

Article

A Novel Data-Driven Specific Emitter Identification Feature Based on Machine Cognition

Mingzhe Zhu , Zhenpeng Feng * and Xianda Zhou

School of Electronic Engineering, Xidian University, Xi'an 710121, China; zhuzmz@mail.xidian.edu.cn (M.Z.); zhouxian999@gmail.com (X.Z.)

* Correspondence: zpfeng6790@protonmail.com

Received: 17 July 2020; Accepted: 11 August 2020; Published: 14 August 2020



Abstract: Machine learning becomes increasingly promising in specific emitter identification (SEI), particularly in feature extraction and target recognition. Traditional features, such as radio frequency (RF), pulse amplitude (PA), power spectral density (PSD), and etc., usually show limited recognition effects when only a slight difference exists in radar signals. Numerous two-dimensional features on transform domain, like various time-frequency representation and ambiguity function are used to augment information abundance, whereas the unacceptable computational burden usually emerges. To solve this problem, some artfully handcrafted features in transformed domain are proposed, like representative slice of ambiguity function (AF-RS) and compressed sensing mask (CS-MASK), to extract representative information that contributes to machine recognition task. However, most handcrafted features only utilizing neural network as a classifier, few of them focus on mining deep informative features from the perspective of machine cognition. Such feature extraction that is based on human cognition instead of machine cognition may probably miss some seemingly nominal texture information which actually contributes greatly to recognition, or collect too much redundant information. In this paper, a novel data-driven feature extraction is proposed based on machine cognition (MC-Feature) resort to saliency detection. Saliency detection exhibits positive contributions and suppresses irrelevant contributions in a transform domain with the help of a saliency map calculated from the accumulated gradients of each neuron to input data. Finally, positive and irrelevant contributions in the saliency map are merged into a new feature. Numerous experimental results demonstrate that the MC-feature can greatly strengthen the slight intra-class difference in SEI and provides a possibility of interpretation of CNN.

Keywords: feature extraction; machine cognition; machine learning; specific emitter identification; saliency detection

1. Introduction

Feature extraction is an important part in specific emitter identification (SEI), but it is more than challenging because in modern complex electromagnetic environment, it is no longer enough to accomplish SEI tasks only relying on some primitive properties of radar signals, like radio frequency (RF), pulse amplitude (PA), pulse width (PW), power spectral density (PSD), and etc. [1–4]. Various two dimensional (2D) transform features, like short time Fourier transform (STFT), Wavelet transform (WT), S transform (ST), Winger-Ville distribution (WVD), ambiguity function (AF), and etc., are used as feature input to classifier in order to represent more comprehensive information in a feature. Although they can achieve good results, the information representation capacity of a transformed feature and its data dimensionality are always a pair of paradox. Based on this problem, some compressive sensing (CS) extraction methods are proposed: Wang, L. and Ji, H. propose a representative slice of ambiguity function (AF-RS) in [5] which selects the most representative slice in a

certain Doppler-lag. Zhu, M. and Zhang, X. propose compressive sensing mask (CS-MASK) in [6] which is a low-dimensional feature with good representation by constraining the reconstruction error between inverse transformation of the selected ambiguity function and the WVD of original signal. Reference [7] performs synchrosqueezing transform on STFT, obtaining a more concentrated energy band and more accurate instantaneous frequency trajectory which are merged into a new feature matrix. These are all excellent handcrafted features that can achieve satisfying recognition rate in different scenarios. However, there are still imperfections which can still be improved: (1) When AF-RS processing complex modulated signals, only a single slice can rarely reflect all the key information in ambiguity domain. As for CS-MASK, it is considerably complicated to find an appropriate optimization method in order to satisfy the constraint condition for various signals. (2) The SST feature has better identification ability for complex modulated signals in comparison to the former methods, whereas its effect fluctuates when used to process simple modulated signals. In addition, some recognition methods in like acoustic and speech identification, like Gaussian Mixed Model (GMM) may also be migrated to SEI signal recognition [8,9]. However, there still exist some problems: GMM is based on a principle that the datasets should be large enough, unfortunately, it is not easy to realize in SEI, because radar signal acquisition needs higher experimental conditions, such as radar, microwave anechoic chamber, etc. An effective feature is required that can achieve high recognition accuracy, even under small number of datasets.

Various handcrafted features are proposed based on the authors' firm conviction that the information contained in these features caters recognition network. Concretely, in AF-RS, it is believed that the slice located near origin in AF domain may represent most useful information, because of its high amplitude and in CS-MASK, a low reconstruction errors is believed to have a good representation of original AF. In the SST feature, it is believed that more accurate instantaneous frequency and suppressed power band are probably the keys for recognition network to distinguish different types of signals. Note that all of these handcrafted features mentioned above are based on the authors' cognition. Nevertheless, does the recognition network really "looks" at this extracted information when it does a recognition task? Does the extracted information still include redundant or even interference information? Few references analyze these questions from the perspective of machine cognition. Linear relevance propagation (LRP) is used in [7] to create a heat map, endeavoring to demonstrate the validation of SST feature, which is a heuristic spirit, even though the interpretation in [7] is quite simple and conjectural.

Machine learning (ML) has made tremendous success in computer vision during recent years, especially in image enhancement, object detection, scene reconstruction, and etc. In the field of object detection, lots of methods are used to segment the foreground object and background in an image [10–12]. Simonyan, K. and Vedaldi, A proposed an image specific class saliency visualization in [12] where a saliency map is created to reflect how different pixels in each grid impacts the final classification result. Inspired by [7,12], a novel data-driven feature that is based on machine cognition (MC- feature) is proposed in this paper combined with specific class saliency visualisation to "tell" us where the network "looks" at and where the network "disinterest", and then manipulate this information into a feature. MC-feature has extraordinary representative capability of intra-class fingerprint information of signals in SEI. This superiority is due to the understanding of the process of feature extraction in machine rather than the complex manipulation on 2D features, which avoids the difficulty of selecting the optimal parameters in AF-RS and CS-MASK. Additionally, MC-feature can reflect the preference to different class signals, which provides a possibility of interpreting clear physical meaning of MC-feature and innate recognition mechanism of machine.

The paper is arranged, as follows: Section 2 introduces two classical 2D transform domain feature as well as two ingenious handcrafted features as the comparative algorithms. Section 3 elaborates the mathematical rationale and procedures of the proposed algorithm. Numerous experimental results are shown and analyzed in Section 4. Section 5 is the paper conclusion.

2. 2D Transform Domain Feature

With the increasing complexity of signal form and modulation mode, particularly signals with time-varying information, traditional one-dimensional features, such as RF, PA, PW, and PSE, have been difficult to satisfy the recognition requirements. Lots of two-dimensional transform domain features have been proposed to allay the limitation of inadequate representation of one-dimensional features. These features can be divided into primitive 2D transformation features and handcrafted 2D transform features, which are introduced in Section 2.1 and Section 2.2, respectively.

2.1. Primitive Transform Domain Feature

Numerous 2D transforms or representation are proposed, such as STFT, WT, ST, WVD, and AF. These features can capture time-varying information of frequency or amplitude of signals, meaning a larger capacity of representation in comparison to one dimensional features. In this section, STFT and AF will be introduced in details, because the MC-feature and two comparative features are based on STFT and AF, respectively.

