutional neural network (CNN) [36] and long short-term memory (LSTM) [37,38], etc., have been among the most effective data-driven techniques to recognize individual emitters. Observe that the above REC methods have achieved superior performance, and usually a long-time duration of signal observation is expected to extract stable features. Whereas, the REC system often suffers from short data and data-hungry problems. In this case, it is also very difficult to construct a compact RFF representation.

Aiming at the challenges above, an end-to-end REC method based on multi-modal features, namely multi-modal generic representation auxiliary learning networks (MGRALN), is proposed. The multi-modal refers to the different time or frequency transformations. We also notice that the same classifier based on different transformation may obtain different results in that multi-modal features have various physical meanings—i.e., feature heterogeneity. Thus, the signals and corresponding transformations are mapped to mutual subspace for the generic RFF representation (auxiliary branch), which is, meanwhile, utilized to assist the deep-learning architecture to mine the RFF representation from the signals (main branch). Finally, with the auxiliary branch guiding the learning of the main branch, multi-modal information is unified as the generic representation, and is incorporated into the main branch for the REC. Finally, when removing the auxiliary branch from the MGRALN, the stable RFF can be available from the main branch, and achieve a signal-to-prediction REC task. The MGRALN is an online end-to-end REC method and has a strong feature extraction ability of benefitting from multi-modal features. Additionally, it is worth noting that this work not only has important academic value and broad application

prospects, but has an important practical significance for improving the electromagnetic spectrum perception ability.

The reminder of this paper is organized as follows. Section 2 makes a detailed description of the signal model and proposed method, which is followed by the numerical results in Section 3. Conclusions are given in Section 4.

2. Methodology

2.1. Signal Model

Consider the signal in a digital communication system,

$$s[k] = s_k e^{j\theta_k + 2\pi f_0 t}, \quad (1)$$

where $s[k]$ is the k -th symbol signal representing the amplitude s_k and the phase θ_k . Thus, the transmitted signal $s_b(t)$ can be expressed as,

$$s_b(t) = \sum_{k=1}^M s[k] q(t - kT_s), \quad (2)$$

where

$$q(t) = \begin{cases} 1, & \text{if } 0 \leq t \leq T_s \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

For RF system, we have

$$r(t) = \Re\{s_b(t)e^{j2\pi f_0 t}\} + n(t), \quad (4)$$

where $r(t)$ represents the received or intercepted signal and $n(t)$ is noise. Generally, using two orthogonal carriers, the complex baseband signal $r_c(t)$ is shown in Figure 2.

$$r_c(t) = r(t)\cos(2\pi f_0 t + \theta) + r(t)\sin(2\pi f_0 t + \theta), \quad (5)$$

The intercepted signal in a communication system is usually converted to an orthogonal double-channel zero intermediate-frequency, also referred to as a complex baseband I/Q signal, through the digital down conversion (DDC). Because the REC is usually under the non-cooperation circumstance, the classification results may be unsatisfied if some demodulation techniques are used to demodulate the intercepted signal. Additionally, the modulation errors may be accumulated in subsequent REC. Thus, this paper does not adopt any demodulation technique to obtain the modulation types of the intercepted signal.

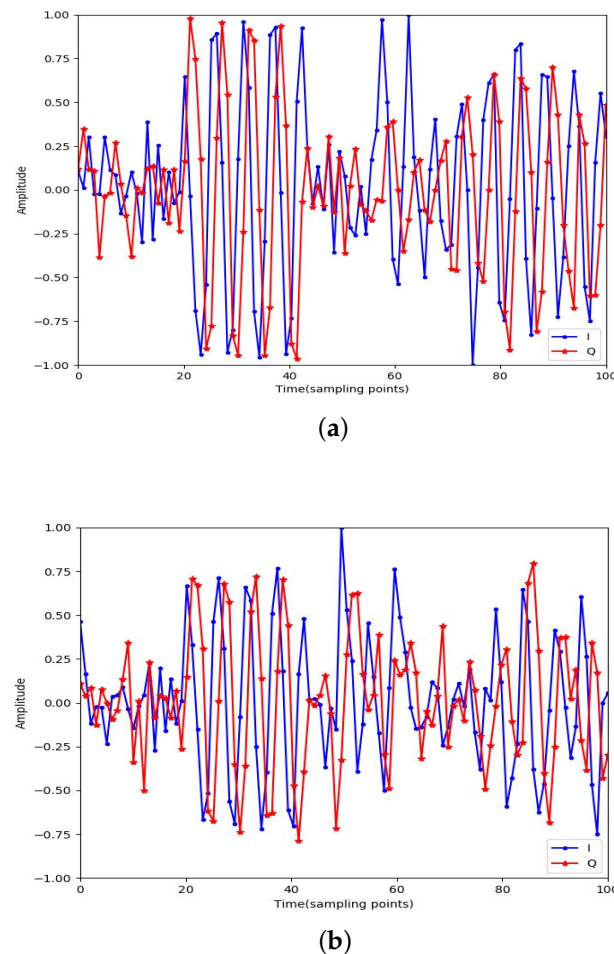


Figure 2. The signal waveform diagram. (a,b) represent the I/Q signals. The number of sampling points are 100, and the amplitude is scaled between 0 and 1. I component is depicted by a blue solid line with ball mark and Q component is represented by a red solid pentacle.

2.2. MGRALN-Based REC

(1) Multi-modal Transformations Selection: Multi-modal transformations selection: The first step for the MGRALN-based REC method is to select the multi-modal transformation, in that the signal transformation determines the distinction of RFF. In principle, the desired transform layer should satisfy the following properties: The inputs of auxiliary branches require high discrimination; all the transformations should be online with well adaptivity; training parameters are not updated layer-wise, but simultaneously. This paper focuses on the guiding model to construct the mutual subspace and to learn the robust RFF representation. Here, we select existing features, as follows:

Signal envelope (SE) presents different transient information, such as the change of signal edge, pulse width, peak position, rising edge and falling edge of signals, etc.

