Wavelet Autoencoder for Radar HRRP Target Recognition with Recurrent Neural Network

Mengjiao Zhang(✉) and Bo Chen

National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China
mengjiao_123@hotmail.com, bchen@mail.xidian.edu.cn

Abstract. A Wavelet Autoencoder model with Recurrent Neural Network (WaveletAE with RNN) is developed for radar automatic target recognition (RATR), with an encoder-decoder layer, in which the weights of decoder are fixed as a set of overcomplete bases derived from mother wavelet. Imposing an sparsity constraint on the hidden units of encoder-decoder layer, interesting structure in the data is discovered, and superior recognition performance is achieved on the measured High Resolution Range Profiles (HRRP) data, showing the effectiveness of the proposed model. Specific results are represented in our experiments.

Keywords: High resolution range profiles (HRRP)
Recurrent neural network (RNN)
Radar automatic target recognition (RATR) · Overcomplete bases
Joint training · Wavelet · Autoencoder (AE)

1 Introduction

An HRRP is the amplitude of coherent summations of the complex time returning from target scatterers in each range cell, which represents as a one-dimensional signature along the radar line of sight (LOS). HRRP infers to some target structure signatures, such as target size, scatterer distribution, etc. Compared with inverse synthetic aperture radar (ISAR) and synthetic aperture radar (SAR) images, HRRP can be acquired more easily and processed more efficiently. Thus, radar HRRP target recognition has received intensive attention [1–5].

One classical RATR approach is comprised of a feature extraction step that computes robust features from target data and a classification step that calculates classes based on features. Extracting proper features can not only reduce system complexity and processing time, but also improve system performance in many aspects [6].

The energy spectrum, bispectrum features and high-order spectra features extracted using Fourier transform and wavelet transform performed well [7–9]. These traditional systems have mainly relied on elaborate models incorporating

Y. Peng et al. (Eds.): ISciDE 2018, LNCS 11266, pp. 262–275, 2018.
https://doi.org/10.1007/978-3-030-02698-1_23

carefully hand engineer features or large amount of prior knowledge [9]. However, without sufficient prior knowledge for the application, those features could be weak and incomplete. Many statistical recognition methods aim at choosing an appropriate model that describes HRRP's statistical property accurately [10]. But most of those statistical models ignore the correlation and dependency between range cells, and the algorithm for updating parameters will cost much time when the scale of dataset is too large. In [11], a linear dynamic model was investigated and [12] presented a hidden Markov model (HMM) to model the time correlations of HRRP. However, both the models are under Gauss Markov Assumptions [13] which will suffer from limited capabilities of Gaussian distributions in each HMM state and it cannot take long-term dependency into account. Recently, nonlinear deep neural networks like Deep Belief Networks (DBN) and Stacked Denoising Autoencoders (SDAE) show superior performance in various real world tasks, and the hierarchical structure with non-linear activation is similar to human visual cortex [14,15]. Yet, they adapt the layer-wise unsupervised greedy learning method and ignore the correlation and dependency as well. Recurrent Neural Network (RNN) [16], another important branch of the deep neural networks family, are mainly designed for handling sequences. In [17], an attention-based Recurrent Neural Network Model is proposed, which is effective to HRRP target recognition.

In this paper, we not only take advantages of the correlation and dependency, but also the spectrum features of HRRP. Instead of a two-stage approach with feature extractor and classifier, a model that directly captures the objective of the ultimate recognition task is taken into consideration. Combining the bases derived from mother wavelet with the RNN model, we propose a wavelet autoencoder (waveletAE) model with RNN for HRRP target recognition, which capitalizes on spectrum features as well as the correlation and dependency between range cells. Unlike other networks, adapted by a layer-wise unsupervised greedy learning method, this kind of model is supervised and can be jointly trained. Detailed experiments have been done to examine and analyze the recognition performance as well as the learned features.

The remainder of the paper is organized as follows. Section 2 is a brief description of HRRP, Fourier Transform and Wavelet Transform. An introduction of RNN is also included. Later, preprocessing of the HRRP and the waveletAE with RNN for HRRP target recognition are discussed in Sect. 3. Several detailed experiments and results are presented in Sect. 4, with the conclusion and future work provided in Sect. 5.