2.1.1. Short Time Fourier Transform

STFT is a widely used time-frequency representation tool in signal processing. STFT transforms signals into time-frequency domain by performing Fourier Transform in a fixed window traversing the entire time domain. The STFT of a given signal $x(t)$ with a window $g(t)$ in Schwartz class, is defined as [13]:

$$V_x^g(t, f) = \int_{\mathbb{R}} x(\tau) g^*(\tau - t) e^{-i2\pi f\tau} d\tau, \quad (1)$$

where $g^*(t)$ is the complex conjugate of $g(t)$, τ is time delay, and f is frequency. Compared with one-dimensional features, STFT can reflect the variation of the frequency of each component with time; hence, STFT holds a larger capacity of information representation. Nonetheless, the time window with fixed width limits time-frequency resolution. Subsequently, to assuage the limitation of STFT, numerous time-frequency transform methods with variable time-frequency resolution are proposed, such as WT and ST. Even though the time-frequency resolution of STFT is limited, the concise and easily implemented mathematical form makes STFT one of the most prevalently used time-frequency analysis tools in signal processing. Here, four typical frequency modulation (FM) signals are taken as examples: single-frequency $x_1(t)$, linear frequency modulation (LFM) $x_2(t)$, second-order frequency modulation $x_3(t)$, and triangle frequency modulation $x_4(t)$, to show effects of STFT. The sampling frequency and sampling time is set to 256 Hz and 1 s respectively. The specific forms of signals are as follows:

$$x_1(t) = \sin(2\pi 20t) \quad (2)$$

$$x_2(t) = \sin(2\pi(20t^2 + 20t)) \quad (3)$$

$$x_3(t) = \sin(2\pi(15t^3 + 20t)) \quad (4)$$

$$x_4(t) = \sin(2\pi 10(\cos(2\pi 0.5t))) \quad (5)$$

Figure 1 shows the waveform and STFT of each signal. It can be seen that STFT can clearly reflect the variation of frequency with respect to time, whereas the time-frequency resolution is fixed no matter when the frequency is low or high.

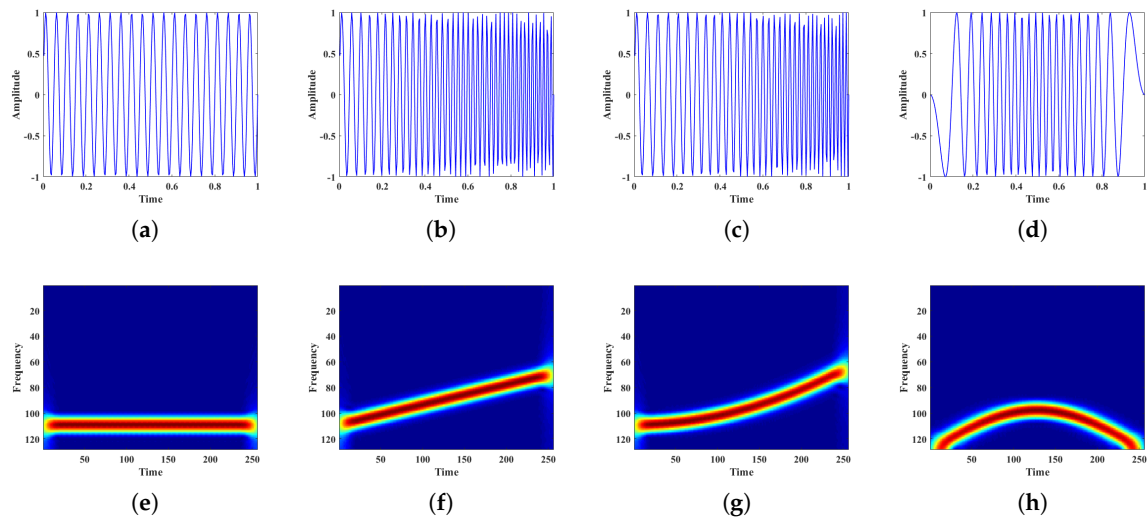


Figure 1. Short time Fourier transform (STFT) of several typical simulation signals. (a–d) are time domain waveform of $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$; (e–h) are corresponding STFT.

2.1.2. Ambiguity Function

AF is the inverse Fourier transform of the instantaneous autocorrelation function of a signal over the time variable. Different from STFT that transforms signals in joint time-frequency domain (TFD), AF transforms signals to time lag-Doppler lag domain [7]. The AF of signal $x(t)$ is defined by:

$$A_x(\tau, \nu) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})e^{-j\nu\tau}dt, \quad (6)$$

where τ is time-lag and ν is Doppler-lag. It is clear from (6) that, when Doppler-lag is 0, the AF degenerates into the integral of autocorrelation function of the signal with respect to time. Therefore, the peak value of AF is located in the origin according to the property of autocorrelation function.

Figure 2 shows the AF of each signal. From Figure 2, the energy is mostly concentrated near the origin in AF domain, and the difference of each signal can be easily detected from sidelobes in disparate shapes.

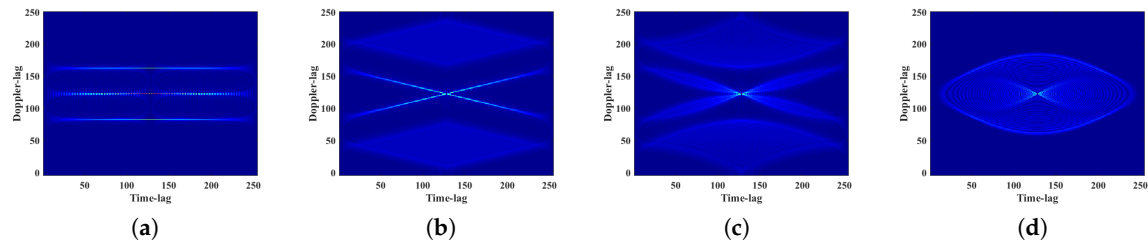


Figure 2. AF of several typical simulation signals. (a–d) are AF of $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$.

2.2. Handcrafted Transform Domain Feature

While the primitive 2D features can represent more informative details of signals, the data quantity of primitive 2D features is usually squared with the data quantity of the processed signal, leading to unacceptable computation burden. With the development of theory of compressed sensing (CS), some algorithms that are based on CS are proposed to alleviate this problem by extracting some most representative parts of primitive 2D features as a new feature with smaller data quantity. However, such compression inevitably jettisons some detailed texture information more or less; hence, we usually attempt to seek a trade off between representative capability and data quantity.

In this section, two excellent handcrafted transform domain features that are based on AF, AF-RS, and CS-MASK, are introduced.

2.2.1. Representative Slice of Ambiguity Function

Because the energy of AF shows a peak value at origin, the slices of AF in a certain Doppler-lag near zero (including zero-slice) could be considered as the major representative feature of radar signals named representative slice.

A class-dependent algorithm to classify radar emitter is proposed by Gillespie and Atlas [14,15], which is used to extract representative features of radar signals, followed by a kernel optimization scheme in the AF domain. Direct Discriminant Ratio (DRR) is used as a criterion to rank the kernel in [14], defined as:

$$DRR(\tau, \nu) = \sum_{i=1}^C \sum_{j=1}^C |\bar{A}^i - \bar{A}^j|^2, \quad (7)$$

here, \bar{A}^i means the "average" auto-ambiguity function of Class i . Frequency offset ν is usually set near zero in order to rank the points along the major direction of ambiguity function distribution. By this norm, AF-RS extracts the slices near the origin in delay-frequency offset domain as a feature set. The most representative slice is selected as the feature, resulting in a great reduction in data quantity. Figure 3 shows the AF-RS of each signal. It is clear that this feature with low data quantity (24×24) can vividly exhibit the difference of each signal in a very visible form.

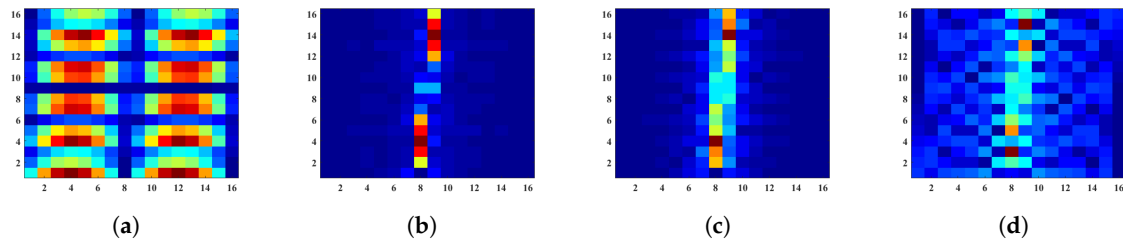


Figure 3. Representative slice of ambiguity function (AF-RS) of several typical simulation signals. (a–d) are AF-RS of $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$.

It is necessary to note that there are still some small flaws that can be improved on AF-RS [7], mainly in: (1) a sole slice at a fixed frequency may neglect other slices which contribute to identification in other Doppler-lag. (2) The selection of representative slice must be implemented by complex feature fusion optimization, which probably makes it unfeasible to achieve real-time optimization. It must be based on recognition rate feedback, which results in large computational complexity.

