Signal Processing 196 (2022) 108497

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Target-attentional CNN for Radar Automatic Target Recognition with HRRP

Jian Chen, Lan Du*, Guanbo Guo, Linwei Yin, Di Wei

National Lab of Radar Signal Processing, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 22 July 2021
Revised 1 February 2022
Accepted 6 February 2022
Available online 10 February 2022

Keywords:
Radar target recognition
High-resolution range profile (HRRP)
One-dimensional convolutional neural network (1-D CNN)
Gated recurrent unit (GRU)
attention mechanism

ABSTRACT

In this paper, a target-attentional convolutional neural network (TACNN) combining the convolutional neural network (CNN) and attention mechanism is proposed for radar high-resolution range profile (HRRP) target recognition. The TACNN takes one-dimensional CNN (1-D CNN) as the feature extractor and has the capability to excavate abundant local structural features of data. However, the HRRP contains non-target areas, where the information is useless or even unfavorable. Furthermore, different parts of HRRP target regions should have differences in contribution to the recognition task. Therefore, it is an inadvisable approach that treats all local features alike and directly uses them for the subsequent target recognition, which is adopted by a lot of models, such as the conventional CNN. To tackle this problem, the TACNN introduces the attention mechanism on the basis of 1-D CNN. In detail, the constructed attention module adaptively assigns a weight to each local feature of HRRP so as to locate the target areas and meanwhile enhance the interest of model in valuable target information. Specially, the attention mechanism in TACNN is realized via a bidirectional gated recurrent unit (Bi-GRU) network, where the attention coefficients used for weighting up local features are generated with full consideration of sequential relationship among different regional features in HRRP. Therefore, the learned attention coefficients in our TACNN can better represent the importance of each local feature to the recognition task, ultimately beneficial for the discovery of target information with more discriminability. Experimental results on measured HRRP data show that the proposed model can get more effectiveness in target recognition than related methods.

1. Introduction

In the wideband radar system, a target does not appear as a "point target" anymore but consists of many scatterers [1] [2]. Since the distribution of scatterers is closely related to target structure, the wideband radar echoes contain a wealth of target structure information, thus can be applied to the higher-level and more difficult tasks than the narrowband radar signals. According to difference of signal forms, the target echoes of wideband radar can mainly be divided into two categories: one-dimensional (1-D) high-resolution range profile (HRRP) and two-dimensional (2-D) image, such as synthetic aperture radar (SAR)/ inverse synthetic aperture radar (ISAR) image. In comparison of 2-D SAR/ISAR image, the acquisition of 1-D HRRP doesn't require a long time for coherent accumulation, nor does it need that the rotation angle of target relative to radar should be greater than a certain value, which may not be guaranteed for the non-cooperative target in battlefield environment; moreover, the 1-D HRRP has more superiorities in computational efficiency and storage requirement. Due to the property of easier acquisition and processing, the HRRP has received attention in radar community.

Since the target size is much larger than the range resolution of wideband radar, the target scatterers are distributed in several different range cells, and numbers of scatterers for different range cells are of difference, as shown in Fig. 1. For each range cell, its echo is the coherent summation of complex returns along radar line-of-sight (LOS) from all scatterers located in this range cell. Then the amplitudes of echoes from all range cells can form a real vector, which is denoted as the HRRP. That is, the HRRP is a vector, in which each element represents the amplitude of coherent summation of complex returns from all scatterers in a range cell.

As introduced at the beginning, the HRRP is rich in target structural information, and thus can be used for the tasks with more challenges, such as the high-resolution modeling of small objects [3–5], target detection with high precision [6,7], and target type recognition [8-10,12,14,15,18,20-23,28], etc. In detail, references [3] and [4] build models for HRRP sequences of small-sized ballistic target and space debris with cone shapes, respectively,

* The corresponding at: 2 Taibai Rd., Xi'an, Shaanxi, 710071, China.
E-mail address: dulan@mail.xidian.edu.cn (L. Du).

https://doi.org/10.1016/j.sigpro.2022.108497
0165-1684/

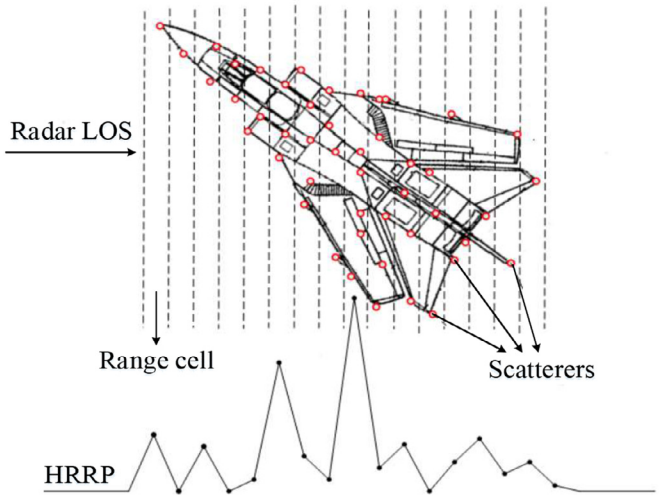

Fig. 1. Illustration of an HRRP sample from a plane target, where the circles on the plane represent the scatterers.

and further evaluate the target's bottom face radius and height. Since the HRRP can reflect more details of small-sized targets, the parameter estimation precisions via the HRRP in [3] and [4] are higher than those obtained via the narrowband radar signals. Liu et al. [6] exploit positive alpha-stable distribution to characterize the distribution of HRRP data from sea clutter, and then obtain the more precise detection threshold for constant false alarm rate (CFAR)-based ship detection, thus leading to a better detection performance. In [9,15,22,28], different models (including local factor analysis, multi-task factor analysis, class factorized complex variational auto-encoder and target-aware recurrent attentional network) are developed to extract more information that can depict essential characteristics of the target from airplane HRRP data, and ultimately realize the recognition of airplane types, such as the recognition of Yark-42 and Cessna Citation S/II both belonging to jet plane. In this paper, we intend to achieve the recognition of target types with HRRP, referred to as the radar HRRP target recognition for convenience.

The basic process of radar HRRP target recognition is summarized in Fig. 2, from which we see that the whole recognition scheme includes two stages, i.e., training stage and test stage. In the training stage, the training HRRPs are firstly pre-processed to overcome the sensitivity problems, such as target-aspect, time-shift and amplitude-scale sensitivities. Then the feature extraction is performed to discover the hidden characteristics of data, followed by a classifier adapted to the extracted features being designed and learned. In the test stage, after pre-processing and feature extraction, the classifier constructed in the training stage can be used to make a prediction to the class attributes of test HRRP samples.

Fig. 2 indicates that the feature extraction plays an important role in HRRP-based radar target recognition and the quality of extracted feature significantly affects the ultimate performance. Therefore, many researchers pay much attention on exploring various feature learning methods. In [10] and [11], the features with physical meaning, such as center of mass, the number of peaks and target size, are extracted for radar HRRP target recognition. In [12], Kim et al. calculate the central moment of HRRP. For the feature extraction in spectral domain, the bispectra feature [13] is investigated and has strong robustness to noise; Du et al. [15] make an analysis on the power spectrum, FFT-magnitude feature, and high-order spectra feature. These HRRP features have the property of convenient calculation but heavily rely on the researchers' prior knowledge to data and specific application scenarios. Moreover, the above methods usually perform feature extraction and classifier learning individually, which suffers the mismatch problem between the feature and backend classifier.