2 Preliminary

2.1 Fourier Transform and Wavelet Transform for HRRP

Generally, the size of the targets or their components is much larger than the wavelength of the HRRP radar. As a result, those complex target such as an aircraft can be divided into many range *cells*. Within the same range cell, radar

signature is the summation of the complex time return from target scatterers. An HRRP $\boldsymbol{x}$ can be represented as:

$$\boldsymbol{x} = e^{j\phi}\left[x_1, x_2, \ldots, x_d\right]^T, \text{ with } x_n = \sum_{i=1}^{I_n} \alpha_{in} e^{j\psi_{in}}, n = 1, 2, \ldots, d \tag{1}$$

where ϕ is the initial phase, d is the number of range cells, I_n is the total number of scatterers in the nth range cell, and α_{in} and ψ_{in} are the reflectivity and phase of the ith scatterer in the nth range cell. The frequency spectrum $\boldsymbol{y}$ of $\boldsymbol{x}$ is given as follow:

$$\boldsymbol{y} = FFT(\boldsymbol{x}) = e^{j\phi}\left[y_1, y_2, \ldots, y_d\right]^T, \text{ with } y_f = \sum_{n=1}^{d} x_n e^{-j2\pi fn/d}, f = 1, 2, \ldots, d \tag{2}$$

where $FFT(\cdot)$ denotes the fast Fourier transform (FFT) and f is the spectral component. There are some advantages of the frequency spectrum of HRRP for target recognition due to its following appealing qualities: 1. FFT is an information preserving procedure, so there is no information lost in the transformation process; 2. Frequency spectrum is translation invariant, and the feature extracted from it can still enjoy this property; 3. A signal represented in the frequency domain may show characteristics which are not normally observed in the time domain (such as stationary) [6]. According to the Wide Sense Stationary-Uncorrelated Scattering model [18], and some mathematical derivation, it is concluded that the correlation function of two frequencies depends only on their difference, i.e. the frequency spectrum is a wide sense stationary process. From Eq. 2, we can see the frequency spectrum still suffers from initial-phase sensitivity. So only the amplitude of frequency spectrum $\boldsymbol{z} = |\boldsymbol{y}|$ is considered here. It is simple to prove $\boldsymbol{z}$ is still a stationary series.

Affected by many factors, HRRPs are non-stationary signals. Since wavelet transform provides high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies [19,20], Wavelet Transform (WT) and more particular the Discrete Wavelet Transform (DWT) is multi-resolution decomposition used to analyze signals and images. The basis function used in wavelet transform are locally supported; they are nonzero only over part of the domain represented. As an alternative to the short time Fourier Transform (STFT), it can overcome problems related to its frequency and time resolution properties.

The DWT is a special case of WT that provides a compact representation of a signal in time and frequency that can be computed efficiently. The DWT coefficients for sequence $f(n)$ is defined by the following equation [21]:

$$W_\phi(j_0,k) = \frac{1}{\sqrt{M}} \sum_n f(k)\phi_{j_0,k}(n)$$
$$W_\psi(j,k) = \frac{1}{\sqrt{M}} \sum_n f(k)\psi_{j,k}(n), for j \geq j_0 \tag{3}$$

where j_0 is an arbitrary starting scale. The $\phi_{j_0,k}(n)$ and $\psi_{j,k}(n)$ in the equations are sampled versions of basis function $\phi_{j_0,k}(x)$ and $\psi_{j,k}(x)$, where $\psi(x)$ is a time function with finite energy and fast decay called the mother wavelet. The complementary inverse DWT is

$$f(n) = \frac{1}{\sqrt{M}} \sum_k W_\phi(j_0,k)\phi_{j_0,k}(n) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_\psi(j,k)\psi_{j,k}(n) \tag{4}$$

Normally, we let $j_0 = 0$ and select M to be a power of 2 (i.e., $M = 2^j$) so that the summation in Eqs. 3 through 4 are performed over $n = 0, 1, 2, \ldots, M-1$,$j = 0, 1, 2, \ldots, J-1$, and $k = 0, 1, 2, \ldots, 2^j - 1$. The DWT analysis can be performed using a fast, pyramidal algorithm related to multirate filterbanks [22].