2.2.2. Compressed Sensing Mask

CS-MASK is an ingenious handcrafted feature extraction method combining CS theory and properties of AF and WVD [6]. CS-MASK seeks for a small-sized region in AF, where most representative information is included by constraining the error between WVD reconstructed from CS-MASK and original WVD.

It can be seen from (6) that the two-dimensional inverse Fourier transform of the Wigner–Ville distribution is the ambiguity function, as shown below:

$$A_x(\tau, \nu) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-j\nu\tau} dt \quad (8)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} WVD_x(t, f) e^{-j(\nu\tau + \tau f)} dt df, \quad (9)$$

where $WVD_x(t, f)$ is Wigner–Ville Distribution (WVD) of signal $x(t)$, a kind of quadratic time-frequency representation tool, defined as:

$$WVD_x(t, f) = \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \quad (10)$$

Set WVD and ambiguity in the form of column vectors in size of $N \times 1$; hence, the matrix form of AF can be written as:

$$\mathbf{A}_{x(N \times 1)} = \mathbf{\Psi}_{(N \times N)} \cdot \mathbf{W}_{D(N \times 1)}, \quad (11)$$

where $\mathbf{\Psi}$ is Fourier transform matrix in size of $N \times N$, and W_D is the matrix form of WVD of $x(t)$. To reconstruct Wigner–Ville distribution from ambiguity function with high accuracy and low data dimension, a measurement matrix Φ in size of $M \times N$ ($M \ll N$) based on CS is used to extract those large non-zero values in \mathbf{A}_x . Feature $\Theta_{(M \times 1)}$, an M sparse measurement of AF, is obtained, as follows:

$$\Theta_{(M \times 1)} = \mathbf{\Psi}_{(N \times N)} \cdot \mathbf{A}_{x(N \times 1)} \quad (12)$$

$$= \Phi_{(M \times N)} \cdot \mathbf{\Psi}_{(N \times N)} \cdot \mathbf{W}_{D(N \times 1)}. \quad (13)$$

The optimal CS-MASK Θ^{Best} can be implemented as following constraint:

$$\Theta^{Best} = \arg \min_{\mathbf{A}} \|\mathbf{A}_x\|_1, \quad (14)$$

$$s.t. \sum_{k=1}^N \sum_{n=1}^M \frac{1}{MN} [\mathcal{F}_{2d}^{-1} \{\Theta^{Best}\} - \mathbf{W}_D] \leq \varepsilon, \quad (15)$$

where \mathcal{F}_{2d}^{-1} is two-dimensional (2D) Fourier inverse transform, ε represents a user-specified bound to constrain the error in an acceptable range. Because the 2D Fourier transform of ambiguity function is WVD, the efficacy of the feature CS-MASK can be examined by the error between its 2D Fourier inverse transform and WVD according to Equation (15).

Figure 4 depicts CS-MASK of each signal. The signal length is 256, the size of the original ambiguity function is 256×256 , in comparison, the size of CS-MASK extracted according to Equations (14) and (15) is 24×24 , meaning a great reduction of the feature size.

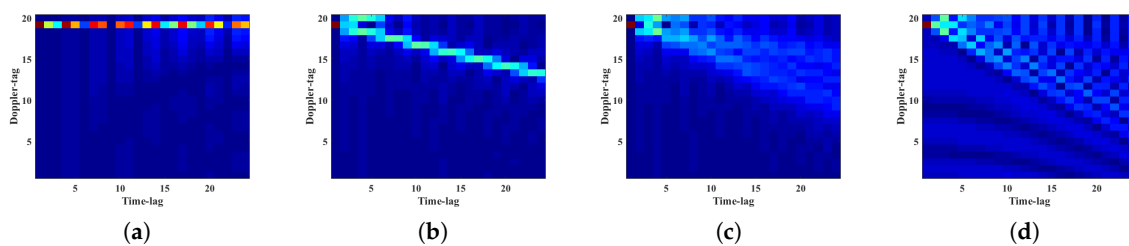


Figure 4. Compressive sensing mask (CS-MASK) of several typical simulation signals. (a–d) are CS-MASK of $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$.

The representative information contained in CS-MASK can be measured by the error between its recovered WVD and original WVD. Figure 5 shows the WVD of the original signal and WVD of the signal reconstructed by CS-MASK. It is really laudable that, even though CS-MASK only takes the zone in size of 24×24 near the origin, it contains the most information to reconstruct the complete WVD.

Nevertheless, there are still some imperfections in CS-MASK [7]: (1) the selection of the mask region is based on the optimization criterion related to reconstructed WVD, so the calculation amount is large, even though the rapid optimization method in two projects is proposed [6]; (2) the size of the optimized feature is not fixed, but it will change with certain factors, such as signal modulation and noise interference.

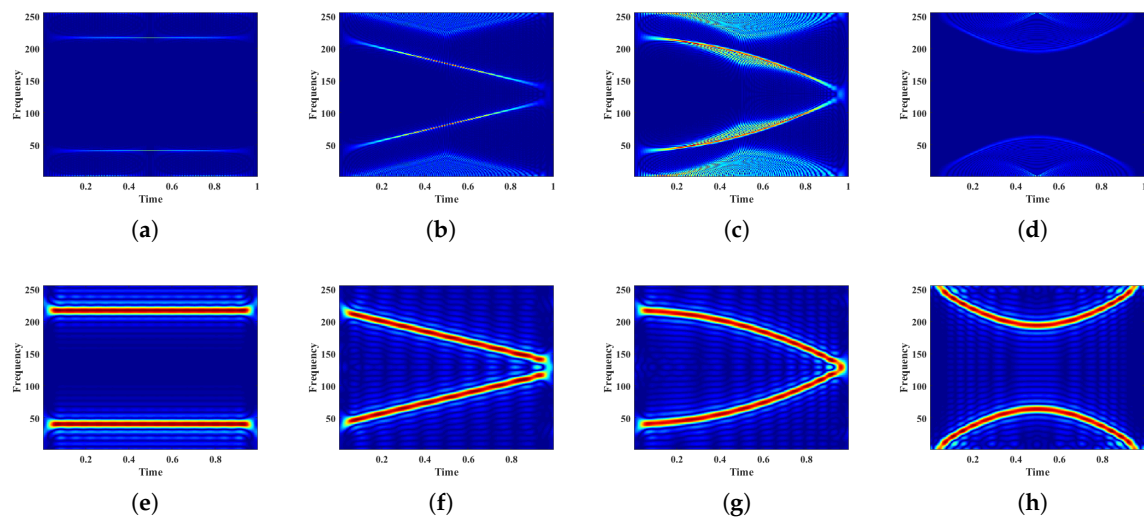


Figure 5. Comparison of inverse-transformed CS-MASK and WVD. (a–d) are WVD of $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$; (e–h) are WVD inverse-transformed from CS-MASK respectively.

3. MC-Feature

In this section, the rationale of MC-feature is elucidated in details. As discussed above, 2D features can be regarded as gray scale images, since the neural network perceive a 2D input from value of each pixel not the object in human cognition. Therefore, a lot of saliency detection methods used in image processing, like LRP, input cropping, deconvolution, and gradient algorithms [16] can be also applied in SEI. Image-specific class saliency visualization (ISCSV), an effective saliency detection technique based on gradient algorithm, is the theoretical basis of our proposed method. Firstly, ISCSV is introduced in Section 3.1. Subsequently, the production of saliency maps is shown in Section 3.2. Finally, the operation procedures of the proposed method are presented in Section 3.3.