Because of the drawbacks of hand-crafted features, many scholars turn to the study of data-driven feature extraction approaches and hope to automatically learn the good representations from HRRP data. In [16], the principal component analysis (PCA) based feature subspaces are constructed for each target-aspect sector and the test HRRP sample is assigned to the subspace with the minimum reconstruction error. By exploiting the dictionary learning based on K singular value decomposition (K-SVD) algorithm, Feng et al. [17] obtain the sparse embeddings of HRRP. Due to possessing the ability of revealing hidden explanation in data, the factor analysis (FA) model [18–20] is widely used, where several independent latent factors form a low-dimensional feature vector. Nevertheless, the above methods are of shallow structures and fail to learn the deep representations. Thanks to the superior expressive capability, a lot of nonlinear deep networks have been proposed and applied in HRRP-based recognition in recent years. In [21], the stacked corrective autoencoder (SCAE) is utilized to extract the robust features from HRRP by taking average profile as the correction term. Liao et al. [22] construct an improved variational autoencoder (VAE), called class factorized VAE (CFVAE), from which the probabilistic latent features can be learned. Pan et al. [23] utilize the deep belief network (DBN) to learn discriminative features with the t-distributed stochastic neighbor embedding (t-SNE) being adopted for the better segmentation of HRRPs from different target-aspect sectors. In [24], the performance of fully connected network (FCN) is evaluated on HRRP data. Considering the successful application of convolutional network in image processing area, the convolutional neural network (CNN) is applied to the radar HRRP target recognition and the structural features learned from multiple layers are visualized in [25].

Although the deep networks hold tremendous potential in data mining, they use the features acquired from all parts of HRRP indiscriminately to perform target recognition and ignore the discrepancies of different local features in recognition contribution, eventually leading to the limited performance improvement. With this consideration, we blend the attention mechanism [26] into one-dimensional CNN (1-D CNN) where the convolution operation only takes place at range dimension, and develop a novel target-attentional CNN (TACNN) with the purpose of catching the discriminative target areas in HRRP. The attention mechanism in TACNN provides a weight for each local regional HRRP feature learned via the 1-D CNN, to make a measurement of importance degree. Since the HRRP features from non-target areas and those from several target areas without discriminability make no sense to the recognition task, their weights should be much smaller than those corresponding to the features from discriminative target areas, which can achieve the suppression of worthless features and the reinforcement of features that convey the valuable target information. Meanwhile, the learned attention weights in our TACNN are data-dependent rather than fixed for all samples, thus making the model well adaptive to the diversity in positions and sizes of different HRRP supports. More specifically, the TACNN learns the attention weights via a bidirectional gated recurrent unit (Bi-GRU) [27]. The Bi-GRU can take advantage of sequential relationship among range cells within an HRRP and is conducive to accurately locate the discriminative target areas.

As discussed above, different from the traditional neural networks where the feature extraction is "blind" and the extracted feature is unexplainable, the proposed TACNN aims to excavate explainable feature, i.e., the feature from discriminative target areas of HRRP, which is consistent with the thought of explainable artificial intelligence. Similarly to our TACNN, some radar HRRP feature

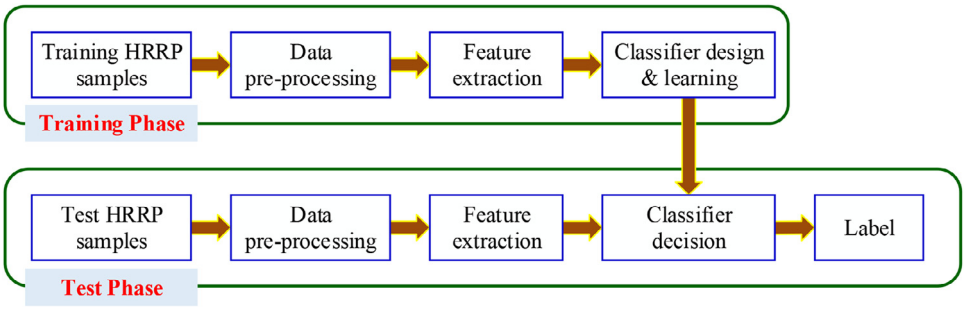

Fig. 2. The flowchart of radar HRRP target recognition.

extraction methods combining the thought of explainable AI have been proposed, such as references [28–30]. Especially for reference [28], an attention mechanism is also introduced to find target-regional features from those of all HRRP local regions, learned via the recurrent neural network (RNN) [31]. Compared with these related approaches, the 1-D CNN in our TACNN performs multiple nonlinear transformations for HRRP, leading to the extraction of deeper and more effective features. Meanwhile, the sliding window operation of multiple convolution kernels on HRRP ensures the diversity of extracted features in each local region of HRRP. However, the feature extraction modules in other related models are of shallow structures (e.g. the RNN used in [28,30]) or lack of diversity (e.g., the deep recurrent gamma belief network in [29]). Thus, the 1-D CNN in our TACNN has more potential to mine the discriminative information from observed HRRPs and further provides a superior feature library for the subsequent location of discriminative target-regional HRRP features. Furthermore, in the mining of discriminative target-regional features based on attention mechanism, the property of 1-D CNN separately extracting features from different HRRP regions avoids the inaccurate learning of attention weights caused by the mixture of target and non-target information. Nevertheless, the RNN based feature extraction methods has the memory characteristic to historical information, i.e., the representations of non-target areas behind HRRP supports will contain the former target information, which misguides the learning of attention weight and easily makes the model focus on the non-target signal, ultimately leading to the real meanings of selected features inconsistent with those we expect. Moreover, the Bi-GRU based attention module in our TACNN takes full account of the relationship among the features of the current, the former and the latter HRRP regions. When the current HRRP local feature is poor resulted from some factors but the former and latter local features are learned well, an accurate attention weight may also be obtained via the Bi-GRU for the current feature, due to the interaction among adjacent local features and the consistency of importance of adjacent local features to recognition task. Nevertheless, the attention mechanism without consideration of correlations among features from different HRRP local regions makes the learning of each attention weight be completely determined by the current HRRP local feature. The bad local feature will directly lead to the wrong attention weight, which cannot be corrected and continues until the end of model learning, such as the attention mechanism achieved by a simple nonlinear transformation in [28]. The capability of Bi-GRU to generate reasonable attention weight can lead to the accurate positioning of discriminative target-regional features, i.e., giving accurate meanings to features from different regions of HRRP. Therefore, the proposed TACNN can obtain the more discriminative HRRP features than the other related methods. Usually, the most discriminative information hidden in HRRP data reflects the target essence, which is largely dependent on the self-structure of target and less affected by target orientation change. As a result, the TACNN should have better robustness to the target orientation change, which is the typical difficulty for most learning-based radar target recognition methods.

In summary, the main contributions of our work cover the following three aspects. 1) We construct a 1-D CNN to fully excavate the structural features hidden in HRRP data, thus offering more opportunities for the final excavation of strongly discriminative information hidden in HRRP. Moreover, due to the property of independent extraction of local features, the 1-D CNN used in our method is more suitable for the integration with attention mechanism to accurately find the meaningful target areas than state-of-the-art methods. 2) The attention module in the proposed method is realized via a Bi-GRU, which can exploit the temporal correlations of HRRP range cells in forward and reverse directions during the learning of attention coefficients, further ensuring the precise attention of the model to HRRP target regions with discriminative information. 3) The capability of TACNN to precisely excavate discriminative target-regional features makes it more robust to the change of target orientation, which is the typical difficulty for the learning-based radar target recognition methods.