2.2 Recurrent Neural Network for HRRP Target Recognition

Recurrent neural networks (RNNs) are powerful models for sequential data. RNNs are inherently deep in time, since their hidden state is a function of all previous hidden states [23]. The network has an input layer $\boldsymbol{x}$, hidden layer $\boldsymbol{h}$ (also called state) and output layer $\boldsymbol{o}$. An HRRP, a 256-dimension vector, can be divided into several segments as a sequence, for echoes of the adjacent range cells have a time difference. In order to transform a vector $\boldsymbol{x}$ into a sequence, we apply a window function on the HRRP sample with the window size $w = 32$ and the overlap $p = 16$, as shown in Fig. 1. $x^{(t)}$ is the input of RNN at time t in the sequence, which can be described as $x^{(t)} = [x(p(t-1)+1) : x(p(t-1)+w)]$. At every time, there is an output to determine the label of the input, and all outputs vote for the final label.

Given an input sequence $\boldsymbol{x} = \left(x^{(1)}, \ldots, x^{(T)}\right)$, a standard recurrent neural network (RNN) computes the hidden vector sequence $\boldsymbol{h} = \left(h^{(1)}, \ldots, h^{(T)}\right)$ and output vector sequence $\boldsymbol{o} = \left(o^{(1)}, \ldots, o^{(T)}\right)$ by iterating the following equations from $t = 1$ to T:

$$h^{(t)} = f\left(\boldsymbol{W}_{xh}x^{(t)} + \boldsymbol{W}_{hh}h^{(t-1)} + \boldsymbol{b}_h\right) \tag{5}$$
$$o^{(t)} = g\left(\boldsymbol{W}_{ho}h^{(t)} + \boldsymbol{b}_o\right) \tag{6}$$

where the $\boldsymbol{W}$ terms denote weight matrices (e.g. $\boldsymbol{W}_{xh}$ is the input-hidden weight matrix), the $\boldsymbol{b}$ terms denote bias vectors (e.g. $\boldsymbol{b}_h$ is hidden bias vector) and f, g is the hidden layer function. Usually, $f(z)$ is selected as the *sigmoid* activation function, and $g(z)$ is $softmax$ function [16]:

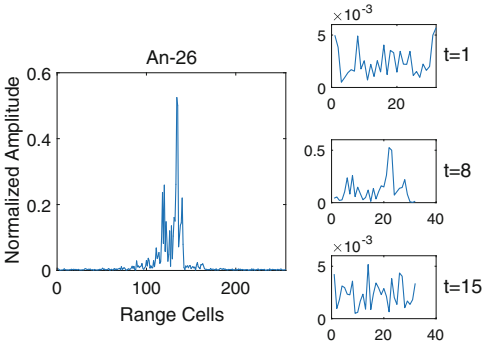


Fig. 1. Illustration of an HRRP sample from an aircraft target. (a) One-dimensional HRRP. (b)The samples of transformed HRRP sequence at time t.

$$f(z) = \frac{1}{1 + e^{-z}} \tag{7}$$

$$g(z_m) = \frac{e^{z_m}}{\sum\limits_k e^{z_k}} \tag{8}$$

The model is trained using the backpropagation through time (BPTT) to maximize the correct probability of the predict output. A loss L measures how far each $\boldsymbol{o}$ is from the corresponding training target $\boldsymbol{y}$, in which N represents the number of the input data:

$$L_1 = -\frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \left(\sum_{i=1}^{K} y_{(ni)}^{(t)} log \left(o_{(ni)}^{(t)} \right) \right) \tag{9}$$

3 Wavelet Autoencoder for HRRP Target Recognition with Recurrent Neural Network

In Sect. 2, we have discussed about the RNN model for HRRP target recognition, which takes the correlation and dependency between range cells, with the one-dimensional input $\boldsymbol{x}$ transformed into a sequence. As mentioned before, spectrum features have several appealing qualities and HRRPs are non-stationary signals, so DWT is more suitable to analyze HRRPs compared to FFT. First at each time, we can take FFT or DWT on original input sequence and then take the transformed result as input of RNN. However, the spectrum transform is an extra process in this way, which is separated from RNN. What we want is a more integrated model—once the input of the model is given, the output is what we want.