3.1. Image-Specific Class Saliency Visualisation

With the rapid development of computing capability of electronic devices, deep network, especially deep Convolutional Neural Networks (CNN) [17,18] now being one of the most prevalent choices for image classification [19,20] and remarkable achievements have been made. However, the inner recognition mechanism of CNN is still in lack of systematic interpretation. Understanding the cognition of CNN has become increasingly important and necessary when CNN is applied in some special scenarios where nearly any nominal error is unacceptable, like driverless automobiles, missile guidance, military radar image processing, and etc.. In a previous work, [21] visualised deep network by seeking an input image which maximises the neuron activity of an optimisation in the image space. Recently, the problem of CNNs visualisation was addressed in [22] by the Deconvolutional Network (DeconvNet) architecture, which aims to approximately reconstruct the input of each layer from its output. Yet, both of these two methods consume large computational resources [21,22]. Ref. [12] proposed a very handy method to obtain image-specific class saliency visualization (ISCSV) by calculating the accumulated gradients of each pixel during back propagation. In this paper, ISCSV is used as the visualization tool to generate a kind of saliency map that reflects the contribution of each pixel in 2D transform domain to recognition network. The details of processing is shown, as follows:

Assume that $V_x^g(t, f)$ is corresponding to a class i . A class score activation function $C_i(\cdot)$ describes the relationship between the input and output of classification CNN. Generally, a CNN contains multiple convolutional and full connection layers with their corresponding activation

functions; hence, $C_i(\cdot)$ is a highly non-linear function. The complete Taylor expansion of $C_i(\cdot)$ in the neighbourhood of $V_x^g(t, f)$ can be expressed as:

$$C_i(V)|_{V=V_x^g} = \sum_{n=0}^{\infty} \frac{C_i^{(n)}(V_x^g)}{n!} (V - V_x^g)^n. \quad (16)$$

Here, we approximate $S_i(\cdot)$ with a linear function in the neighbourhood of $V_x^g(t, f)$ by computing the first-order Taylor expansion:

$$\begin{aligned} S_i(V)|_{V=V_x^g} &\approx C_i(V_x^g) + C_i'(V_x^g)(V - V_x^g) \\ &= C_i'(V_x^g)V + C_i(V_x^g) - C_i'(V_x^g)V_x^g. \end{aligned} \quad (17)$$

Usually, there is a loss function $J(C_i, L)$ with respect to S_i and true labels L to measure the error between the current output of NN and target labels. Noting that the second and third item in right side of Equation (17), $C_i(V_x^g) - C_i'(V_x^g)V_x^g$, can be regarded as a constant, the partial derivative of $C_i(\cdot)$ and J with respect to the STFT V at V_x^g can be expressed as:

$$\left. \frac{\partial C_i}{\partial V} \right|_{V=V_x^g} = C_i'(V_x^g) \quad (18)$$

$$\begin{aligned} \left. \frac{\partial J}{\partial V} \right|_{V=V_x^g} &= \frac{\partial J}{\partial C_i} \frac{\partial C_i}{\partial V} \\ &= \frac{\partial J}{\partial S_i} C_i'(V_x^g). \end{aligned} \quad (19)$$

It is clear from Equation (19) that the magnitude of the derivative indicates which pixels need to be changed the least to affect the class score the most; therefore, it is visible to detect the positive contributing pixels in 2D feature as well as irrelevant pixels if the gradient of each pixel can be obtained.

3.2. Saliency Map

Based on successful applications of ISCSV in saliency detection, it is considered that the inner recognition mechanism can probably be reflected by ISCSV. As mentioned in Section 1, various transformed features with 2 dimension can be regarded as an equivalent to the input in Equation (16). In the beginning of forward propagation of CNN, each pixel in 2D transformed feature is perceived by different convolutional kernel. The partial derivative of the loss between network output and label with respect to network input is propagated back from layer-to-layer in the form of neuron error δ . An overview of back propagation as well as its detailed algorithmic presentation can be found in [23]. Here, we take a simple network model to explain how to obtain saliency maps of 2D transform feature.

Figure 6 shows the structure of a simple neuron network. $\omega_{ij}^{(q)}$ denotes the weight from j th neuron in $(q-1)$ th layer to i th neuron in q th layer, $b_i^{(q)}$ denotes the bias of i th neuron in q th layer, and $s_i^{(q)}$ is the weighted input of i th neuron in l th layer defined as:

$$s_i^{(q)} = \sum_{j=1}^{N-1} \omega_{ij}^{(q)} x_j^{(q-1)}, i, j = 1, 2, \dots, N, q = 1, 2, \dots, Q, \quad (20)$$

$x_j^{(q)}$ denotes the output of j th neuron in q th layer defined as:

$$x_i^{(q)} = \sigma(s_i^{(q)} + b_i^{(q)}), \quad (21)$$

where $\sigma(\cdot)$ denotes the activation function of q th layer. The output of this network y_i equals the output of last layer $x_i^{(Q)}$. This process is called forward propagation.

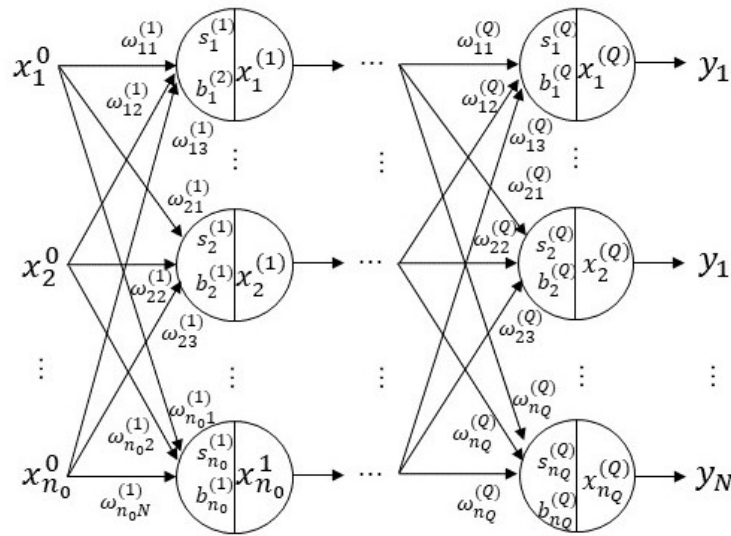


Figure 6. Structure of a neural network.

Now, we assume that there are P learning samples $x_p^{(0)}$ (STFT of processed radar signals) with their corresponding labels L_P as below:

$$x_p^{(0)} = [x_{p1}^{(0)}, \dots, x_{pn_0}^{(0)}], \quad (22)$$

$$L_p = [L_{p1}, \dots, L_{pn_Q}]. \quad (23)$$

After forward propagation, the output of last layer $x_p^{(Q)}$ can be obtained, and then a loss $E_p(L_p, x_p^{(Q)})$ is used to measure the error between the output of network and true labels. The loss function J can be formulated as the summation of each error:

$$J = \sum_{p=1}^P E_p. \quad (24)$$

Now, the partial derivative of the loss function J with respect to each weight $\omega_{ij}^{(q)}$ can be expressed as:

$$\frac{\partial J}{\partial \omega_{ij}^{(q)}} = \sum_{p=1}^P \frac{\partial E_p}{\partial \omega_{ij}^{(q)}}. \quad (25)$$

Accordingly, we can set $q = 1$ to obtain the gradient G_p of J with respect to weights $\omega_{ij}^{(1)}$ connecting the input data to the network:

$$G_p = \frac{\partial J}{\partial \omega_{ij}^{(1)}} = \sum_{p=1}^P \frac{\partial E_p}{\partial \omega_{ij}^{(1)}}, \quad (26)$$

where G_p is a matrix with the same size of the input data. Usually, the NN will be trained lots of epochs, and the gradient is accumulated in every epoch; hence, the final saliency map H_p can be calculated by the following formula:

$$H_p = \sum_{n=1}^K G_{p,n} \quad (27)$$

where n denotes the number of epoch, and K is the maximum of epoch.

3.3. MC-Feature

The proposed method can be compartmentalized into three main parts: initial training, saliency map production, and re-training. Firstly, the 2D transform feature of radar signals are divided into training set and testing set, then the training set is fed into a network1 as input to train it, and the testing set is used to measure the performance of network1. Secondly, all of the data are sent to forward propagate the well-trained network1 again, and then back propagate the error to the first hidden layer. In this way, a set of saliency maps for each 2D transform feature can be obtained. Thirdly, the saliency maps of original training set are sent to a new network with identical structure of network1 to do training. The performance of this new network can be measured by the saliency maps of original testing set. Figure 7 shows the flowchart of the proposed algorithm and Algorithm 1 elaborates the detailed steps of this algorithm.