Ambiguity function (AF) is mainly used to measure the distinguishability of the target in distance and velocity dimensions. The ambiguity function decreases rapidly along the axis of the frequency offset. We select several frequency offset slices of ambiguity function to act as transform layers. The slices AF0, AF2 and AF4 represent the features when the frequency offset is set to 0 Hz, 2 Hz, and 4 Hz, respectively.

Power spectrum density (PSD) represents the change of signal power with the frequency, which is defined as the Fourier transform of the autocorrelation function of radar signals.

(2) The MGRALN Model: The MGRALN for REC model is shown in Figure 3. A hierarchy of convolutions is utilized to unify their distribution; thus, the RFF representation benefits from data and transform pipelines. After this, the radar signal and RFF features are aligned by a pixelwise convolution, which are together input to CNN architecture to recognize radar emitter individuals. Given complex baseband signals set $(X, Y) = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i = [I_i, Q_i]$, y_i represents the groundtruth emitter. $X \in \mathbb{C}^{N \times 2 \times L}$, $Y \in \mathbb{R}^{N \times 1}$, N denotes the number of signals, and L represents the sampling length. The convolutional feature of the main branch (i.e., signal pipeline) can be expressed as,

$$M_0 = \mathcal{F}_1(\mathbf{W}_0; x_i) \quad (6)$$

where $\mathcal{F}_1$ denotes a stack of the convolution-BN-swish assemble units, and $\mathbf{W}_0$ denotes learnable parameters.

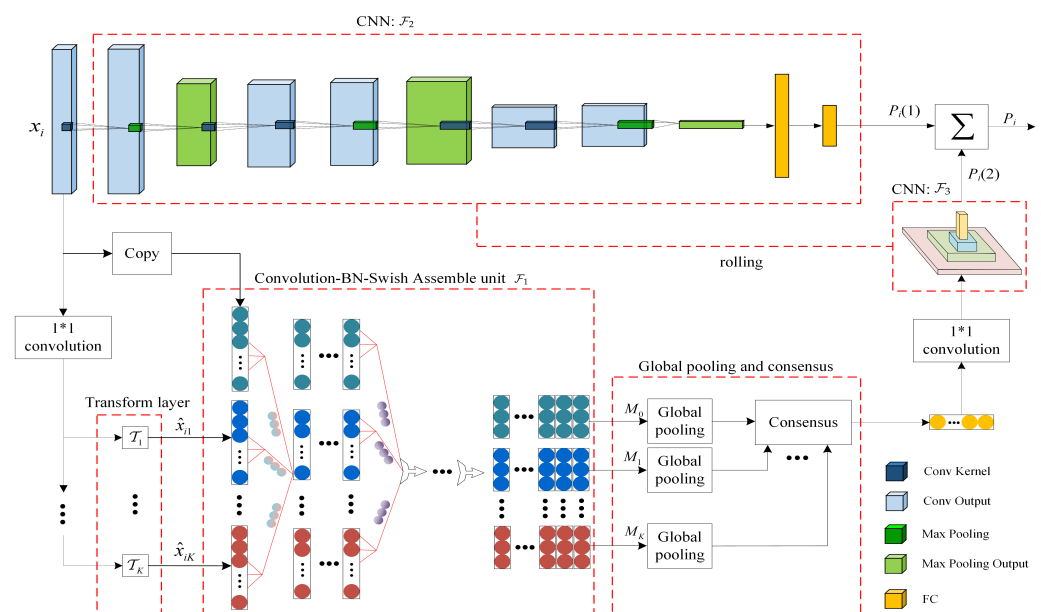


Figure 3. The MGRALN model for REC. Radar signal is supposed to transform layers, convolution-BN-Swish assembly, and temporal global pooling and consensus unit. After this, with learnable parameters shared between signal-based CNN and consensual response-based CNN, the final prediction is available by fusing $P_i(1)$ and $P_i(2)$.

The signal transformation follows similar processing, thus,

$$M_k = \mathcal{F}_1(\mathbf{W}_k; \hat{x}_{ik}) \quad (7)$$

and

$$\hat{x}_{ik} = \mathcal{T}_k(\text{conv}(\mathbf{W}_x; x_i)) \quad (8)$$

where k represents the branch index, $k \in [1, K]$, K is the total number of auxiliary branches, $\mathbf{W}_k$ denotes the learnable parameters, $\mathcal{T}$ denotes the transform layer, and conv is a convolution layer.

To avoid enormous parameters, the parameters are shared among all the paths,

$$\mathcal{F}_1(\mathbf{W}) = \mathcal{F}_1^0(\mathbf{W}_0) = \mathcal{F}_1^1(\mathbf{W}_1) = \dots = \mathcal{F}_1^K(\mathbf{W}_K) \quad (9)$$

After randomly initializing the weights, the signals and multi-modal paths are activated. Then, the convolutional responses are concatenated, and the concatenated responses are supplied to the consensual function as well as a softmax layer. Denote $\mathcal{M}$ as the response of consensual function, and we construct three types, as follows,

Element-wise Average (EA):

$$\mathcal{M} = \frac{M_0 + M_1 + \dots + M_K}{K} \quad (10)$$

Element-wise Multiplication (EM):

$$\mathcal{M} = (M_0 \odot M_1 \odot \dots \odot M_K)^{\frac{1}{K+1}} \quad (11)$$

where $\odot$ represents element-wise multiplication.

Element-wise Concatenation (EC):

$$\mathcal{M} = [M_0, M_1, \dots, M_K] \quad (12)$$

where $[M_0, M_1, \dots, M_K]$ represents element-wise concatenation by channels.

Next, the obtained consensual responses and radar signals are together input to a CNN architecture for the preliminary prediction,

$$P_i(1) = \mathcal{H}(\mathcal{F}_2(\mathbf{W}'; x_i)) \quad (13)$$

$$P_i(2) = \mathcal{H}(\mathcal{F}_3(\mathbf{W}''; \mathcal{T}_{\mathcal{M}})) \quad (14)$$

where

$$\mathcal{T}_{\mathcal{M}} = \text{conv}(\mathbf{W}_{\mathcal{M}}, \mathcal{M}) \quad (15)$$

and

$$\mathcal{H} = \frac{e^l}{\sum_{o=0}^{O-1} e^o} \quad (16)$$

where O is the number of radar emitters, $\mathcal{F}_2$ and $\mathcal{F}_3$ are the CNNs, $\mathcal{H}$ is a softmax function, and W represents its learnable parameters, satisfying $W = W''$ —i.e., $\mathcal{F}_2 = \mathcal{F}_3$.