As we all know, both FFT and DWT can be seen as the linear combination of a set of bases, and the signal can be reconstructed by another set of bases, which is similar to autoencoder (AE). In the encoder, the deterministic mapping f_θ transforms an input vector $\boldsymbol{x}$ into hidden representation $\boldsymbol{s}$, and then decoder maps $\boldsymbol{s}$ back to $\hat{\boldsymbol{x}}$, reconstruction of $\boldsymbol{x}$, with function $g_{\theta'}$, $\theta = \{W, b\}$, and $\theta' = \left\{W', b'\right\}$. Instead of automatically learning the weights of AE, the weights of decoder can be set as the fixed overcomplete bases derived from the mother wavelet, W_{idwt}, and the reconstruction $\hat{\boldsymbol{x}}^{(t)}$ at each time is a linear mapping from $s^{(t)}$:

$$s^{(t)} = f(\boldsymbol{W}_{xs} x^{(t)} + \boldsymbol{b}_s) \tag{10}$$

$$\hat{x}^{(t)} = \boldsymbol{W}_{sx} s^{(t)} = \boldsymbol{W}_{ifft(idwt)} s^{(t)} \tag{11}$$

This can be called wavelet bases autoencoder (WaveletAE). Theoretically, infinite bases can be obtained with a mother wavelet, and only several are involved for DWT. However, we can fix the weights of decoder $\boldsymbol{W}_{sx}$ as a set of overcomplete bases derived from mother wavelet to reconstruct the input as $\hat{\boldsymbol{x}}$ and learn the transform matrix $\boldsymbol{W}_{xs}$ from $\boldsymbol{x}$ to $\boldsymbol{s}$. The loss of WaveletAE over all HRRP samples $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ can be written as follow:

$$L_2 = -\frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \|x_n^{(t)} - \hat{x}_n^{(t)}\|_2^2 \tag{12}$$

AE is forced to learn a *compressed* representation of the input. If there is structure in the data, for example, if some of the input features are correlated, then this algorithm will be able to discover some of those correlations. The argument above relied on the number of hidden units s being small, however, the bases are overcompelete in WaveletAE. But even when the number of hidden units is large (perhaps even greater than the number of input pixels), we can still discover interesting structure, by imposing other constraints on the network. In particular, if we impose a sparsity constraint on the hidden units, then the AE will still discover interesting structure in the data, even if the number of hidden units is large. We choose the following penalty term [24]:

$$penalty = \sum_{j=1} \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \tag{13}$$

$\hat{\rho}_j$ is the average activation of hidden unit j (averaged over the training set):

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} s_i^{(t)} \tag{14}$$

and ρ is a sparsity parameter, typically a small value close to zero (say $\rho = 0.01$).

If we put $\boldsymbol{s}$ as the input of RNN, an integrated model, combining RNN with the wavelet bases autoencoder is derived. This model handles the time sequence

and takes the spectrum features into account, called WaveletAE with RNN, as shown in Fig. 2. The goal of the model is to minimize reconstruction error to get the distinctive physical structure and maximize the correct probability of the predict output.

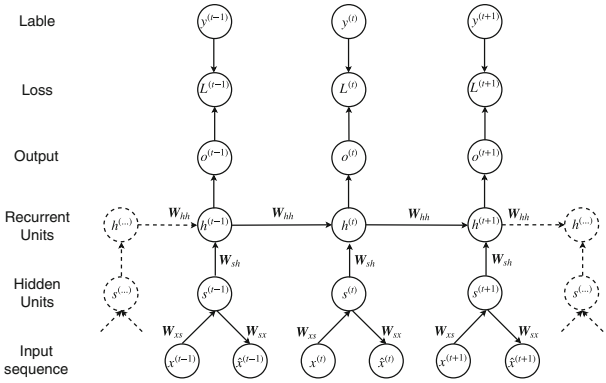


Fig. 2. WaveletAE with RNN

The calculation of recurrent units and hidden units are the same as RNN. But the loss function is a little bit different, which contains the loss of cross entropy L_1 and reconstruction L_2.

$$\begin{aligned} L &= L_1 + L_2 + penalty \\ &= -\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\sum_{i=1}^{K} y_{(ni)}^{(t)} log\left(o_{(ni)}^{(t)}\right) + \|x_n^{(t)} - W_{sx}s_t^{(n)}\|_2^2\right) + penalty \end{aligned} \quad (15)$$

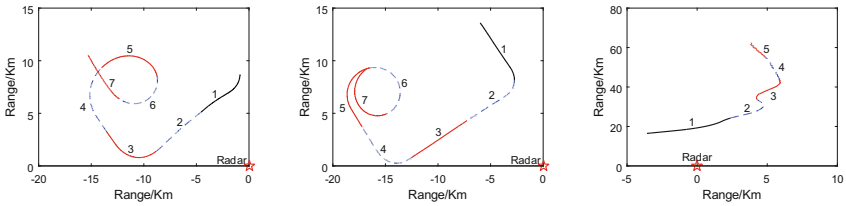
4 Experimental Results

In this section, we first study the methodology and introduce the measured HRRP data used in our experiments. And the detailed model settings and experiments to analyze the recognition performance of the model is presented and studied, respectively.