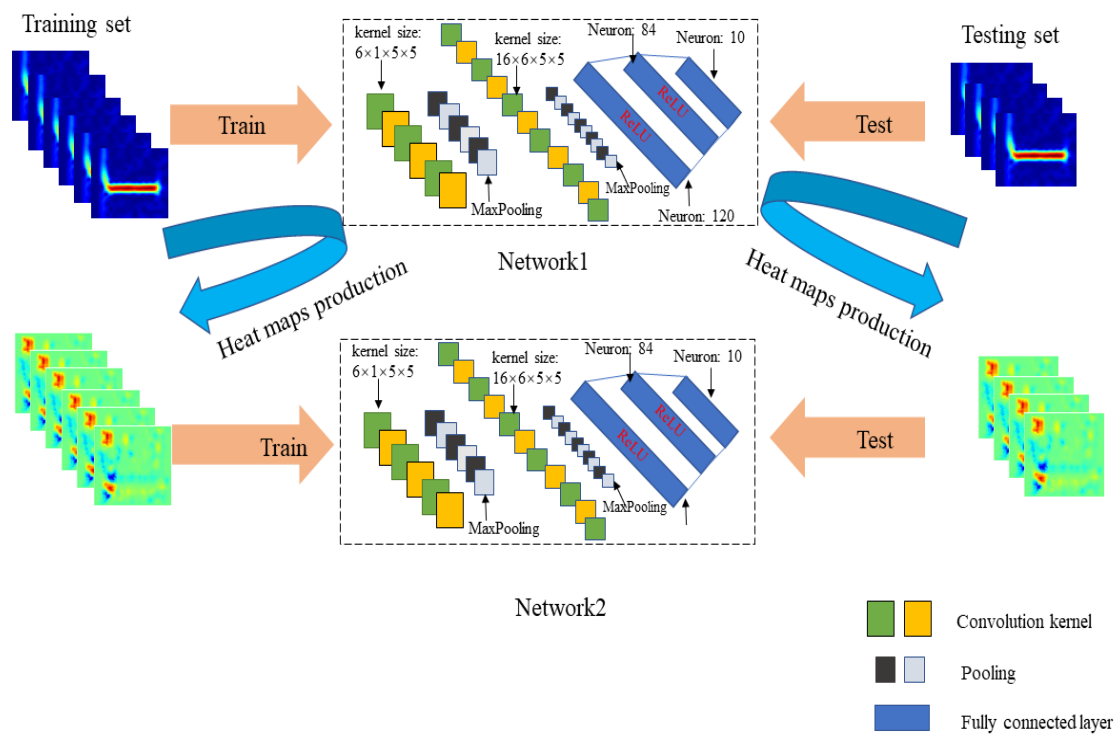


Figure 7. Data flow diagram of the proposed algorithm.

The procedures of the proposed algorithm (Algorithm 1)

Algorithm 1: Machine Cognition Based Feature Extraction

Input: \mathbf{V} the transform feature of radar signals in size of (N, C, W, H) as well as corresponding labels \mathbf{L} in size of $(N, 1)$, N, C, W, H denote the number, channel, width, and height of the \mathbf{V} , training rate λ represents a ratio of the number of training set to testing set

Step1: Data Partition: Training set:

$\mathbf{V}_{tr} = \mathbf{V}(1:\lambda N), \mathbf{L}_{tr} = \mathbf{L}(1:\lambda N)$; Testing set: $\mathbf{V}_{te} = \mathbf{V}(\lambda N + 1:N), \mathbf{L}_{te} = \mathbf{L}(\lambda N + 1:N)$

Step2: Feed the network1 with \mathbf{V}_{tr} and \mathbf{L}_{tr} :

do:

$\text{output1} = \text{network1.model}(\mathbf{V}_{tr})$ \triangleright train the network with both forward and back propagation

$\text{loss} = J(\text{output1}, \mathbf{L}_{tr})$ \triangleright loss function

$\text{loss.backward}()$

while $\text{loss} > \mu$ || $\text{epoch} < K$ $\triangleright \mu$ is the lowest acceptable loss, and K is the maximum of epochs

Obtain a well trained network1

Step3: Saliency maps production:

$\text{score} = \text{network1.forward}(\mathbf{V})$ \triangleright the classification result of network1

$\text{score.backward}(\mathbf{L})$

$\mathbf{H}_p = \mathbf{V}.\text{grad.data}$

Step4: Retrain a new network2 whose structure is identical with network1

Feed the network2 with $\mathbf{H}_p(1:\lambda N)$ and $\mathbf{L}(1:\lambda N)$

Repeat Step2

Obtain a well trained network2

Step5: Test the network2 with $\mathbf{H}_p(\lambda N + 1:N)$ and $\mathbf{L}(\lambda N + 1:N)$

Output: network2 and \mathbf{H}_p

4. Experimental Results

In this section, the detailed experimental results are presented and analyzed. Section 4.1 introduces the detailed information about the data used in our experiment. Section 4.2 depicts the structure of the recognition CNN. Section 4.3 shows the analysis of experimental results, including the interpretation of CNN and comparison of recognition rate by different features.

4.1. Data Information

In real scenarios, there are numerous complex types of radar signals for different applications. According to types of radar transmission, signals can be divided into impulse radar signal, swept frequency radar signal, and continuous-wave radar signal [24]. According to the modulation mode, the radar signal can be divided into analog modulated signal and digital modulated signal. Analog modulation includes linear frequency modulation, quadrature amplitude modulation (QAM), triangular frequency modulation, and etc. [25]. Digital modulation includes binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), 8 phase shift keying (8PSK), and etc. Besides, ultra wide bandwidth (UWB) signal has been used successfully in radar systems for many years [26]. In this experiment, four datasets are selected. In each dataset, the signals are set with the same parameters (such as radio frequency and frequency modulation mode) and transmitted by 10 radars of same type produced by the same manufacturer. 10 types of 500 approximate single-frequency signals from 10 civil aviation meteorological radars are involved in Database I. Database II contains 10 types of 500 single-frequency signals from 10 radar generators. Database III and database IV include 10 types of 500 (LFM) signals gleaned from radar signal generators. The time sampling points of each signal is in length of 500. The reasons why these databases are selected are because (1) they are typical and representative radar signals in SEI; and, (2) limited by experimental conditionals, signals with more complex modulation are unavailable for us. We randomly select a signal from each dataset to show its STFT in Figure 8.

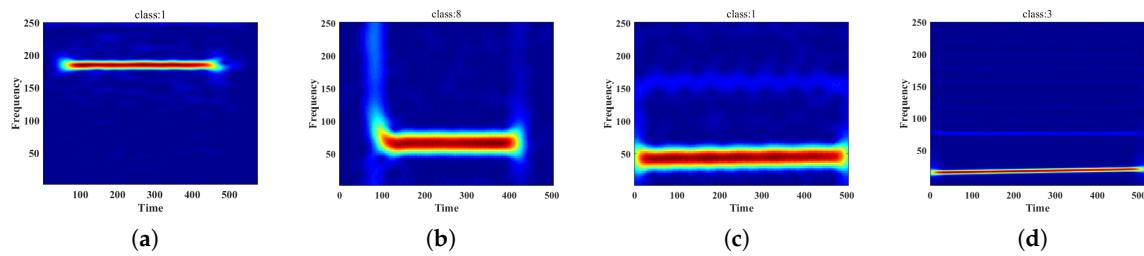


Figure 8. STFT spectrum of each database, the signal in (a) is from DatasetI, the signal in (b) is from DatasetII, the signal in (c) is from DatasetIII, the signal in (d) is from DatasetIV.

4.2. Recognition CNN Structure

The structure of CNN influences the final recognition effect greatly, hence, some extraordinary artificial intelligence professions delved into network structure and numerous effective network structures were designed, like LeNet-5 designed by Yann LeCun [27], AlexNet proposed by Hinton and Alex Krizhevsky [28], VGG proposed by Oxford Visual Geometry Group [29], GoogleNet proposed by Christian Szegedy [30], and etc.. It should be noted that this paper focuses on studying “what” is learned by a certain network other than the difference between various networks. Accordingly, LeNet-5, a very simple and effective CNN is used as classifier in this paper. Figure 9 shows the structure of LeNet-5 which contains 7 hidden layers: 2 convolutional layers, 2 pooling layers, and 3 fully connected layers. It should be pointed that this kind of parameters setting is only suitable for STFT and MC-feature. As for AF-RS and CS-MASK, the parameters of convolutional kernel and fully connected layers should be adjusted according to the size of these two features.