Finally, the emitter is predicted by fusing the available preliminary prediction above, i.e.,

$$P_i = P_i(1) + P_i(2) \quad (17)$$

(3) The learning principle: To achieve an effective REC task, a predefined loss for the MGRALN model is expected to realize the parameters' update. For the REC, the deep learning networks in the MGRALN expect to minimize a cross-entropy loss,

$$\mathcal{L}_R = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{o=0}^{O-1} y_{io} \log P_{io} \quad (18)$$

In order to capture generic RFF, an additional learning principle is defined for the consistency. For a radar signal, the consensual loss is given by

$$\mathcal{L}_c(x_i) = \sum_{r=1}^K \frac{\langle M_0, M_r \rangle}{\|M_0\| \|M_r\|} + \sum_{v=1}^K \sum_{q=1}^K \frac{\langle M_v, M_q \rangle}{\|M_v\| \|M_q\|} \quad (19)$$

where $\langle \rangle$ denotes the inner product operator, $\| \|$ denotes the modulus operator, and $r, v, q \in \{1, 2, \dots, K\}$. In Equation (19), the first term on the left side keeps the distribution consistency between the signal and each transformation, while the second term guarantees a high correlation between the transformations.

Considering the consensual loss evaluated on all the training data, we have

$$\mathcal{L}_c = \sum_{r=1}^N \mathcal{L}_c(x_i) \quad (20)$$

The final objective function is given by,

$$\mathcal{L} = U(\mathcal{L}_R, \mathcal{L}_c) \quad (21)$$

where U represents the associated learning function.

$$\begin{aligned} \min_{\mathbf{W}_x, \mathbf{W}, \mathbf{W}', \mathbf{W}'', \mathbf{W}_{\mathcal{M}}} \mathcal{L} = & \min_{\mathbf{W}_x, \mathbf{W}, \mathbf{W}', \mathbf{W}'', \mathbf{W}_{\mathcal{M}}} \frac{1}{N} \sum_{i=0}^{N-1} \left\{ \lambda \sum_{r=1}^K \frac{\langle M_0, M_r \rangle}{\|M_0\| \|M_r\|} \right. \\ & \left. + \lambda \sum_{v=1}^K \sum_{q=1}^K \frac{\langle M_v, M_q \rangle}{\|M_v\| \|M_q\|} - \sum_{o=0}^{O-1} y_{io} \log P_{io} \right\} \end{aligned} \quad (22)$$

where λ is a compromise between the recognition and construction of mutual subspace.

However, the loss function above would confuse the MGRALN architecture, because it not only recognizes radar emitter individuals that require the distinctive features, but constructs RFF representation in the auxiliary pipeline. A suboptimal cross-training method can be proposed by, alternatively, optimizing $\mathcal{L}_R$ and $\mathcal{L}_C$. Actually, with CNN, it is difficult to ascertain whether the network training would be synchronized under the two learning criteria. This paper proposes associated learning, given by Equation (22). Next, take the derivative of the associated loss $\mathcal{L}$ with respect to training parameters $\mathbf{W}_x, \mathbf{W}, \mathbf{W}', \mathbf{W}''$ and $\mathbf{W}_{\mathcal{M}}$,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{W}_x} = & \sum_{i=0}^{N-1} \frac{\partial \mathcal{L}_R}{\partial P_i(2)} \frac{\partial P_i(2)}{\partial \mathcal{T}_{\mathcal{M}}} \frac{\partial \mathcal{T}_{\mathcal{M}}}{\partial \mathcal{M}} \sum_{k=0}^K \frac{\partial \mathcal{M}}{\partial M_k} \frac{\partial M_k}{\partial \mathcal{T}_k} x_{ik} \\ & + \lambda \sum_{k=0}^K \frac{\partial \mathcal{L}_C}{\partial M_k} \frac{\partial M_k}{\partial \mathcal{T}_k} x_{ik} \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{W}} = & \sum_{i=0}^{N-1} \frac{\partial \mathcal{L}_R}{\partial P_i(2)} \frac{\partial P_i(2)}{\partial \mathcal{T}_{\mathcal{M}}} \frac{\partial \mathcal{T}_{\mathcal{M}}}{\partial \mathcal{M}} \sum_{k=0}^K \frac{\partial \mathcal{M}}{\partial M_k} \frac{\partial M_k}{\partial \mathcal{T}_k} x_{ik} \\ & + \lambda \sum_{k=0}^K \frac{\partial \mathcal{L}_C}{\partial M_k} \mathcal{T}_k \end{aligned} \quad (24)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}'} = \sum_{i=0}^{N-1} \frac{\partial \mathcal{L}_R}{\partial P_i(1)} x_{ik} \quad (25)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}''} = \frac{\partial \mathcal{L}_R}{\partial P_i(2)} \mathcal{T}_{\mathcal{M}} \quad (26)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{\mathcal{M}}} = \frac{\partial \mathcal{L}_R}{\partial P_i(2)} \frac{\partial P_i(2)}{\partial \mathcal{T}_{\mathcal{M}}} \mathcal{M} \quad (27)$$

The learnable parameters of MGRALN are updated,

$$\begin{pmatrix} \mathbf{W}_x \\ \mathbf{W} \\ \mathbf{W}' \\ \mathbf{W}'' \\ \mathbf{W}_{\mathcal{M}} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_x \\ \mathbf{W} \\ \mathbf{W}' \\ \mathbf{W}'' \\ \mathbf{W}_{\mathcal{M}} \end{pmatrix} - \eta \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial \mathbf{W}_x} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{W}} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{W}'} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{W}''} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{W}_{\mathcal{M}}} \end{pmatrix} \quad (28)$$

The overall loss of MGRALN is a weighted sum of the cross-entropy loss and consensual loss, and the gradients of $\mathcal{L}$ with respect to training parameters $\mathbf{W}_x, \mathbf{W}, \mathbf{W}', \mathbf{W}''$ and

W_M are updated, simultaneously. After training, if the signal is input to the model only (removing auxiliary branches), learned model parameters still enable one to recognize the radar emitter individuals in an end-to-end manner, as shown in Figure 4.