In Table 1, we list all notations in our TACNN model to assist readers clearly understand the matrix dimensions and definitions of notations.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce the attention mechanism and explain its necessity in radar HRRP target recognition based on the discussion of HRRP formation principle. Then the detailed description of the proposed TACNN is presented in Section 3. In Section 4, the experiments on measured HRRP data are conducted to validate the effectiveness of our TACNN, and the conclusions are finally summarized in Section 5.

2. Preliminaries

2.1. Problem formulation

According to the scattering center model [32], the target can be represented by a set of scatterers located in several range cells along the radar LOS. Suppose that the radar transmitted signal is $s(t)e^{j2\pi f_c t}$ with $s(t)$ denoting the signal envelope and f_c the radar center frequency. After dechirping and matched filtering, the nth received echo of the pth range cell in baseband, i.e., $x_{pn}(t)$ ($p = 1, \cdots, P$ with P denoting the number of range cells occupied by the target), can be expressed as

$$\begin{aligned} x_{pn}(t) &= \sum_{v=1}^{V_p} \sigma_{pv} s\left(t - \frac{2R_{pnv}}{c}\right) \exp\left\{-j\frac{4\pi f_c}{c}(R_n + r_{pnv})\right\} \\ &= e^{j\psi_n} \sum_{v=1}^{V_p} \sigma_{pv} s\left(t - \frac{2R_{pnv}}{c}\right) e^{j\varphi_{pnv}}, \end{aligned} \tag{1}$$

Table 1
List of notations in the TACNN model.

Notations	Meanings	Notations	Meanings
N	The number of observed HRRP samples	$\mathbf{W}_q^l \in \mathbb{R}^{h^l \times 1 \times Q_{l-1}}$	The qth ($q = 1, \cdots, Q_l$) convolution kernel in the lth ($l = 1, \cdots, L$) convolutional layer which contains Q_{l-1} channels with each channel being a vector of size $h^l \times 1$
C	The number of target categories	$\mathbf{F}_q^l \in \mathbb{R}^{H^l \times 1}$	The qth ($q = 1, \cdots, Q_l$) channel of the output of the lth convolutional layer, which is a H^l – dimensional vector
P	The number of range cells occupied by the target in an HRRP	$\mathbf{F}_{t,q} \in \mathbb{R}^{d \times 1}$	The tth ($t = 1, \cdots, T$) segment of $\mathbf{F}_q^L$ via the sliding window method
D	Dimension/Range cell number of the HRRP ($D > P$)	$\mathbf{FC}_t \in \mathbb{R}^S$	The input of the tth timestep in the Bi-GRU module
L	The number of convolutional layers in the TACNN	$\vec{\mathbf{h}}_t \in \mathbb{R}^M$, $\overleftarrow{\mathbf{h}}_{T-t+1} \in \mathbb{R}^M$	The forward and reverse M – dimensional hidden states outputted by the Bi-GRU, corresponding to the tth timestep, respectively
Q_l	The number of convolution kernels in the lth ($l = 1, \cdots, L$) convolutional layer	$\mathbf{h}_t = [\vec{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_{T-t+1}] \in \mathbb{R}^{2M}$	The hidden state outputted by Bi-GRU at timestep t
d, b	The window length and the stride length used for the CNN feature segmentation, respectively	$\vec{\mathbf{r}}_t \in \mathbb{R}^M$	Reset gate of the forward GRU at timestep t
T	The number of input timesteps for the Bi-GRU	$\vec{\mathbf{z}}_t \in \mathbb{R}^M$	Update gate of the forward GRU at timestep t
S	Input dimension of each timestep in the Bi-GRU module	$\vec{\mathbf{W}}_r \in \mathbb{R}^{M \times (M+S)}$	Projection matrix of the reset gate for the forward GRU
$s(t)$	Envelope of radar transmitted signal	$\vec{\mathbf{W}}_z \in \mathbb{R}^{M \times (M+S)}$	Projection matrix of the update gate for the forward GRU
f_c	Radar center frequency	$\vec{\mathbf{b}}_r$	Bias of the reset gate for the forward GRU
$x_p(n)$	The target echo from the pth ($p = 1, \cdots, P$) range cell in the nth ($n = 1, \cdots, N$) HRRP sample	$\vec{\mathbf{b}}_z$	Bias of the update gate for the forward GRU
σ_{pv}	The strength of the vth scatterer in the pth range cell with $v = 1, \cdots, V_p$ and V_p being the scatterer number in the range cell	a_t	Attention coefficient corresponding to the tth timestep
R_{pnv}	Radial distance between radar and the vth scatterer in the pth range cell of the nth HRRP	$\tilde{\mathbf{F}} \in \mathbb{R}^{H_L \times 1 \times Q_L}$	The weighted HRRP feature
R_n	Radial distance between radar and the target reference center for the nth HRRP	$\mathbf{S} \in \mathbb{R}^C$	Score vector to be fed into the softmax classifier in the recognition module of TACNN
r_{pnv}	Radial distance between target center and the vth scatterer in the pth range cell for the nth HRRP	$\hat{\mathbf{y}} \in \mathbb{R}^C$	The predicted label of training HRRP samples
$\mathbf{x}_s(n)$	The target signal in the nth HRRP	$\mathbf{y} \in \mathbb{R}^C$	The true label of training HRRPs
$\mathbf{x} \in \mathbb{R}^D$	The observed HRRP sample	$\mathbf{y}^* \in \mathbb{R}^C$	The predicted label of test HRRPs

where σ_{pv} represents the strength of the vth scatterer in the pth range cell with $v = 1, \cdots, V_p$ and V_p denoting the number of scatterers in the range cell; R_{pnv}, R_n, and r_{pnv} represent the radial distances between radar and the vth scatterer in the pth range cell of the nth echo, radar and the target reference center for the nth echo, and target center and the vth scatterer in the pth range cell for the nth echo, respectively; c denotes the speed of light, $\psi_n = -\frac{4\pi f_c}{c} R_n$ and $\varphi_{pnv} = -\frac{4\pi f_c}{c} r_{pnv}$ are the initial phase and remaining phase of the vth scatterer in the pth range cell for the nth echo. Given that the radial displacement of all scatterers in each range cell is approximately unvaried, the $x_{pn}(t)$ can then be approximated as

$$x_{pn}(t) \approx x_p(n) = e^{j\psi_n} \sum_{v=1}^{V_p} \sigma_{pv} e^{j\varphi_{pnv}}. \tag{2}$$

And for the target echo in the nth HRRP sample $\mathbf{x}_s(n)$, it is represented as $\mathbf{x}_s(n) = [|x_1(n)|, |x_2(n)|, \cdots, |x_P(n)|]^{\mathrm{T}}$with

$$|x_p(n)| = \sqrt{\left(\sum_{v=1}^{V_p} \sigma_{pv} \sin \varphi_{pnv}\right)^2 + \left(\sum_{v=1}^{V_p} \sigma_{pv} \cos \varphi_{pnv}\right)^2}, \tag{3}$$

$|\cdot|$ denoting the modular arithmetic, and $[\cdot]^{\mathrm{T}}$ denoting the transposition operation.

For different targets, their structure differences lead to the discrepancies in scatterer's magnitude (i.e., σ_{pv} with $v = 1, \cdots, V_p$ and $p = 1, \cdots, P$) and distribution (affecting the value of φ_{pnv}), finally manifested by the distinctions of HRRPs. Thus, the HRRP can be exploited to identify the category of unknown targets. However, the HRRP used in practice is intercepted from the received radar echo signal by using a distance window with its length larger than the number of range cells the target possessing, due to the difficulty in determinations of starting and terminal positions for the target area [33]. Therefore, the HRRP includes not only the target echo $\mathbf{x}_s(n)$ but also a certain redundant information as shown in Fig. 3.