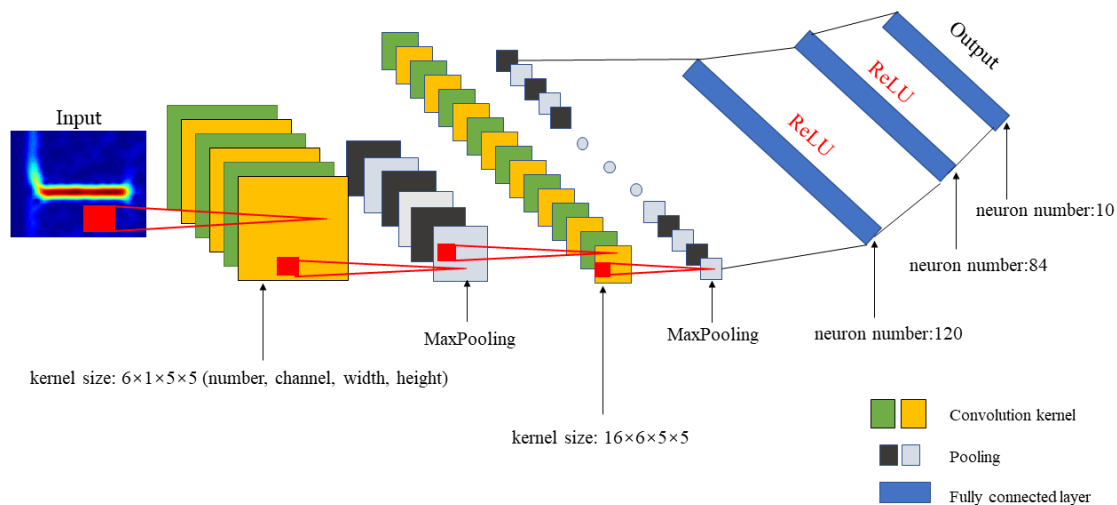


Figure 9. Structure of LeNet 5.

4.3. Results Analysis

Figures 10–13 exhibit AF-RS, CS-MASK, STFT, and MC-feature of two signals belong to different class in each dataset. It should be noted that the bright red and dark blue pixels in the proposed method represent high magnitude of gradient, meaning a great impact on classification of network, whereas the light green pixels mean nominal contribution to classification. In this way, we can obtain the information that network really “looks” in the process of recognition. In comparison to AF-RS and CS-MASK, which usually modify the area where the energy is concentrated in the transformed domain, the proposed features are visibly more representative, shown in Figures 10–13. It is clear that the divergence of other three features of two different class signals is not very apparent, while the divergence of the proposed method can even be distinguished by eyes.

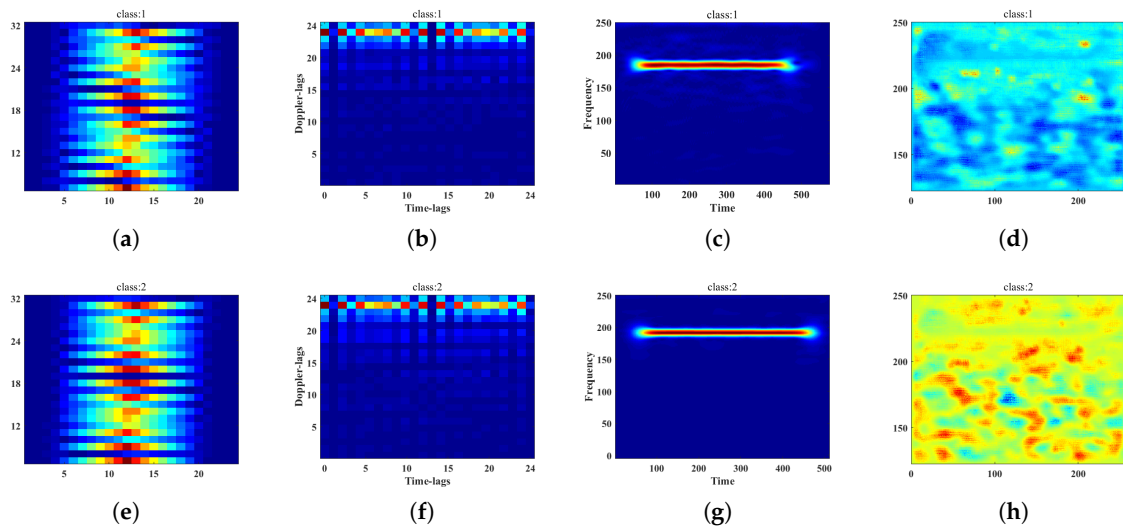


Figure 10. (a–d) are the AF-RS, CS-MASK, STFT, and machine cognition (MC)-feature feature of a signal of class 1 in dataset I; (e–h) are respectively features of a signal of class 2 in dataset I.

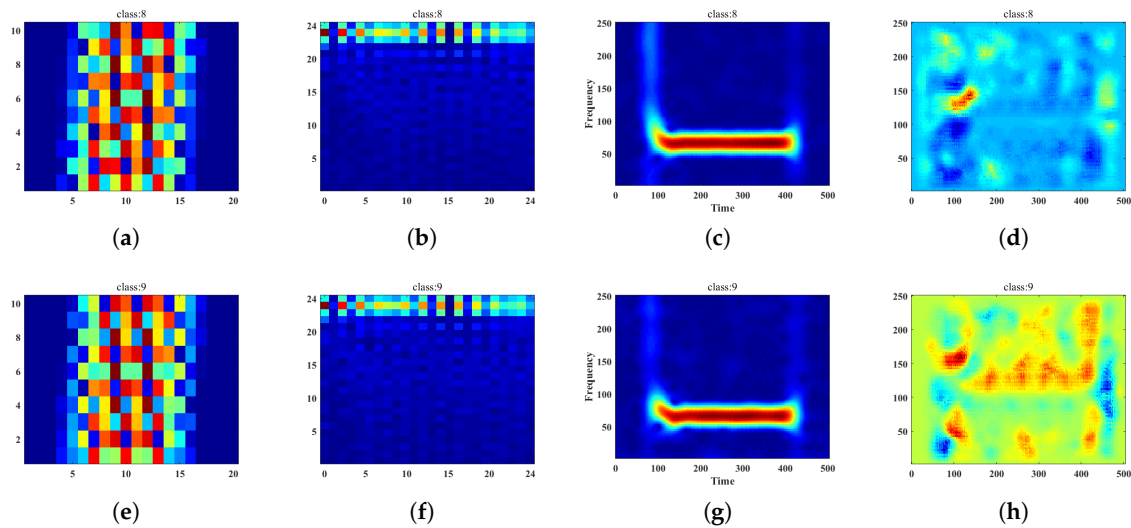


Figure 11. (a–d) are the AF-RS, CS-MASK, STFT, and MC-feature feature of a signal of class 8 in dataset II; (e–h) are respectively features of a signal of class 9 in dataset II.

In order to verify the superiority of the proposed algorithm, AF-RS, CS-MASK, STFT, and Machine Cognition Based feature are sent into LeNet 5 to recognize. We set $\mu = 0.01$, $K = 400$, and training rate from 5% to 60%. Tables 1–4 show the recognition rate of each dataset. Figure 14 exhibits the recognition rate and training rate curves of different features in each data set. In general, the proposed feature obtains a very high recognition rate when compared with other features in each dataset, even the training rate is low. For dataset I, AF-RS cannot provide the network with enough useful information, leading to a result that 10 types of signals are divided into only 2 class; hence, the recognition rate retains approximately 20% no matter how training rate changes. For dataset II and dataset III, AF-RS and CS-MASK contain some information that is conducive to the recognition task of network; however, the recognition rate declines obviously when the training rate is below 20%. In addition, even the training rate is over 40%, the network cannot learn more information from AF-RS and CS-MASK. It is probably because these two features may lose some detailed texture information at cost of dimensionality compression. Similar phenomenon also appears in dataset IV that the recognition rate is only 10% of AF-RS and CS-MASK when the training rate is 5%, which means that the network

learns nothing from these two features. In four datasets, the performance of STFT is relatively stable, but still much lower than that of the proposed feature, especially when the training rate is less than 30%.