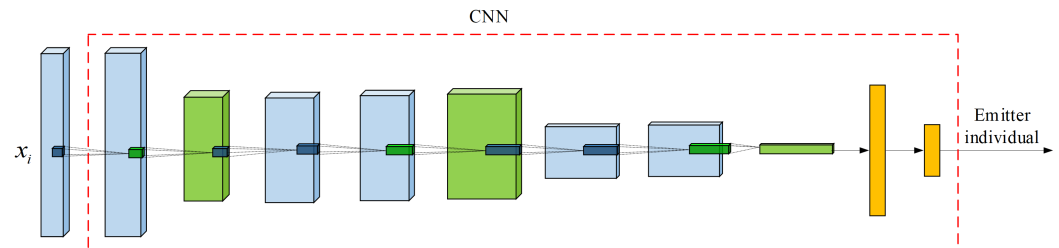


Figure 4. The MGRALN without auxiliary branches.

3. Simulation Results

3.1. Data Collection and Implementation Details

Our data are collected from seven airplanes, and each of them contains 200 files. For each airplane, we sample 100 snapshots, including I and Q components. Five snapshots are collected per second, according to the hardware sampling rate of 12.5 MHz and the signal duty ratio of $0.01/0.2 = 5\%$. The radar scanning cycle is 6 s, and 100 snapshots signals are scanned three times by the main lobe of radars, which results in strong signals. Dataset I consists of 608 samples collected from seven airplanes and dataset II contains 6123 samples collected from 15 airplanes. The civil aviation radar signal of the receiving and processing system is shown in Figure 5. Note that the receiver does not apply any modulation classification technique to obtain the modulation format. We consider the duration of each sequence to be $8 \mu s$ and $46 \mu s$. Meanwhile, we consider the difference of the amounts of samples between classes, equilibrium (Taking dataset I for instance, and seeing radar 1, 2, 3 and 6, or radar 5 and 7) and imbalance (See radar 4 and 5), as shown in Figure 6.

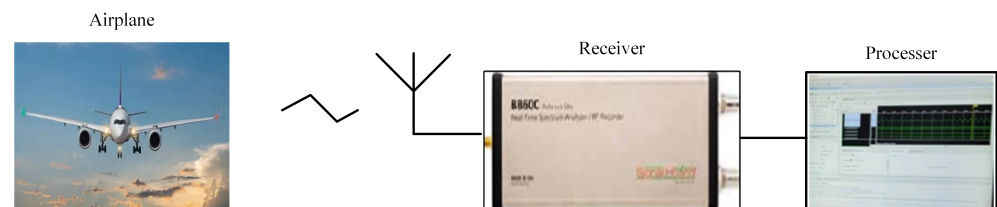


Figure 5. Civil aviation radar signal of receiving and processing system.

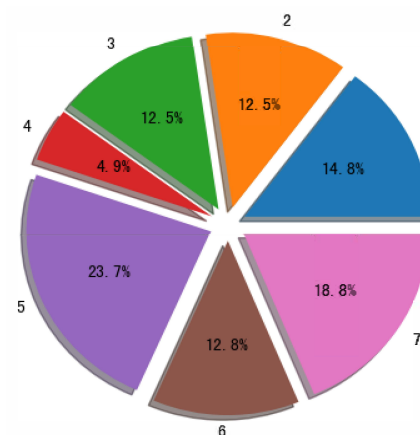


Figure 6. Samples' distribution of different emitters. Equilibrium and disequilibrium sample collected from 7 radars.

For dataset I and dataset II, CNNs follow different design schemes. For dataset I, the CNN consists of three assembly units with 64, 128 and 256 convolutional kernels, respectively. Three convolutional kernels in the Convolution-BN-Swish assembly are of 128, 256, 100, respectively. The number of the last convolutional kernel is 100. The main reason is to guarantee a mutual feature $\mathcal{M}$ in the mutual subspace to have the same dimension with x_i . For dataset II, the CNN contains five similar assembly units with 64, 128, 256, 512 and 1024 convolutional kernels, respectively, and the corresponding convolutional kernels in the Convolution-BN-Swish assembly are 256, 512 and 500, respectively. To increase the receptive fields in the temporal dimension, the kernel size of the first convolutional layer in CNN is 1, and the stride size, meanwhile, is adjusted to 2. The spatial size of the pooling layers is 2. For simplification, the tradeoff parameter λ in the associated loss is set to 1, and the dropout ratio is set to 0.5. In the training stage, MGRALN is trained by a Geforce RTX 3090 Ti GPU. Meanwhile, the learning parameters are optimized using the adaptive moment estimation with a learning rate of 0.0001 [39]. Because the radar signals are unidimensional, the mandatory relation of the relative spatial position might be unduly captured with 2D or 3D convolution operators. Due to lack of spatial correlation in random and unpredictable radar signals, the unidimensional convolution offers obvious advantages [40]. Since a pooling operator tends to make spatial sizes of the convolutional feature half that of the input, the number of convolutional kernels would be doubled to enhance the representational ability of the model [41]. We thus conform to the practice, and the convolutional responses are arranged to form a large-scale majority of channels. We also improve the global averaged pooling to summarize the temporal convolutional responses, such as increasing the temporal receptive field.

3.2. Recognition Performance of MGRALN

(1) Consensus comparisons: signal and single transformation.

The consensual function emerges as a key factor to the MGRALN, due to the vital influence on recognition performance; actually, its construction has been nimbler and more challenging, which is a very fascinating, relevant and important module. In this section, we give a detailed ablation study on the contribution of the consensual function $\mathcal{M}$ to its recognition performance. The form of consensual function is an open problem, and here we compare the correct accuracy of the MGRALN with the SE transformation on the dataset I under a different consensus including (i) EA, (ii) EM, and (iii) EC. As shown in Figure 7, it can be noted that the MGRALN with EA consensus in terms of accuracy achieves a superior performance, and outperforms those of MGRALN with EM and MGRALN with EC. In the traditional machine learning REC task, SE seems to not be a comprehensive representation of RFF.