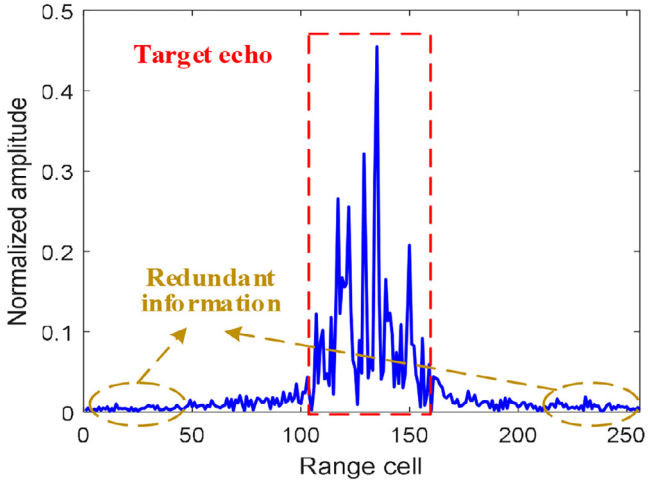

Fig. 3. An HRRP example from the airplane Yark-42, with partial redundant information from the non-target area and a part of target echo from the target area being marked with the ellipses and rectangle, respectively.

There is a broad consensus on the fact that the redundant information has no use for the recognition task. Even worse, the HRRP non-target areas will display in the form of noise when the observed target is far away from radar, and in this case, the information in non-target regions is harmful to target recognition. In

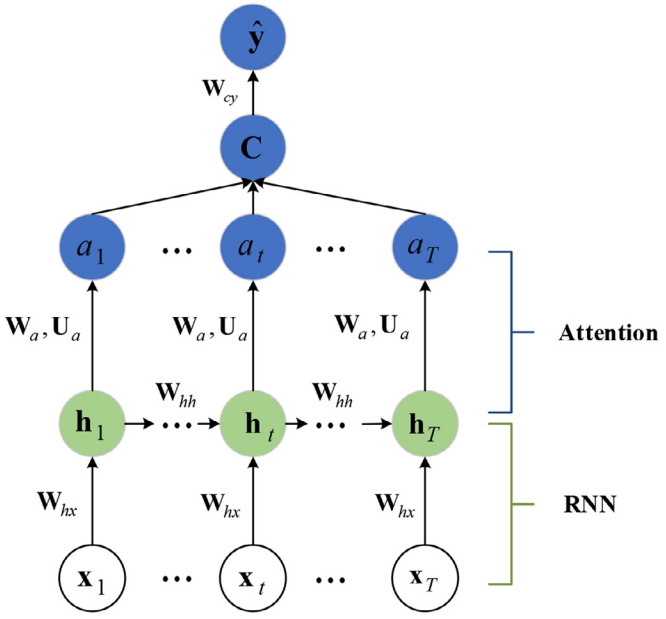

Fig. 4. Graphical representation of the TARAN.

addition, the target echoes in some range cells may have little discriminative information that can be utilized to identify different targets, and thus one should also inhibit them to reduce the classifier burden. Nevertheless, the recent methods, such as SCAE, CF-VAE, DBN, etc., treat equally the information in any area of HRRP, thus often causing the target to distinguish with difficulty. Therefore, our work aiming to construct a target-attentional model is of great significance in the improvement of recognition performance.

2.2. Attention mechanism and TARAN

Attention mechanism comes from the study of human vision, which selectively pays attention to partial information and ignores other visible information. The attention mechanism in machine learning is similar to that of human beings, with the core goal being to select the information more critical to the current task within such a huge amount message. In [34], the attention mechanism is firstly applied to the machine translation, and afterwards, it appears in many tasks of image processing [35], speech recognition [36], and nature language processing [37].

Considering that HRRP consists of non-target area, indiscriminative target area and discriminative target area with only the last one being valuable to recognition task, the attention mechanism in radar HRRP target recognition is mainly used for the location of features from discriminative target areas. Thus the reasonable attention should possess two characteristics. The first one is that the attention mechanism should only focus on the excavation of features from a part of target areas. The second one is that the features processed by the reasonable attention mechanism should have more inter-class differences than unprocessed features and features processed by unreasonable attention mechanism.

To pick out the discriminative target regional features, the TARAN integrating the RNN with attention mechanism is proposed with the graphical illustration being shown in Fig.4. Suppose that the observed HRRP is represented as $\mathbf{x} \in \mathbb{R}^D$ with $\mathbb{R}^D$ denoting the D-dimensional real vector space. It is composed of target echoes, i.e., $\mathbf{x}_s$ in Eq. (3), and noise. The TARAN partitions each HRRP sample $\mathbf{x}$ into several segments $\{\mathbf{x}_t\}_{t=1}^T$ (T denotes the number of data segments) with the same length, each of which is successively used as the input at a timestep of the RNN. Assuming that the dimension of each HRRP segment is d ($d < D$), i.e., $\mathbf{x}_t \in \mathbb{R}^d$. For the tth sequential input $\mathbf{x}_t$, it is encoded by the RNN as

$$\mathbf{h}_1 = f(\mathbf{W}_{hx}\mathbf{x}_1), \mathbf{h}_t = f(\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1}), t = 2, \cdots, T \tag{4}$$

where $\mathbf{h}_t \in \mathbb{R}^M$ is the latent feature of $\mathbf{x}_t$, $\mathbf{W}_{hx} \in \mathbb{R}^{M\times d}$ is the encoding matrix, $\mathbf{W}_{hh} \in \mathbb{R}^{M\times M}$ is the connection matrix between two features from adjacent timesteps, and $f(\cdot)$ refers to the nonlinear transformation, e.g., sigmoid function. From Eq. (4), we know that the encoding matrix $\mathbf{W}_{hx}$ is shared among all HRRP segments. Therefore, the lengths (dimensions) of different HRRP segments, corresponding to the column number of $\mathbf{W}_{hx}$, should be guaranteed to be the same during the segmentation of HRRP, and the similar reason for the equi-length latent features from different HRRP segments. We also see that the acquisition of hidden feature $\mathbf{h}_t$ in Eq. (4) is related to not only the input at the current timestep $\mathbf{x}_t$ but also the previous feature $\mathbf{h}_{t-1}$. This information transmission gives a consideration to the temporal dependence among different timestep inputs, enabling the RNN to learn the more discriminative representations. Instead of directly projecting the features to the sample label adopted in the traditional RNN [38], the TARAN learns a weight to measure the importance of information from different timesteps by imposing a nonlinear transformation on the latent features. Then the weighted summation of all timestep features is treated as the final representation of an HRRP used for the prediction of sample label. In detail, the above process can be mathematically expressed as

$$\begin{aligned} &e_t = \mathbf{U}_a \tanh(\mathbf{W}_a\mathbf{h}_t), a_t = \frac{e_t}{\sum_{t=1}^T e_t}, \mathbf{C} = \sum_{t=1}^T a_t\mathbf{h}_t, \\ &\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}_{cy}\mathbf{C}), \end{aligned} \tag{5}$$

where $\mathbf{W}_a \in \mathbb{R}^{V\times M}$ and $\mathbf{U}_a \in \mathbb{R}^{1\times V}$ are the transformation matrices in the attention mechanism, a_t ($t = 1, \cdots, T$) represents the attention coefficient (weight) of feature from the tth HRRP segment, and $\mathbf{C} \in \mathbb{R}^M$ is the weighted summation of all features and considered as final feature of the model input to the softmax classifier. $\hat{\mathbf{y}} \in \mathbb{R}^C$ denotes the predicted label of the HRRP $\mathbf{x}$ with C denoting the number of targets (classes), which is obtained by first linearly transforming $\mathbf{C}$ with transformation matrix $\mathbf{W}_{cy}$ and then performing nonlinear transformation for each element of $\mathbf{W}_{cy}\mathbf{C}$ with softmax function. Therefore, we have

$$\mathbf{W}_{cy} \in \mathbb{R}^{C\times M} = \begin{bmatrix} w_{11}^{cy} & w_{12}^{cy} & \cdots & w_{1M}^{cy} \\ w_{21}^{cy} & w_{22}^{cy} & \cdots & w_{2M}^{cy} \\ \vdots & \vdots & \vdots & \vdots \\ w_{C1}^{cy} & w_{C2}^{cy} & \cdots & w_{CM}^{cy} \end{bmatrix}, \tag{6}$$

and all elements in the linear transformation matrix $\mathbf{W}_{cy}$ can be estimated after the model learning.