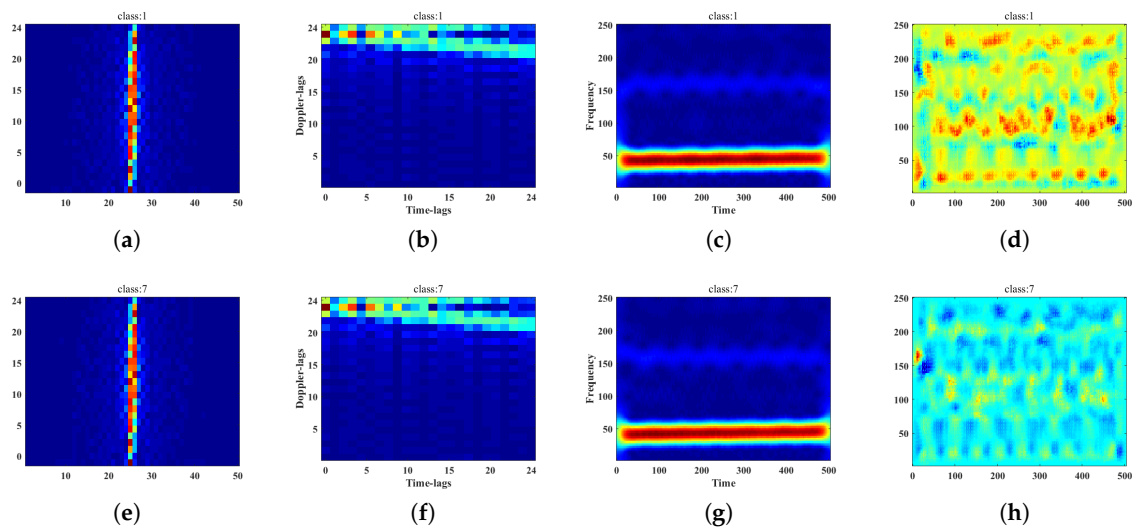


Figure 12. (a–d) are the AF-RS, CS-MASK, STFT, and MC-feature feature of a signal of class 1 in dataset III; (e–h) are respectively features of a signal of class 7 in dataset III.

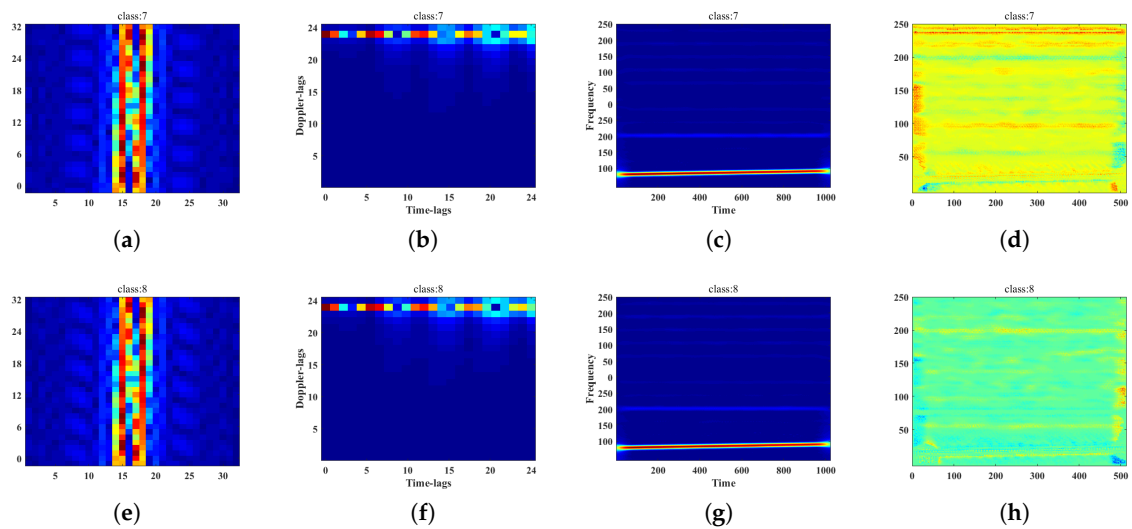


Figure 13. (a–d) are the AF-RS, CS-MASK, STFT, and MC-feature feature of a signal of class 7 in dataset IV; (e–h) are respectively features of a signal of class 8 in dataset IV.

Table 1. The recognition rate of each feature for Dataset I.

Training Rate(%)	Feature			
	AF-RS	CS-MASK	STFT	MC-Feature
5	22%	39%	66%	95%
10	24%	42%	76%	98%
20	21%	41%	83%	99%
30	22%	61%	75%	97%
40	27%	63%	80%	99%
50	22%	62%	79%	98%
60	23%	72%	81%	99%

Table 2. The recognition rate of each feature for DatasetII.

Feature Training Rate(%)	AF-RS	CS-MASK	STFT	MC-Feature
5	44%	45%	46%	91%
10	61%	63%	67%	98%
20	64%	67%	68%	99%
30	69%	69%	69%	98%
40	67%	67%	78%	97%
50	78%	74%	80%	97%
60	77%	69%	92%	97%

Table 3. The recognition rate of each feature for DatasetIII.

Feature Training Rate(%)	AF-RS	CS-MASK	STFT	MC-Feature
5	46%	30%	52%	89%
10	57%	36%	74%	89%
20	70%	47%	84%	96%
30	75%	61%	85%	97%
40	76%	64%	84%	100%
50	73%	61%	95%	98%
60	85%	71%	96%	100%

Table 4. The recognition rate of each feature for DatasetIV.

Feature Training Rate(%)	AF-RS	CS-MASK	STFT	MC-Feature
5	10%	9%	66%	89%
10	58%	58%	82%	94%
20	60%	60%	85%	99%
30	56%	66%	88%	97%
40	57%	63%	88%	98%
50	58%	66%	92%	100%
60	56%	65%	93%	100%

The experimental results demonstrate the superiority of the proposed method. Furthermore, they also indicate that some information based on human cognition, like the area concentrated energy, may be not the key to a classification network, and some more informative parts behind the primitive 2D features are not mined. From the perspective of machine learning, deep informative information can be extracted in the proposed algorithm.