4.1 Measured HRRP Data

The results presented in this paper are based on measured data of three real airplanes. The center frequency and bandwidth of the radar are 5.52 GHz and 400 MHz, respectively. The projections of plane trajectories on to ground plane are segmented as displayed in Fig. 3.

The detailed size of each airplane and the parameters of the measured radar are listed in Table 1.



(a) An-26 (b) Cessna (c) Yark-42

Fig. 3. The projections of three target trajectories onto the ground plane

Table 1. Parameters of radar and planes

Radar parameters	Center frequency		5.52 GHz
	Bandwidth		400 MHz
Airplane	Length (m)	Width (m)	Height (m)
A-26	23.80	29.20	9.83
Cessna citation S/II	14.40	15.90	4.57
Yark-42	36.38	34.88	9.83

According to the preconditions for choosing the training and testing dataset: (a) The training dataset cover almost all of the target-aspect angles of the test dataset; (b) The elevation angles of the test dataset are different from those of the training dataset. Therefore, the fifth and sixth parts of An-26, the sixth and seventh parts of Cessna and the second and fifth parts of Yark-42 are chosen as training data with the remaining parts taken as testing data. There are 26000 HRRPs in each part except for the fifth part of Yark-42, which has 10000 HRRPs. In addition, the dimension of HRRP is 256. In our experiment, there are 7800 HRRP samples for training and 2600 each class; the number of test samples is 5200.

As discussed in literature [25], it is a prerequisite for radar target recognition to deal with the target-aspect, time-shift, and amplitude-scale sensitivity. Similar to the previous study [25,26], HRRP training samples should be aligned by the time-shift compensation techniques in ISAR imaging to avoid the influence of time-shift sensitivity. Each HRRP sample is normalized by L_2 normalization algorithm to avoid the amplitude-scale sensitivity. In the following experiments, all of the HRRPs are assumed to have been aligned and normalized. And considering the echoes from different range cells as a sequence, the HRRP sample is separated into 15 time steps as mentioned in Sect. 2. So, for an HRRP sample, there are 15 time steps, and at each time the dimension of input $x^{(t)}$ is 32.

4.2 Model Setting

The proposed model consists of the input layer, encoder-decoder layer, recurrent layer, and the output layer. The number of the recurrent units in the model is

15, and weights are shared at each time. The dimension of the input at each time is 32 and output 3; the size of one recurrent unit $h^{(t)}$ is usually 4–30. And for hidden layer and recurrent layer, the activation is *sigmoid* function.

For initialization, all the weights are set to a matrix of small values, using truncated normal (zero mean and 0.1 variance) except the decoder matrix W_{sx}. W_{sx} is a set of fixed bases. It can be derived from mother wavelet with different scale and shift or consists of some FFT bases of different points. Besides the bases of wavelet transform, more bases are included in W_{sx}, which are considered overcomplete. Several mother wavelet bases are discussed in our experiment, for example, $biorNr.Nd$ means biorthogonal wavelets, Nr, Nd are the reconstruction order and decomposition order respectively; dbN are the wavelets belong to Daubechies family, N means the order.

The model is trained in several epochs, in which all data from training set are sequentially presented. For optimization, we use the Adam optimizer [27], and reduce the value of the learning rate after several epochs to accelerate the convergence. The number of epoch is 300 and sparsity parameter ρ is 0.01.

4.3 Recognition Performance

To access the recognition performance of the proposed model on measured data, we carry out some experiments on other methods compared with our model. First we use the popular classification technique, Support Vector Machine (SVM), to classify the measured data directly; both linear SVM (LSVM) and kernel SVM (KSVM) are included. Then the dimension reduction methods are also applied to extract low-dimension features which contain most information of the data, and we take the Linear Discriminant Analysis (LDA), PCA into consideration, followed by a classifier SVM. Several statistic method, like Probabilistic principal component analysis (PPCA) to model the data with Gaussian latent variable and Hidden Markov Model (HMM) to model the time relations of the data, are also considered here to demonstrate our method's validity. At last, We measured the performance of some neural networks, e.g. Multilayer Perceptron (MLP), Deep Belief Network(DBN) and RNN. The superior recognition performance of our model is shown in Table 2.