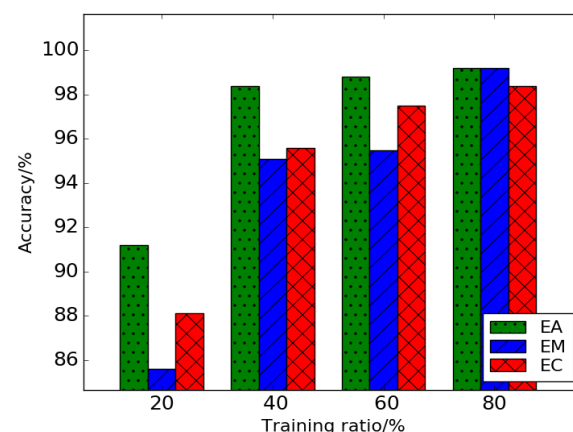


Figure 7. The accuracy performance of various consensual functions, including EA, EM and EC-based MGRALN, respectively.

It can be found that MGRALN still enables one to recognize the radar emitter individuals when using the generic RFF representation generated by the signal and SE transformation to act as the auxiliary pipeline. In the following experiments, we select the EA as the default consensual function.

(2) Accuracy comparisons of MGRALN when $K = 2$.

This section compares the MGRALN with the signal and single transformation, that is, the generic RFF representation, which is utilized to assist the training of the signal pipeline and is available by fusing the signal and one transformation. Specifically, we choose only one signal and one transformation, i.e., SE, AF or PSD, as the input of the auxiliary branch. Figure 8 shows the histograms of the recognition accuracy of MGRALN with traditional SVM methods and the deep-learning based CNN method on dataset I. Observe that less training data will lead to network over-fitting, thereby resulting in inferior recognition performance. With the increase in the training ratio, the MGRALN gradually fits the data distribution, and recognition accuracy is improved gradually. Compared with the feature-based SVM and CNN methods, MGRALNs improve the recognition performance.

It can be also observed that PSD_SVM or PSD_CNN performs better than PSD_MGRALN when $K = 2$. The main reason can be ascribed to two factors. On the one hand, for AF0_MGRALN and SE_MGRALN, the generic representation is achieved by the signal and SE or signal and AF0. The former and the latter are temporal operators. Whereas, the generic representation of PSD_MGRALN is constructed by the signal and PSD feature by unifying time and frequency features, both of which essentially are with a large distribution gap.

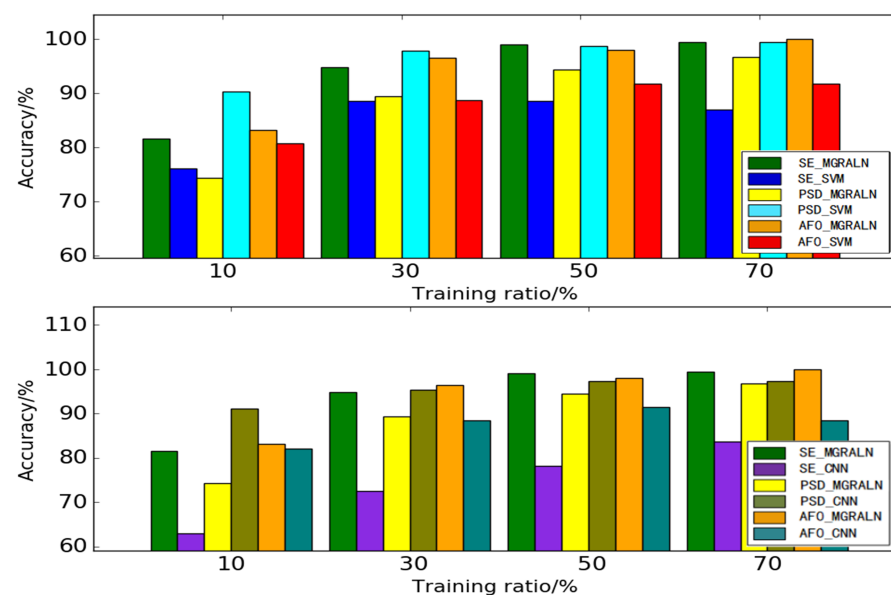


Figure 8. The accuracy performance of MGRALN when $K = 2$ with SVM and CNN classifiers.

(3) Accuracy comparisons of MGRALN when $K = 3$.

As shown in Figure 9a,b, the training ratio of 30% is the critical point that determines where the data-driven MGRALN method outstrips the traditional SVM method. It is worth noting that the performance improvement is large, especially when the AF0 acts as an auxiliary branch. It can be ascribed that the transform has high overlaps with the radar emitter signal. Compared with the PSD and SE, the AF0 seems not to be a high-quality distinctive transform. However, when it is utilized to act as the transform layer, the performance of the MGRALN, instead, matches that of which is based on the PSD. We conjecture that AF has a large overlap with the radar emitter signal in the mutual subspace. Since SE seems to be a relatively weak representation of the specific radar emitter, and we compare the recognition performance of the SE_MGRALN with the support vector machine (SE_SVM) and CNN (SE_CNN). We can observe that SE_MGRALN still makes a

better recognition performance. In our work, the CNN is actually one of the effective tools we can use to break the distribution isolation. That is, any data-driven technique can act as an available tool to achieve this goal.