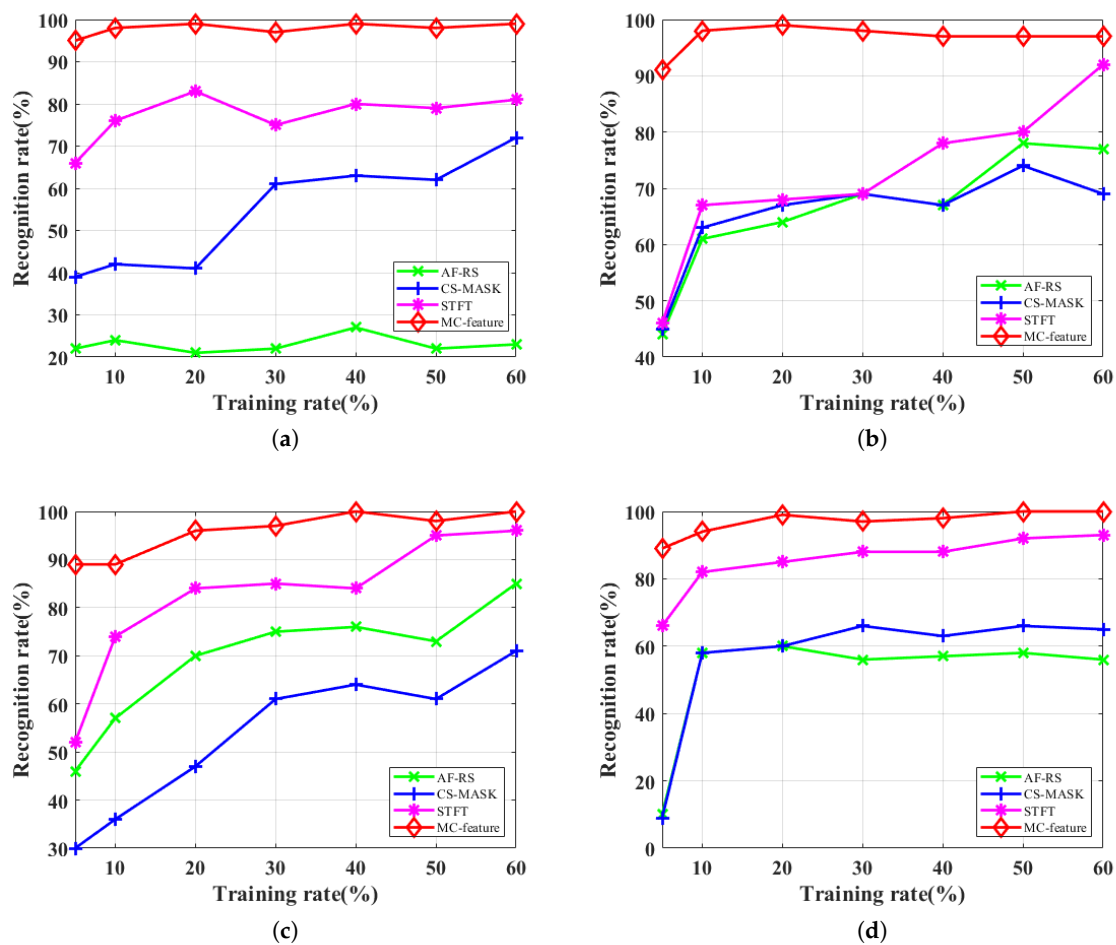


Figure 14. Comparison of recognition rate of four features in Dataset I, II, III, IV. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV.

5. Conclusions

In this paper, we propose a novel SEI feature from perspective of machine cognition instead of human cognition. The MC-feature can obtain considerably higher recognition accuracy in comparison to other handcrafted features, particularly under scant data samples, which can greatly alleviate the model immaturity that is caused by insufficient training radar signals. In addition, even though MC-feature still remains to be deeply interpreted, it demonstrates that the information network relies on for identification is much different from various handcrafted features based on human cognition. It illuminates that it is quite necessary and potential to understand the inner mechanism of recognition network. In the future, the clear and profound interpretation of the exact physical meaning of MC-feature will be the focus of our study. Once the specific physical meaning of MC-feature can be explained, the recognition network will no longer be a "black box", but an analytical mathematics tool, which will widely broaden the application of machine learning in many scenarios. In the future, the research on clear physical meaning of MC-feature of complex radar signals, like UWB and frequency hopping signals, will be the focus of our research team.

Author Contributions: The contributions of the authors are as follows. Methodology, M.Z.; Conceptualization, Z.F.; Data Curation, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shaanxi Province (2019JM-412) and the National Natural Science Foundation of China (61701374).

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

References

1. Yang, L.B.; Zhang, S.S.; Xiao, B. Radar Emitter Signal Recognition Based on Time-frequency Analysis. In Proceedings of the IET International Radar Conference 2013, Xi'an, China, 14–16 April 2013; pp. 1–4.
2. Lu, J.; Xu, X. Multiple-Antenna Emitters Identification Based on a Memoryless Power Amplifier Model. *Sensors* **2019**, *19*, 5233. [[CrossRef](#)] [[PubMed](#)]
3. Zhu, M.; Zhou, X.; Zang, B.; Yang, B.; Xing, M. Micro-Doppler Feature Extraction of Inverse Synthetic Aperture Imaging Laser Radar Using Singular-Spectrum Analysis. *Sensors* **2018**, *18*, 3303. [[CrossRef](#)] [[PubMed](#)]
4. Wang, X.; Huang, G.; Zhou, Z.; Tian, W.; Yao, J.; Gao, J. Radar Emitter Recognition Based on the Energy Cumulant of Short Time Fourier Transform and Reinforced Deep Belief Network. *Sensors* **2018**, *18*, 3103. [[CrossRef](#)] [[PubMed](#)]
5. Wang, L.; Ji, H.; Shi, Y. Feature Extraction and Optimization of Representative-slice in Ambiguity Function for Moving Radar Emitter Recognition. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, TX, USA, 14–19 March 2010; pp. 2246–2249.
6. Zhu, M.; Zhang, X.; Qi, Y.; Ji, H. Compressed Sensing Mask Feature in Time-Frequency Domain for Civil Flight Radar Emitter Recognition. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2146–2150.
7. Zhu, M.; Feng, Z.; Zhou, X. Specific Emitter Identification Based on Synchrosqueezing Transform for Civil Radar. *Electronics* **2020**, *9*, 658. [[CrossRef](#)]
8. Ayhan, B.; Kwan, C. Robust Speaker Identification Algorithms and Results in Noisy Environments. In Proceedings of the International Symposium on Neural Networks, Minsk, Belarus, 25–28 June 2018.
9. Wang, D.; Brown, G.J. Computational Auditory Scene Analysis: Principles, Algorithms and Applications. *J. Acoust. Soc. Am.* **2008**, *124*, 13.
10. Misaghi, H.; Moghadam, R.A.; Mahmoudi, A.; Salemi, A. Image Saliency Detection By Residual and Inception-like CNNs. In Proceedings of the 2018 6th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, 23–25 October 2018; pp. 94–99.
11. Ramik, D.M.; Sabourin, C.; Moreno, R.; Madani, K. A Machine Learning Based Intelligent Vision System for Autonomous Object Detection and Recognition. *Appl. Intell.* **2014**, *40*, 94–99. [[CrossRef](#)]
12. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv* **2013**, arXiv:1312.6034.
13. Auger, F.; Flandrin, P.; Lin, Y. Time-Frequency Reassignment and Synchrosqueezing: An Overview. *IEEE Signal Process. Mag.* **2013**, *30*, 32–41. [[CrossRef](#)]
14. Gillespie, B.W.; Atlas, L.E. Optimizing Time-Frequency Kernels for Classification. *IEEE Trans. Signal Process.* **2001**, *49*, 485–496. [[CrossRef](#)]
15. Gillespie, B.W.; Atlas, L.E. Optimization of Time and Frequency Resolution for Radar Transmitter Identification. In Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Phoenix, AZ, USA, 15–19 March 1999; pp. 1341–1344.
16. Islam, N.U.; Lee, S. Interpretation of Deep CNN Based on Learning Feature Reconstruction With Feedback Weights. *IEEE Access* **2019**, *7*, 25195–25208. [[CrossRef](#)]
17. Kim, J.; Kim, J.; Kim, H.; Shim, M.; Choi, E. CNN-Based Network Intrusion Detection Against Denial-of-Service Attacks. *Electronics* **2020**, *9*, 916. [[CrossRef](#)]
18. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
19. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
20. Ciregan, D.; Meier, U.; Schmidhuber, J. Multi-column Deep Neural Networks for Image Classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3642–3649.
21. Erhan, D.; Bengio, Y.; Courville, A.; Vincent, P. *Visualizing Higher-Layer Features of a Deep Network*; Technical Report 1341; University of Montreal: Montreal, QC, Canada, 2009.
22. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *arXiv* **2013**, arXiv:1311.2901.

23. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations By Back-Propagating Errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
24. Wanga, X.; Li, J.; Yang, Y. Comparison of Three Radar Systems for Through-the-Wall Sensing. In *Radar Sensor Technology XV*; Spie-the International Society for Optical Engineering: Bellingham, WA, USA, 2011.
25. Yao, Y.; Li, X.; Wu, L. Cognitive Frequency-Hopping Waveform Design for Dual-Function MIMO Radar-Communications System. *Sensors* **2020**, *20*, 415. [[CrossRef](#)]
26. Hamran, S.E. Radar Performance of Ultra Wideband Waveforms. In *Radar Technology*; InTech: Rijeka, Croatia, 2009.
27. LeCun, Y.; Bengio, Y. Convolutional Networks for Images, Speech, and Time-Series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1998.
28. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2012.
29. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
30. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).