Table 2. Classification performance of the proposed model with several traditional methods.

Methods	LSVM	KSVM (RBF)	LDA	PCA	PPCA	HMM	DBN	MLP	RNN with different inputs			WaveletAE with RNN
									$x^{(t)}$	FFT($x^{(t)}$)	DWT($x^{(t)}$) (bior2.4)	W_{xs} (bior.2.4)
ACRR	86.81	89.27	83.88	86.22	89.19	87.05	89.68	87.78	89.89	91.63	93.70	**94.42**

Several observations can be made from Table 2. First, the average correct recognition rates(ACRR) of LSVM here only serves as a simple baseline, thus

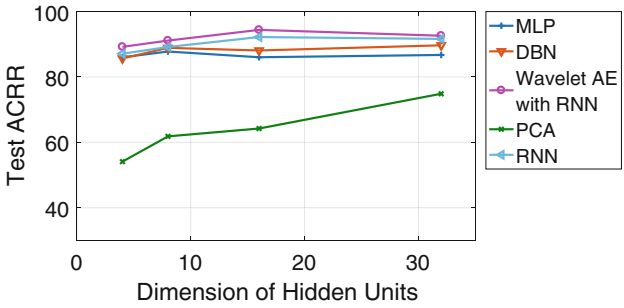


Fig. 4. The ACRRs with dimensionality of feature via different models.

it does not employ any feature extraction. KSVM is the SVM with kernel function which can used for non-linearly separate data. Using radial basis function kernel (RBF), SVM can achieve a higher accuracy. With single HRRP as input, the ACRR of PCA is 86.22%, of which the feature dimension is 77, containing 99% information of the input. While the performance of LDA is inferior, for the feature dimension is C-1, as C is the number of categories, and in our experiment, C is 3. A probabilistic formulation of PCA from a Gaussian latent variable model, called PPCA reaches a desirable result as well as the HMM. But LDA, PCA, and PPCA, which are feature exaction method, should be followed by a classifier. And when we use the SVM as the classifier, we should pick the relatively better one. The process of these methods consist of two stages, feature exaction and classifier selection. And the neural networks, DBN and MLP with a single hidden layer can also obtain effective performance, and the ACRR of RNN is better. However, our model is the most effective and efficient, which can directly predict object categories from input by a joint learning method.

We also plot the ACRRs of the methods varying with the dimensionality of extracted features in Fig. 4. From Fig. 4, we can see that networks works more stable and better than PCA, especially when the dimensionality is small, as the network can capture more discriminative features because of its nonlinear characteristic. And WaveletAE with RNN gives the leading performance, since the model makes use of the time dependency of a single HRRP data. What's more, the model also takes advantages of spectrum features with the decoder weights fixed as the overcomplete bases, which can be jointly trained.

We especially take the experiments on RNN and WaveletAE with RNN as following:

(1) $x^{(t)}$ as the input of the RNN model;
(2) the L_2 norm of FFT on $x^{(t)}$ as the input of RNN model;
(3) DWT with different mother wavelet (dbN, bior$n.d$) on $x^{(t)}$ as the input of RNN model;
(4) $x^{(t)}$ as the input of the proposed model shown in Fig. 2, and the W_{xs} is derived from the mother wavelet same as (3).

Table 3 depicts the ACRRs of the experiments mentioned above. To look insight into the performance of the methods on different targets, we also list the confusion matrix for RNN with different inputs and the proposed model with different W_{sx} (Table 4).

Table 3. Classification performance of the proposed model with RNN model

Methods	RNN with different inputs						WaveletAE with RNN			
	$x^{(t)}$	FFT($x^{(t)}$)	DWT($x^{(t)}$)				W_{xs}			
			bior1.5	bior2.4	db2	db3	bior1.5	bior2.4	db2	db3
Av.R	89.89	91.63	92.47	93.70	91.31	91.88	93.46	**94.42**	93.72	93.56

In Table 3, we show the superior recognition performance of the proposed model. Compared with RNN with different inputs, our method reaches the best result. Although take $x^{(t)}$ as the input of RNN model achieved a good performance 89.89%, FFT on $x^{(t)}$ has a better result, which shows the appealing qualities of FFT mentioned in Sect. 2. Moreover, substituting DFT for FFT increases the classification performance, while the best results achieved by the proposed model. Using the same mother wavelet, the performance of RNN with DWT on $x^{(t)}$ is inferior to the proposed model about $1\% \sim 2\%$.