MGRALN differs from many existing REC approaches in that it expects explicit insight into the number of auxiliary branches, and what kinds of combinations of auxiliary branches can achieve higher-quality RFF. This section introduces dual transform layers, and demonstrates their effect on measured radar emitter data. Considering the comprehensiveness, dual auxiliary branches are selected by random sampling with replacement, i.e., both items are randomly selected from PSD, AF0 and SE. In order to eliminate the influence of network over-fitting, we set the training ratio to 80%, and explore the recognition performance of the MGRALN when $K = 3$, i.e., signal and dual transformations. Table 1 shows the recognition results on the dataset I. The diagonal entries denote that a signal and two transformations is of the same type as the inputs, in essence, which is sort of boosting in the ensemble learning. Despite this encouraging progress, the MGRALNs when $K = 3$ are, still, marginally inferior to those when $K = 2$, which can be ascribed to two aspects. One possible reason is that the form of $\mathcal{L}_c$ is an open problem, directly affecting the quality of the auxiliary RFF representation. Additionally, under the two learning rules, $\mathcal{L}_R$ and $\mathcal{L}_c$, it, however, is a very challenging thing to guarantee the synchronization of model learning under the two learning principle.

Taken together, the cross-entropy loss and the consensual loss have a perceptible effect on the recognition performance of REC, because the former is the foundation of the latter, and, both $\mathcal{L}_R$ and $\mathcal{L}_c$ determine the final recognition performance. This paper, in its current form, does not fall into the trap of dedicating too much on the learning principles, but focuses its best efforts on how to construct the stable RFF formulation.

(4) Accuracy comparisons of MGRALN without auxiliary pipelines.

It can be observed that MGRALN has the ability to project the signal and its transformations to the mutual subspace. Naturally, the proposed MGRALNs with the auxiliary pipeline have been demonstrated to yield a competitive performance and use transform layers with a back-propagate ability to embed the transformations inside the learning of the MGRALNs. Although such a practice can avoid extracting the RFF features offline, there is still a lack of an intuitionistic demonstration of the effects of auxiliary pipelines. In view of this, for a pure signal-prediction module, the recognition accuracy of the MGRALN (maximum accuracy) is shown in Table 2. Observe that the MGRALN removing the AF0 pipeline, achieves a superior recognition accuracy on dataset I. For dataset II, the MGRALN removing the SE pipeline is superior to that of other forms.

Table 3 illustrates the accuracy comparisons of the end-to-end MGRALN with the mainstream methods including SVM, LSTM and a combination of CNN and LSTM (CNN-LSTM). On the dataset I and dataset II, the recognition accuracy of the MGRALN obviously outperforms the SVM recognition method by a large margin of 22.2% and 3.6%, respectively. Compared with deep learning radar emitter recognition method inputting signals to LSTM, MGRALN exceeds the accuracy improvement of 37.7% and 12.3%. Further, considering the advantage of CNN and LSTM, we compare the MGRALN with the CNN concentrated with a LSTM, and the MGRALN makes a 5.3% and 1.4% accuracy boost.

Compared with CNN, the computation complexity of the proposed MGRALN focuses mainly on the Convolution-BN-Swish assembly structure. For deep-learning architectures, floating point operations (FLOPs) are utilized to calculate the complexity of deep learning architectures. For the convolutional layer, the FLOPs are $2 \times C_i \times K_i \times C_o \times L_o$, where C_i is the number of channels in the input, K_i is the kernel size, C_o is the number of channels in the output, and L_o is the feature size of the output. Specifically, the total FLOPs difference for dataset I: $O(3.43 \times 10^6)$, and $O(8.74 \times 10^7)$. For the MGRALN without auxiliary branches, the computation complexity of MGRALN is the same with CNN in the test. Compared with MGRALN and CNN, the SVM has a lower computational complexity.

Table 1. Accuracy comparisons of the MGRALN when $K = 3$ on the dataset I.

Accuracy \ T	T	AF0	PSD	SE
	AF0	98.4%	98.4%	95.9%
	PSD	98.4%	99.2%	95.1%
	SE	95.9%	95.1%	96.7%

Table 2. Accuracy comparisons of MGRALN without auxiliary pipelines on the dataset I and dataset II when the training ratio is 80%.

Accuracy \ T	T	AF0	PSD	SE	max
	Dataset I	99.2%	93.4%	98.4%	99.2%
	Dataset II	88.5%	88.7%	97.9%	97.9%

Table 3. Accuracy comparisons of MGRALN with the mainstream methods on the dataset I and dataset II when the training ratio is 80%.

Accuracy \ Method	Method	MGRALN	SVM	LSTM	CNN-LSTM
	Dataset I	99.2%	77.0%	61.5%	93.9%
	Dataset II	97.9%	93.9%	85.6%	96.5%

The proposed MGRALN can achieve a superior radar emitter recognition by bringing the weak signal-to-recognition methods to high-level end-to-end prediction. From a scientific standpoint, the MGRALN attempts to demonstrate the validity of the distribution unification. From the perspective of engineering applications, the MGRALN makes an end-to-end REC in the complex electromagnetic environment. From the perspective of the model, the MGRALN is a new model for the online noncorporate REC, which can be treated as the alternative scheme; or a more complementary approach with the mainstream approach, e.g., multi-path REC classifiers.