From Eq. (5) we know that the normalized attention coefficient has the ability to further increase the "contrast" between the small-valued feature (e.g., the feature from non-target area) and large-valued feature (e.g., the feature from target echo), with the result that the features more valuable to target recognition can play a more prominent role. Notwithstanding, the learning of attention coefficient heavily relies on the latent features. For TARAN, the features from HRRP non-target areas at the posterior timesteps also contain the prior target information due to the information delivering property of RNN, thus interfering the judgment of attention mechanism to the feature importance and making the representation of HRRP still includes a wealth of redundant information. In this paper, we attempt to construct a more reasonable attention feature extraction network to extract more precise target information from HRRP.

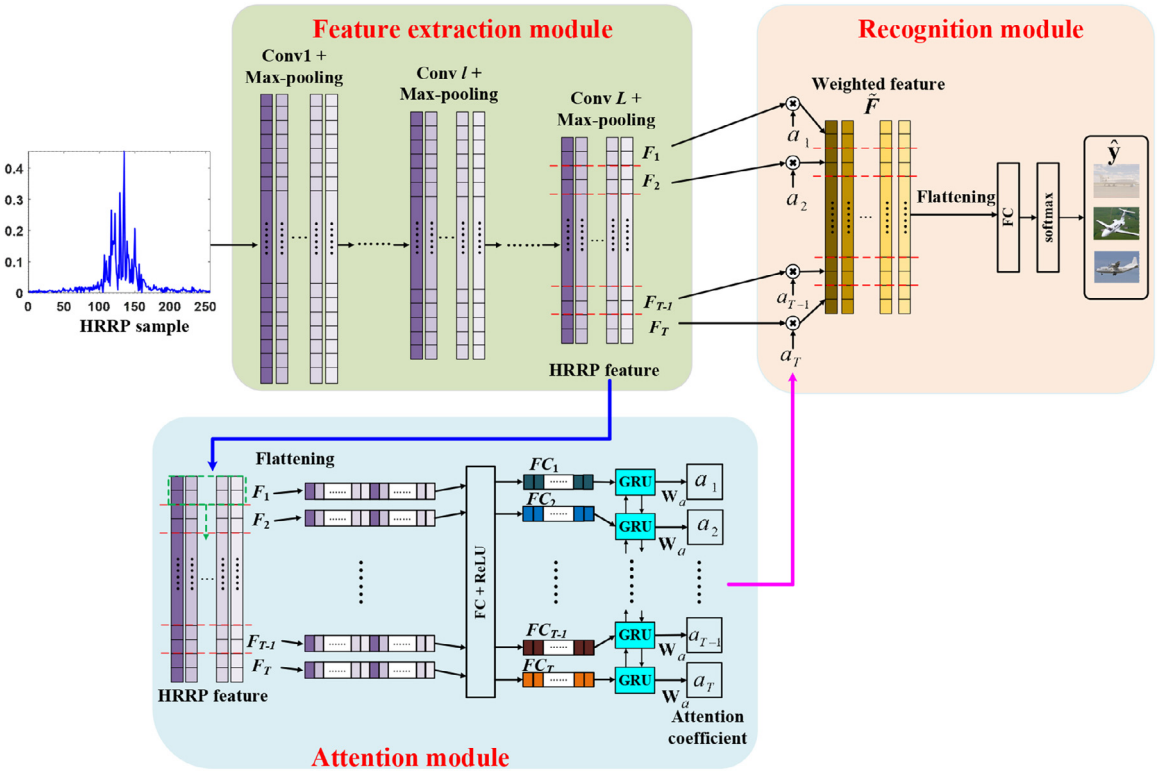

Fig. 5. Architecture of the TACNN. In the feature extraction module, the HRRP feature is obtained via a 1-D CNN, where each column represents a feature vector obtained by performing convolution operation at range dimension with a convolution kernel. Then the HRRP feature is divided into several segments by sliding window and different segments characterize the features from different areas of HRRP. After dimensionality reduction via a fully connected network, each local regional feature is treated as a timestep input of the Bi-GRU, which outputs the corresponding attention coefficient. At last, the recognition module weights up each local feature and feeds the weighted feature to the softmax classifier to make a class prediction for the HRRP.

3. Target-attentional CNN

As discussed in Section 2.1, the HRRP includes the non-target information other than the target signal. The majority of research efforts neglect their importance difference in terms of recognition task. In addition, the different range cells from target area corresponds to the different parts of target and should has the distinct contribution levels to recognition. Considering these issues, we design an attentional convolutional network, i.e., TACNN, which assigns each local regional feature of HRRP a weight in a data-driven way to accentuate useful target areas.

The overview of our TACNN is shown in Fig. 5. Apparently the TACNN consists of three components: 1) a feature extraction module constructed by a 1-D CNN to acquire the rich features separately from different parts of HRRP, 2) an attention module realized via a Bi-GRU intending for accurately expressing the importance of each local reginal feature, and *3)* a recognition module which outputs the label based on the weighted hidden features. The detailed introductions to the proposed model are given as follows.

3.1. Feature extraction module

We take advantage of the 1-D CNN to perform representation learning of HRRP due to its powerful capability in mining abundant local structural features of data. Meanwhile, the 1-D CNN utilizes the small-scale convolution kernel to slide on the input, to which the response in a local receptive field is irrelevant to the information from input's other regions, thus can avoid the problem that the feature in HRRP non-target area contains the target information, eventually affecting the accurate learning of attention coefficient.

The 1-D CNN is composed of multiple layers based on convolution and subsampling operations. Suppose that a 1-D CNN built for the HRRP contains L convolutional layers and layer l ($l = 1, \cdots, L$) includes Q_l convolution kernels (filter). For the qth ($q = 1, \cdots, Q_l$) convolution kernel in the layer l, denoted as $\mathbf{W}_q^l$, it has Q_{l-1} ($Q_0 = 1$) channels, each of which is a vector of size $h^l \times 1$ with h^l representing the vector dimension, i.e., $\mathbf{W}_q^l \in \mathbb{R}^{h^l \times 1 \times Q_{l-1}}$. Let $\boldsymbol{F}^{l-1} \in \mathbb{R}^{H^{l-1} \times 1 \times Q_{l-1}}$ ($l = 1, \cdots, L$) represent the input of the lth convolutional layer (i.e., the output of layer $l-1$, and $\boldsymbol{F}^0 \in \mathbb{R}^D$ is the HRRP sample) with Q_{l-1} being its channel number and H^{l-1} the vector dimension of each channel. The convolutional result of $\mathbf{W}_q^l$ and $\boldsymbol{F}^{l-1}$, denoted as $\boldsymbol{F}_q^l \in \mathbb{R}^{H^l \times 1}$, can be calculated as

$$\boldsymbol{F}_q^l = f\left(\boldsymbol{F}^{l-1} * \mathbf{W}_q^l + \mathbf{b}_q^l\right), \tag{7}$$

where $*$ denotes the convolution operator, $\mathbf{b}_q^l \in \mathbb{R}^{H^l \times 1}$ denotes the bias, and $f(\cdot)$ a nonlinear active function (ReLU [39] is used in this paper), respectively. Then final output of layer l, i.e., $\boldsymbol{F}^l$, is the stack of those from all channels, i.e., $\boldsymbol{F}^l = \{\boldsymbol{F}_q^l\}_{q=1}^{Q_l}$, with its size being $H^l \times 1 \times Q_l$. For multi-channel $\boldsymbol{F}^{l-1}$ and $\mathbf{W}_q^l$, their convolution, i.e., $\boldsymbol{F}^{l-1} * \mathbf{W}_q^l$, is the convolution summation of all channels and the detailed calculation approach is shown in Appendix A.