Table 4. Confusion matrices of different methods.

Methods	RNN with different inputs									WaveletAE with RNN		
	$x^{(t)}$			FFT($x^{(t)}$)			DWT($x^{(t)}$)(bior2.4)			W_{xs}(bior2.4)		
	An-26	Cessna	Yark-42	An-26	Cessna	Yark-42	An-26	Cessna	Yark-42	An-26	Cessna	Yark-42
An-26	90.33	2.39	7.28	96.73	1.72	1.56	93.76	2.13	4.11	94.91	3.07	2.03
Cessna	18.65	81.00	0.35	16.6	80.50	2.90	9.50	89.60	0.90	8.05	91.10	0.85
Yark-42	1.25	0.42	98.33	2.00	0.33	97.67	1.67	0.58	97.75	1.92	0.83	97.25
Av.R	89.89			91.63			92.47			**94.42**		

For RNN, FFT and DWT on $x^{(t)}$ are extra processes of the model, which are two-stage methods, while for the proposed model, it can be jointly trained to predict the categories from input directly. The model has an encoder-decoder layer, in which the weights of decoder are fixed as a set of overcomplete bases derived from the mother wavelet, containing more bases than DWT. Imposing a sparsity constraint on hidden layer $s^{(t)}$, the model can find interesting structure in the data, even if the number of hidden units is large. Table 4 list the confusion matrices for different methods, and only mother wavelet *bior*2.4 is included, since it is the leading one for both RNN and proposed model. It can be concluded that An-26 and Cessna are more separable in the proposed model. Fig. 5 shows W_{sx} and the bases of inverse DWT. It is apparent that W_{sx} contains more bases than IDWT. Fig. 5 also demonstrates the hidden units in the proposed model and DWT on one of the HRRP samples $x^{(t)}$. Obviously, the hidden unites contains

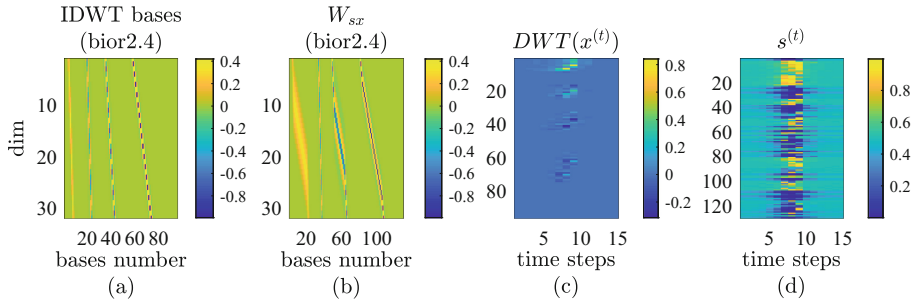


Fig. 5. (a) The inverse DWT bases of bior2.4. (b) The weights of decoder W_{xs}, which is derived from mother wavelet bior2.4. (c) The DWT on input in each time. (d) The hidden units of encoder-decoder layer in each time.

more information than DWT on $x^{(t)}$, for the reason that the bases of DWT is limited.

In summary, the proposed model is a more effective and efficient model than RNN, which takes a set of overcomplete bases into consideration, and obtain better recognition performance for radar HRRP target recognition. Besides, the model can be jointly trained with a specific objective function rather than takes two stages of extracting features and classification.

5 Conclusions

In this paper, a wavelet autoencoder with RNN is analyzed and utilized for radar HRRP target recognition, which extract features by a joint learning method, achieving better recognition performance than the RNN model with different inputs. We fixed the weights of decoder in our model as a set of overcomplete bases, and imposed an sparsity constraint on the hidden units, in order to discover interesting structure in the data, even if the number of hidden units is large. The experimental results on measured HRRP data show that our model achieves superior recognition than other methods.