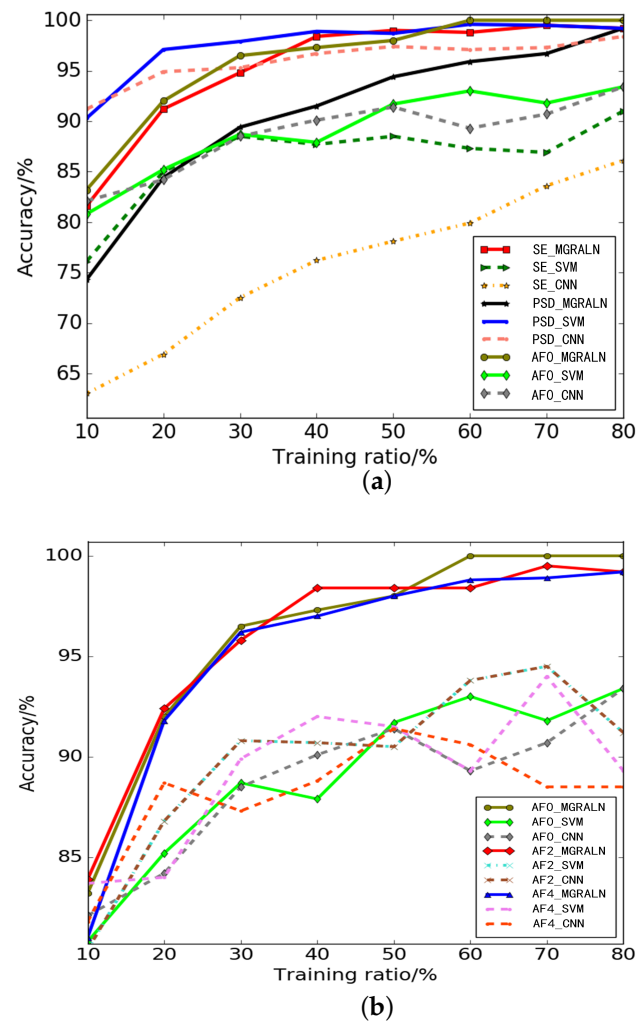


Figure 9. Accuracy comparisons of MGRALN with SVM and CNN when training ratio varies from 10% to 80%. (a) Represents recognition results of MGRALN when transformation is SE, PSD or AF0, and (b) represents recognition results of MGRALN with different AF slices.

4. Conclusions and Future Work

This paper aims to solve the radar emitter classification (REC) problem because the unstable RFF inclines to forge, in order to be easily covered as well as to be damaged by noise and radar signals. This motivates us to attempt to construct a more stable RFF representation by fusing multi-modal features to facilitate deep-learning model learning, which can be embedded inside of the process to further mine the radar signal. The resultant model is the MGRALN.

Compared with SVM, CNN, and LSTM, etc., the advantage of the MGRALN is its superior recognition performance on two measured civil aviation radar datum. Additionally, our MGRALN is an online end-to-end prediction architecture in the REC task, especially in the cases where only radar signals are available, and no demodulation techniques are used.

Despite its validity in the civil airplane scenario, the proposed MGRALN is restricted by the design of the auxiliary branches, in that their construction is an open problem, which determines the generic representation. Additionally, the proposed MGRALN cannot completely replace the current existing REC methods, in that it is just a feasible path. From the perspective of signal mining, it provides sparse yet distinctive signal representation for signal analysis, which may be applied in the other signal recognition tasks, e.g., automatic modulation classification. Directions for future works include the construction of learning principle, more insight into the internal operation and interpretability of RFF mechanism,

boosting [42], and multi-path features or classifiers' fusion, refs. [43,44] including the complementation with CNN, LSTM, and BiLSTM, etc. Additionally, some attempts may focus on deeper model construction and attention mechanisms, including the transformer [45] based REC.

Author Contributions: Conceptualization, Z.Z. and L.L.; methodology, Z.Z. and Z.Y.; formal analysis, L.L.; Funding acquisition, Z.Z. and L.L.; Investigation, S.L.; Supervision, L.L.; Validation, Z.Y. and S.L.; Visualization, Z.Z. and S.L.; Writing—original draft, Z.Z.; Writing—review and editing, Z.Z. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the National Natural Science Foundation of China under Grants # 62203343, # 62071349 and # U21A20455.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to express their appreciation to the editors for their rigorous and efficient work and the reviewers for their helpful suggestions, which greatly improved the presentation of this paper.

Conflicts of Interest: The authors declare that they have no conflict of interest to report regarding the present study.

References

1. Dobre, O. Signal identification for emerging intelligent radios: Classical problems and new challenges. *IEEE Instrum. Meas. Mag.* **2015**, *18*, 11–18. [\[CrossRef\]](#)
2. Kim, L.S.; Bae, H.B.; Kil, R.M.; Jo, C.H. Classification of the trained and untrained emitter types based on class probability output networks. *Neurocomputing* **2017**, *248*, 67–75. [\[CrossRef\]](#)
3. Li, X.; Huang, Z.; Wang, F.; Wang, X.; Liu, T. Toward Convolutional Neural Networks on Pulse Repetition Interval Modulation Recognition. *IEEE Commun. Lett.* **2018**, *22*, 2286–2289. [\[CrossRef\]](#)
4. Sun, J.; Shi, W.; Yang, Z.; Yang, J.; Gui, G. Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems. *IEEE Trans. Veh. Technol.* **2019**, *68*, 10348–10356. [\[CrossRef\]](#)
5. Polak, A.C.; Dolatshahi, S.; Goekel, D.L. Identifying wireless users via transmitter imperfections. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1469–1479. [\[CrossRef\]](#)
6. Reising, D.R.; Temple, M.A.; Jackson, J.A. Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1180–1192. [\[CrossRef\]](#)
7. Qi, P.; Zhou, X.; Zheng, S.; Li, Z. Automatic modulation classification based on deep residual networks with multimodal information. *IEEE Trans. Cogn. Commun. Netw.* **2021**, *7*, 21–33. [\[CrossRef\]](#)
8. Ramkumar, B. Automatic modulation classification for cognitive radios using cyclic feature detection. *IEEE Circuits Syst. Mag.* **2009**, *9*, 27–45. [\[CrossRef\]](#)
9. Cao, R.; Cao, J.; Mei, J.; Yin, C.; Huang, X. Radar emitter identification with bispectrum and hierarchical extreme learning machine. *Multimed. Tools Appl.* **2019**, *78*, 28953–28970. [\[CrossRef\]](#)
10. Xu, Q.; Zheng, R.; Saad, W.; Han, Z. Device fingerprinting in wireless networks: challenges and opportunities. *IEEE Commun. Surv. Tutorials* **2016**, *18*, 94–104. [\[CrossRef\]](#)
11. Zheng, J.; Lv, Y. Likelihood-based automatic modulation classification in ofdm with index modulation. *IEEE Trans. Veh. Technol.* **2018**, *67*, 8192–8204. [\[CrossRef\]](#)
12. Wei, W.; Mendel, J.M. Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Trans. Commun.* **2000**, *48*, 189–193. [\[CrossRef\]](#)
13. Hameed, F.; Dobre, O.A.; Popescu, D. On the likelihood-based approach to modulation classification. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 5884–5892. [\[CrossRef\]](#)
14. Al-Sa'd, M.; Boashash, B.; Gabbouj, M. Design of an optimal piece-wise spline wigner-ville distribution for TFD performance evaluation and comparison. *IEEE Trans. Signal Process.* **2021**, *69*, 3963–3976. [\[CrossRef\]](#)
15. Zhang, M.; Diao, M.; Guo, L. Convolutional neural networks for automatic cognitive radio waveform recognition. *IEEE Access* **2017**, *5*, 11074–11082. [\[CrossRef\]](#)
16. Zhang, Z.; Wang, C.; Gan, C.; Sun, S.; Wang, M. Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD. *IEEE Trans. Signal Inf. Process. Over Netw.* **2019**, *5*, 469–478. [\[CrossRef\]](#)