To increase the content covered by a convolution kernel and the sparsity of the hidden features, a subsampling layer is usually embedded after the convolution computation. In this paper, we adopt the max-pooling operation, which removes all elements other than the maximum one in the pooling region for each feature vector of each convolutional layer.

The feature extraction module of our TACNN makes full use of the CNN's merits and is capable of distilling various features for each part of the HRRP, thus laying the foundation for the superior

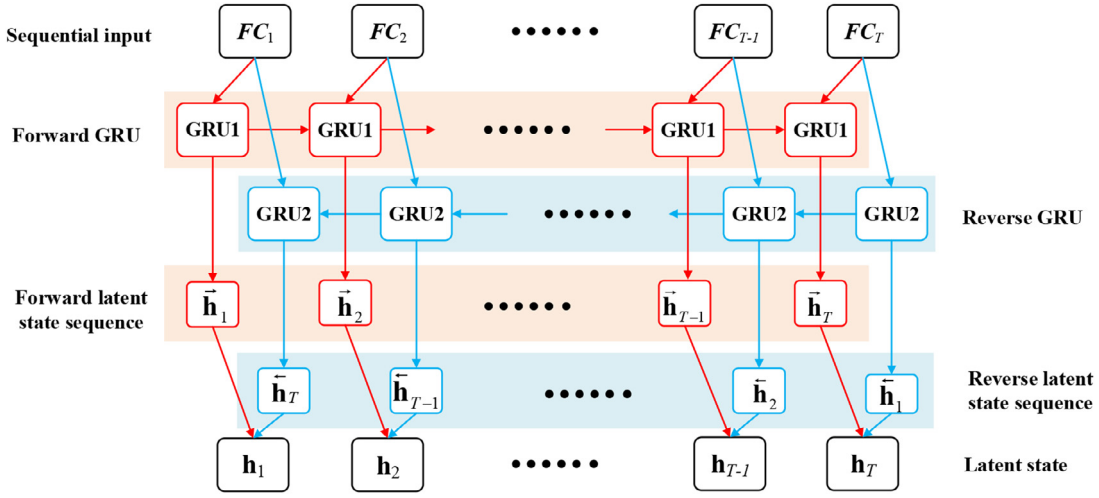

Fig. 6. Structure of the Bi-GRU. The Bi-GRU includes two separate GRUs, which input the sequence in forward and reverse directions, respectively, and then output the forward and reverse latent state sequences, followed by the concatenation of two latent states corresponding to the same timestep input in Bi-GRU.

recognition performance. In addition, the weight sharing and local connection of CNN make it better than the FCNs in those application scenarios hard to collect sufficient training samples, such as the radar target recognition field where the radar sampling rate is often limited.

3.2. Attention module

In this module, as shown in Fig. 5, the HRRP feature extracted by 1-D CNN is firstly divided into several segments via the sliding window technique. Let d represent the window length and b the stride length. The HRRP feature $\boldsymbol{F}^L \in \mathbb{R}^{H^L \times 1 \times Q_L}$ can then be partitioned into several segments, and for the qth channel of the tth segment $\boldsymbol{F}_{t,q}$, it can be represented as

$$\boldsymbol{F}_{t,q} = \boldsymbol{F}_q^L[(t-1)\times b+1 : (t-1)\times b+d], \tag{8}$$

where $q = 1, \cdots, Q_L$, $t = 1, \cdots, T$ with $T = \frac{H^L - d}{b}$, $\boldsymbol{F}_q^L \in \mathbb{R}^{H^L \times 1}$ denotes the qth channel of $\boldsymbol{F}^L$, and $\boldsymbol{F}_t \in \mathbb{R}^{d \times 1 \times Q_L} = \{\boldsymbol{F}_{t,q}\}_{q=1}^{Q_L}$ denotes the tth segment of $\boldsymbol{F}^L$. After feature partition, $\boldsymbol{F}_t$ is firstly flattened as a vector along the channel dimension and subsequently fed into a FCN to perform dimensionality reduction, with its output being denoted as $\boldsymbol{FC}_t \in \mathbb{R}^S$ ($S \ll d \times Q_L$).

For each $\boldsymbol{FC}_t$ ($t = 1, \cdots, T$), it is the abstract representation of echoes from a section of HRRP range cells. As discussed above, the range cell echoes in the support of HRRP reflect the target structural information and are useful for the recognition task; conversely, those from the non-target areas are harmful and should be discarded. To this end, we resort to a Bi-GRU network to help the model focus on the discriminative target areas by assigning the larger weights to the features from target echoes. The structure of Bi-GRU is presented in Fig. 6. We see that the Bi-GRU includes two separate GRUs, which take the forward sequence $\{\boldsymbol{FC}_1, \boldsymbol{FC}_2, \cdots, \boldsymbol{FC}_T\}$ and reverse sequence $\{\boldsymbol{FC}_T, \boldsymbol{FC}_{T-1}, \cdots, \boldsymbol{FC}_1\}$ as the inputs, respectively, and output the forward and reverse latent state sequences, represented as $\{\vec{\mathbf{h}}_1, \vec{\mathbf{h}}_2, \cdots, \vec{\mathbf{h}}_T\}$ and $\{\overleftarrow{\mathbf{h}}_1, \overleftarrow{\mathbf{h}}_2, \cdots, \overleftarrow{\mathbf{h}}_T\}$. Then for the input of Bi-GRU at timestep t, i.e., $\boldsymbol{FC}_t$, its hidden state $\mathbf{h}_t$ is the concatenation of forward and reverse latent states, which are the outputs of two GRUs at timesteps t and $T-t+1$, i.e., $\mathbf{h}_t = [\vec{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_{T-t+1}]$.

Similarly to RNN, besides the current input itself, the output hidden state for the tth timestep in each GRU of Bi-GRU is also related to the hidden state from the previous timestep, as shown in Fig. 6. The calculation of hidden state for each GRU of Bi-GRU is provided in Appendix B. Eventually, for the output of Bi-GRU at timestep t ($t = 1, \cdots, T$), its corresponding attention coefficient can be obtained via the expression of

$$a'_t = \mathbf{W}_a \mathbf{h}_t,\ a_t = \exp\left(a'_t\right) / \sum_{t=1}^{T} \exp\left(a'_t\right), \tag{9}$$

with $\mathbf{W}_a \in \mathbb{R}^{1 \times 2M}$ being the transformation matrix. From Fig. 6 and Eq. (9) we know that the attention coefficient at the current timestep is related to the inputs at the current, the former and the latter timesteps.