References

1. Chen, B., Liu, H., Chai, J., Bao, Z.: Large margin feature weighting method via linear programming. IEEE Trans. Knowl. Data Eng. **21**(10), 1475–1488 (2009)
2. Jacobs, S.P.: Automatic target recognition using high-resolution radar range-profiles. Washington University (1997)
3. Feng, B., Du, L., Liu, H., Li, F.: Radar HRRP target recognition based on K-SVD algorithm. In: 2011 IEEE CIE International Conference on Radar (Radar), vol. 1, pp. 642–645. IEEE (2011)
4. Molchanov, P., Egiazarian, K., Astola, J., Totsky, A., Leshchenko, S., Jarabo-Amores, M.P.: Classification of aircraft using micro-doppler bicoherence-based features. IEEE Trans. Aerosp. Electron. Syst. **50**(2), 1455–1467 (2014)

5. Chen, B., Liu, H., Bao, Z.: Analysis of three kinds of classification based on different absolute alignment methods. Xiandai Leida (Mod. Radar) **28**(3), 58–62 (2006)
6. Wang, P., Dai, F., Pan, M., Du, L., Liu, H.: Radar HRRP target recognition in frequency domain based on autoregressive model. In: 2011 IEEE Radar Conference (RADAR), pp. 714–717. IEEE (2011)
7. Zhang, X.-D., Shi, Y., Bao, Z.: A new feature vector using selected bispectra for signal classification with application in radar target recognition. IIEEE Trans. Signal Process. **49**(9), 1875–1885 (2001)
8. Du, L., Liu, H., Bao, Z., Xing, M.: Radar HRRP target recognition based on higher order spectra. IEEE Trans. Signal Process. **53**(7), 2359–2368 (2005)
9. Wang, T., Wu, D.J., Coates, A., Ng, A.Y.: End-to-end text recognition with convolutional neural networks. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 3304–3308. IEEE (2012)
10. Liu, H., Du, L., Wang, P., Pan, M., Bao, Z.: Radar HRRP automatic target recognition: algorithms and applications. In: 2011 IEEE CIE International Conference on Radar (Radar), vol. 1, pp. 14–17. IEEE (2011)
11. Penghui, W., Lan, D., Mian, P., Xuefeng, Z., Hongwei, L.: Radar HRRP target recognition based on linear dynamic model. In: 2011 IEEE CIE International Conference on Radar (Radar), vol. 1, pp. 662–665. IEEE (2011)
12. Pan, M., Lan, D., Wang, P., Liu, H., Bao, Z.: Multi-task hidden Markov modeling of spectrogram feature from radar high-resolution range profiles. EURASIP J. Adv. Signal Process. **2012**(1), 86 (2012)
13. Wang, X., Takaki, S., Yamagishi, J.: A comparative study of the performance of HMM, DNN, and RNN based speech synthesis systems trained on very large speaker-dependent corpora. In: 9th ISCA Speech Synthesis Workshop, vol. 9, pp. 125–128 (2016)
14. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
15. Lee, H., Ekanadham, C., Ng, A.Y.: Sparse deep belief net model for visual area V2. In: Advances in Neural Information Processing Systems, pp. 873–880 (2008)
16. Mikolov, T., Karafiát, M., Burget, L., Černockỳ, J., Khudanpur, S.: Recurrent neural network based language model. In: Eleventh Annual Conference of the International Speech Communication Association (2010)
17. Liu, H., Xu, B., Chen, B., Jin, L.: Attention-based recurrent neural network model for radar high-resolution range profile target recognition. J. Electron. Inf. Technol., December 2016
18. Kim, K.-T., Seo, D.-K., Kim, H.-T.: Efficient radar target recognition using the MUSIC algorithm and invariant features. IEEE Trans. Antennas Propag. **50**(3), 325–337 (2002)
19. Tzanetakis, G., Essl, G., Cook, P.: Audio analysis using the discrete wavelet transform. In: Proceedings of the Conference in Acoustics and Music Theory Applications, vol. 66 (2001)
20. Jie, W., Jianjiang, Z., Jiehao, Z.: A radar target recognition method based on wavelet power spectrum and power offset. In: International Asia Conference on Informatics in Control, Automation and Robotics, pp. 130–134 (2010)
21. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice-Hall Inc., Upper Saddle River (2007)
22. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)

23. Graves, A., Mohamed, A., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6645–6649. IEEE (2013)
24. Ng, A.: Sparse autoencoder (2011)
25. Lan, D., Liu, H., Bao, Z., Zhang, J.: A two-distribution compounded statistical model for radar HRRP target recognition. IEEE Trans. Signal Process. **54**(6), 2226–2238 (2006)
26. Lan, D., Liu, H., Bao, Z.: Radar HRRP statistical recognition: parametric model and model selection. IEEE Trans. Signal Process. **56**(5), 1931–1944 (2008)
27. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)