17. Li, B.; Wang, W.; Zhang, X.; Zhang, M. Deep learning based automatic modulation classification exploiting the frequency and spatiotemporal domain of signals. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8.
18. Zhang, Z.; Luo, H.; Wang, C.; Gan, C.; Xiang, Y. Automatic modulation classification using cnn-lstm based dual-stream structure. *IEEE Trans. Veh. Technol.* **2020**, *69*, 13521–13531. [[CrossRef](#)]
19. Graves, A.; Fernández, S.; Schmidhuber, J. Bidirectional LSTM networks for improved phoneme classification and recognition. In Proceedings of the 15th International Conference on Artificial Neural Networks, Bratislava, Slovakia, 15–18 September 2005; Volume II, pp. 799–804.
20. Sepas-Moghaddam, A.; Etemad, A.; Pereira, F.; Correia, P.L. Facial emotion recognition using light field images with deep attention-based bidirectional LSTM. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 3367–3371.
21. Liu, S.; Yan, X.; Li, P.; Hao, X.; Wang, K. Radar emitter recognition based on SIFT position and scale features. *IEEE Trans. Circuits Syst. II Exp. Briefs* **2018**, *65*, 2062–2066. [[CrossRef](#)]
22. Doan, V.-S.; Huynh-The, T.; Hua, C.-H.; Pham, Q.-V.; Kim, D.-S. Learning Constellation Map with Deep CNN for Accurate Modulation Recognition. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6.
23. Kumar, Y.; Sheoran, M.; Jajoo, G.; Yadav, S.K. Automatic Modulation Classification based on Constellation Density using Deep Learning. *IEEE Commun. Lett.* **2020**, *24*, 1275–1278. [[CrossRef](#)]
24. Wang, A.; Li, R. Research on Digital Signal Recognition Based on Higher Order Cumulants. In Proceedings of the 2019 International Conference on Intelligent Transportation, Big Data and Smart City (ICITBS), Changsha, China, 12–13 January 2019; pp. 586–588.
25. Liu, M.; Zhao, Y.; Shi, L.; Dong, J. Research on Recognition Algorithm of Digital Modulation by Higher Order Cumulants. In Proceedings of the 2014 Fourth International Conference on Instrumentation and Measurement, Computer, Communication and Control, Harbin, China, 18–20 September 2014; pp. 686–690.
26. 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.
27. Xie, L.; Wan, Q. Cyclic Feature-Based Modulation Recognition Using Compressive Sensing. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 402–405. [[CrossRef](#)]
28. Javed, Y.; Bhatti, A. Emitter recognition based on modified x-means clustering. In Proceedings of the IEEE Symposium on Emerging Technologies, Islamabad, Pakistan, 18 September 2005; IEEE: Piscataway, NJ, USA, 2005; pp. 352–358.
29. He, B.; Wang, F.; Liu, Y.; Wang, S. Specific emitter identification via multiple distorted receivers. In Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
30. Qin, M.; Li, X.; Yao, Y. An algorithm of digital modulation identification based on instantaneous features. *J. Theor. Appl. Inf. Technol.* **2013**, *50*, 396–400.
31. Triantafyllakis, K.; Surligas, M.; Vardakis, G.; Papadakis, S. Pasma: An automatic modulation classification system based on Random Forest. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; pp. 1–3.
32. Narayan, Y. Comparative analysis of SVM and naive Bayes classifier for the SEMG signal classification. *Mater. Today Proc.* **2020**, *37*, 3241–3245. [[CrossRef](#)]
33. Güner, A.; Alçın, Ö.F.; Şengür, A. Automatic digital modulation classification using extreme learning machine with local binary pattern histogram features. *Measurement* **2019**, *145*, 214–225. [[CrossRef](#)]
34. Wei, Y.; Fang, S.; Wang, X. Automatic modulation classification of digital communication signals using SVM based on hybrid features, cyclostationary, and information entropy. *Entropy* **2019**, *21*, 745. [[CrossRef](#)]
35. Peng, S.; Jiang, H.; Wang, H.; Alwageed, H.; Zhou, Y.; Sebdani, M.M.; Yao, Y. Modulation classification based on signal constellation diagrams and deep learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 718–727. [[CrossRef](#)]
36. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* **1995**, 3361, 1–14.
37. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
38. Schuster M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [[CrossRef](#)]
39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the ICLR, San Diego, CA, USA, 8–9 July 2014; pp. 1–15.
40. Sun, J.; Xu, G.; Ren W.; Yan, Z. Radar emitter classification based on unidimensional convolutional neural network. *IET Radar Sonar Navig.* **2018**, *12*, 862–867. [[CrossRef](#)]
41. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the CVPR, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
42. Zhao, S.; Wang, W.; Zeng, D.; Chen, X.; Zhang, Z.; Xu, F.; Liu, X. A Novel Aggregated Multipath Extreme Gradient Boosting Approach for Radar Emitter Classification. *IEEE Trans. Ind. Electron.* **2022**, *69*, 703–712. [[CrossRef](#)]

-
43. Huang, H.S.; Liu, L.; Tseng, V.S. Multivariate time series early classification using multi-domain deep neural network. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 90–98.
 44. Liu, C.; Hsaio, W.; Tu, Y. Time series classification with multivariate convolutional neural network. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4788–4797. [[CrossRef](#)]
 45. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17, Long Beach, CA, USA, 4–7 December 2017; pp. 5998–6008.