As shown in Fig. 6, the Bi-GRU takes full account of the forward and reverse temporal correlations among features from different local regions of HRRP during the calculation of attention coefficients, with which the more reasonable attention coefficients can obtained, ultimately leading to the accurate excavation of discriminative information from HRRP's target areas. In detail, on the one hand, the attention coefficient's calculation via the Bi-GRU for each HRRP local feature not only considers the current HRRP local feature, but also the former and latter HRRP local features. Therefore, although there is not a good learning for the current HRRP local feature due to some factors, an accurate attention weight may also be obtained via the Bi-GRU based attention mechanism, when its previous and posterior local features are learned well (this result can be guaranteed since the extractions of different HRRP local features are independent to each other), because of the interaction among adjacent local features during the calculation of attention coefficient and the consistency of importance of adjacent local features to recognition task. Thanks to the joint optimization of 1-D CNN and Bi-GRU, the accurate attention coefficient further prompts the learning of corresponding local feature in a right direction in the subsequently iterative updating. Compared with the attention mechanism without consideration of temporal correlations among local features, such as the conventional attention approach used in TARAN where the attention weight accuracy is completely dependent on the quality of current local feature, the Bi-GRU based attention mechanism is more likely to ultimately learn the good HRRP local features and accurate attention coefficients. On the other hand, the Bi-GRU considers the bi-directional temporal characteristics among different HRRP local features and can effectively avoid the attention of model to the non-target areas behind the support of HRRP, which often occurs in the unidirectional GRU. The attention coefficient of each HRRP local feature is calculated based on the corresponding hidden state outputted by the Bi-GRU, as shown in Eq. (9). For the hidden state of non-target area behind the support of the HRRP in Bi-GRU, it consists of

not only the target and non-target information carried by the previous states, but also the pure non-target information carried by posterior states. This approach can reduce the proportion of target discriminative information in the hidden state and thus promote the model to yield a tiny weight. However, the unidirectional GRU only takes account of the former local features when it learns the current hidden state, and thus its learned hidden states from non-target areas behind the HRRP support contain the prior target information with larger proportion, which easily makes the model pay attention to the non-target areas. Therefore, the Bi-GRU can better strengthen the information from target areas and suppress the information from non-target areas.

In summary, the Bi-GRU based attention mechanism has the stronger ability to accurately locate the local HRRP features from the discriminative target areas than the conventional and unidirectional GRU based attention mechanisms, thus being more potential to acquire the higher recognition accuracy.

3.3. Recognition module

This module aims to make a prediction to the class membership of the weighted feature of an HRRP. If the weighted HRRP feature is denoted as $\tilde{\boldsymbol{F}}$, we then have $\tilde{\boldsymbol{F}} \in \mathbb{R}^{H_L \times 1 \times Q_L} = \{\tilde{\boldsymbol{F}}_q\}_{q=1}^{Q_L}$ and $\tilde{\boldsymbol{F}}_q \in \mathbb{R}^{H_L \times 1} = [\tilde{\boldsymbol{F}}_{1,q}, \cdots, \tilde{\boldsymbol{F}}_{T,q}]^{\mathrm{T}}$, where Q_L represents the channel number of $\tilde{\boldsymbol{F}}$, H_L represents the feature dimension of each channel, $\tilde{\boldsymbol{F}}_q$ is the qth feature vector in $\tilde{\boldsymbol{F}}$ with it consisting of T components and the tth subvector feature $\tilde{\boldsymbol{F}}_{t,q} \in \mathbb{R}^{d \times 1 \times Q_L}$ being expressed as

$$\tilde{\boldsymbol{F}}_{t,q} = a_t \boldsymbol{F}_{t,q}. \tag{10}$$

After the flattening operation for $\tilde{\boldsymbol{F}}$ along the channel dimension, the flattened feature is linearly transformed to performs dimension reduction and then a score vector $\mathbf{S} \in \mathbb{R}^C$ is outputted with C denoting the number of target classes. Finally, the predicted label vector $\hat{\mathbf{y}} \in \mathbb{R}^C$ can be obtained via the softmax normalization as follows:

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{S}) = \left[\frac{\exp(S_1)}{\sum_{c=1}^{C} \exp(S_c)}, \cdots, \frac{\exp(S_C)}{\sum_{c=1}^{C} \exp(S_c)} \right]^{\mathrm{T}}, \tag{11}$$

where S_c represents the cth element in $\mathbf{S}$.

As shown in Fig. 2, the radar HRRP target recognition includes training and test stages. For our TACNN, we utilize the training HRRP samples to learn the model parameters in the training stage, by minimizing the cross entropy defined as

$$L(\theta) = -\sum_{n=1}^{N} \sum_{c=1}^{C} y_{n,c} \log \hat{y}_{n,c}, \tag{12}$$

where $\hat{y}_{n,c}$ and $y_{n,c}$ represent the cth element in the predicted and true labels for the nth ($n = 1, \cdots, N$ with N denoting the sample number) HRRP sample, respectively; θ represents the model parameters embedded in $\hat{y}_{n,c}$, including the convolution kernels and biases from the 1-D CNN module, transformation matrices of the FCN and Bi-GRU from the attention module, and transformation matrix from the recognition module. In the model learning, the parameters are iteratively updated via the stochastic gradient descent (SGD) algorithm, which can be expressed as $\theta' = \theta - \varepsilon \nabla_\theta L(\theta)$, with θ' denoting the updated model parameter, ε the learning rate, and $\nabla_\theta L(\theta)$ the gradient of the loss function with respect to θ. For more clarity, the overall training procedure is summarized in Algorithm 1. In the test stage, the class label of the unknown HRRP, denoted as $\mathbf{y}^* = [y_1^*, \cdots, y_c^*, \cdots, y_C^*]^{\mathrm{T}}$, can be directly acquired by going through the feedforward propagation once the TACNN training ends. Then the test sample can be assigned to the c^*th class via the expression of $c^* = \arg\max_c \{y_1^*, \cdots, y_c^*, \cdots, y_C^*\}$.

Algorithm 1
The training procedure of the TACNN.

1: Initialize the model parameter θ, including the convolution kernels and biases from the 1-D CNN module, transformation matrices of the FCN and Bi-GRU from the attention module, and transformation matrix from the recognition module. Initialize the learning rate ε, numbers of the epochs I, and batch size N_b with the batch number can be calculated as N/N_b.
2: **for** the number of epochs **do**
3: **for** the number of batches **do**
4: Sample a batch of N_b samples from HRRP training dataset, and then find their corresponding class labels.
5: Get the HRRP feature $\boldsymbol{F}$ according to Eq. (7).
6: Get the attention coefficients $\{a_1, \cdots, a_T\}$ according to Eq. (8)∼Eq. (9).
7: Compute the weighted feature $\tilde{\boldsymbol{F}}$ according to Eq. (10) and then obtain the predicted labels $\hat{\mathbf{y}}$ according to Eq. (11).
8: Bring $\hat{\mathbf{y}}$ into Eq. (12) to get the loss function $L(\theta)$.
9: Update the parameters of the TACNN by SGD on the loss function $L(\theta)$.
10: **end for**
11: **end for**

4. Experiments

4.1. Data introduction and experimental setup

This paper constructs a deep model, i.e., TACNN, for radar target HRRP data to realize the recognition of target types. To validate the effectiveness of the proposed method, we apply the TACNN model to data from three real airplanes, including Yark-42, Cessna Citation S/II, and An-26. In detail, the data of three airplanes are measured by inverse synthetic aperture (ISAR) C-band radar on the ground, which transmits pulse signal with linear frequency modulation. In detail, the radar center frequency is 5520 MHz, the pulse repetition frequency is 400 Hz, the signal bandwidth is 400 MHz, and thus the resolution is 0.375 m (the resolution ΔR can be calculated as $\Delta R = c/2B$ with c and B denoting the speed of light and signal bandwidth, respectively). In addition, the sampling frequency of signal after dechirping is 10 MHz. For intuition, we present the parameters of radar and targets in Table 2. As shown in Table 2, the sizes of targets used in the experiment are much larger than the radar range resolution. Therefore, each airplane contains a large number of range cells, and each echo signal of the airplane contains multiple fluctuating peaks formed by echoes from different range cells, as shown in Fig. 7, rather than the single peak in the low range-resolution radar signal. The fluctuation characteristic of peaks in target echo can fully represent the structural feature of observed airplane. Consequently, the range resolution of radar used in the experiment (i.e., 0.375 m) is high enough to obtain HRRPs rich in structural information of the airplanes, with which the recognition of airplane types can be achieved.

During HRRP data measurement, three types of airplanes are in stable flight state (i.e., there are almost no head lowering, head raising, left tilt, right tilt and other actions for each target), and the ways to flight are of cooperation. For example, the flight paths contain “circles”, so as to ensure that the radar can obtain as much HRRP data from different target attitudes as possible and then realize the adequate learning for each target, avoiding the effect of insufficient target HRRP data on the judgment of model recognition performance. Therefore, in summary, this section aims to exploit airplane HRRP data, obtained by the wideband radar with high range resolution on the ground, to realize the type recognition of airplane targets in flight, based on the TACNN model.

Since the proposed TACNN is a learning mechanism-based recognition method, we should first divide all obtained HRRP echoes into training HRRP set for model learning and test HRRP set for model performance assessment. Due to the attitude change of target and sensitivity of HRRP to target attitude, the target attitudes in the training stage should be as diverse as possible and

Table 2
Parameters of three airplanes and radar in the experiment.

Radar parameters	Center frequency	5520*MHz*	
	Pulse repetition frequency	400*Hz*	
	Bandwidth	400*MHz*	
	Resolution	0.375*m*	
	Sampling frequency after dechirping	10 *MHz*	
Airplanes	**Length (*m*)**	**Width (*m*)**	**Height (*m*)**
Yark-42	36.38	34.88*m*	9.83*m*
Cessna Citation S/II	14.40	15.90*m*	4.57*m*
An-26	23.80	29.20*m*	9.83*m*

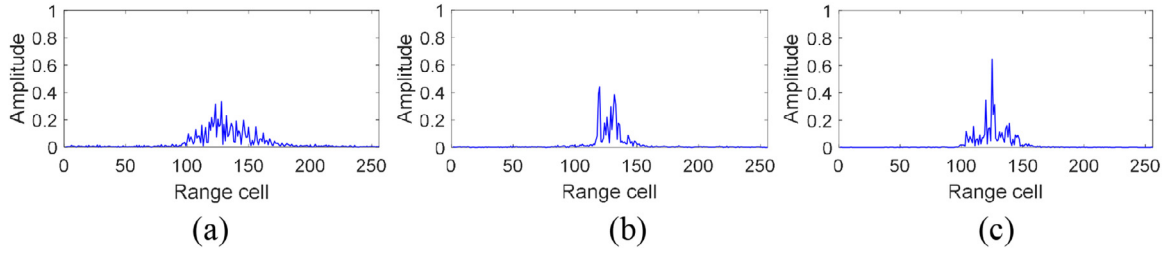

Fig. 7. HRRP examples from different airplane targets. (a), (b), and (c) are the HRRP samples from Yark-42, Cessna Citation S/II, and An-26, respectively.

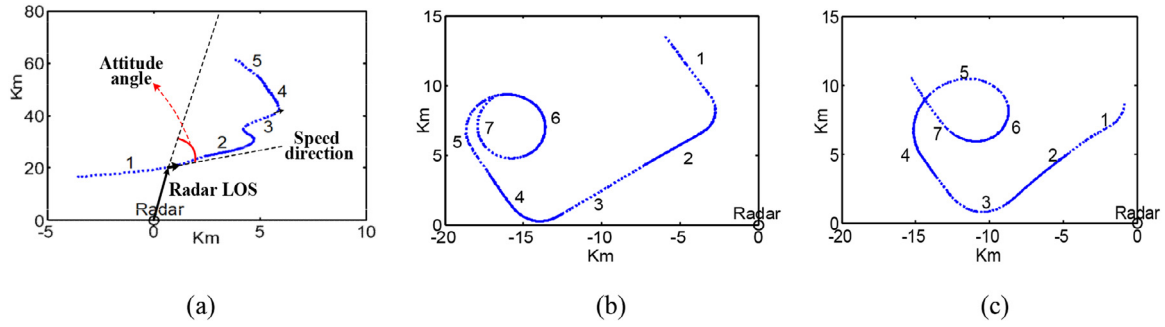

Fig. 8. Projections of airplane trajectories onto the ground. For each target, its trajectory is divided into different segments, several of which are selected as the training data segments, and others as the test data segments. (a) presents the two-dimensional track chart of Yark-42 and the calculation approach of target attitude. (b) and (c) are the two-dimensional track charts of Cessna Citation S/II, and An-26, respectively.

almost cover all those in the test stage to realize the full learning of target, further mitigating the influence of target attitude sensitivity on the judgment of model performance. For the flat airplane target in stable flight state, the change of pitch angle has little effect on the formulation of HRRP echoes, and thus the target attitude change of each airplane relative to radar mainly embodies as the change of included angle between the directions of radar LOS and airplane centerline. At any target location, the direction of radar LOS is the direction that the radar points to the target position. In the stable flight, the direction of target centerline is consistent with that of the target speed, and moreover, the target speed direction at one location can be approximated by the direction that the current target location points to the location of next moment. Then according to the specific coordinate value of target location in radar coordinate, the attitude of target can be estimated via the cosine theorem. To clearly show the calculation approach of target attitude, we plot the projections of three airplane trajectories onto the ground in Fig. 8, where the coordinates of each target in radar coordinate are given and the attitude calculation of target can be shown. Based on the estimated target attitudes, we can select reasonable HRRP echoes from the whole HRRP dataset as the training set and the rest of HRRP echoes as the test set for each target, according to above mentioned division criterion of training and test data. In detail, with target attitude computation, 35,840 HRRPs from the 2th and 5th segments of Yark-42, 51,200 HRRPs from the 6th and 7th segments of Cessna Citation S/II, 51,200 HRRPs from the 5th and 6th segments of An-26 are selected as the training data, and other HRRPs as the test data. For each HRRP, its size is of 256×1, i.e., $C = 3$ and $D = 256$. Moreover, from the two-dimensional track charts of Cessna Citation S/II and An-26, we see that both the 6th, 7th segments of Cessna Citation S/II and 5th, 6th segments of An-26 form a circle, and they indeed contain relatively complete attitudes compared with other segments, which demonstrates the rationality of the data division.

Besides target-attitude sensitivity, it is the prerequisite that the time-shift and amplitude-scale sensitivity problems of training and test HRRP echoes should be overcome before the model learning with training HRRPs and model validation with test HRRPs [18]. Specifically, for each HRRP echo in training and test sets, we adopt the centroid alignment [40] and L_2 normalization [20] to deal with its time-shift and amplitude-scale sensitivities, respectively.

After data partition for each target and preprocessing for each HRRP echoes, we conduct the training of TACNN by utilizing the training HRRP echoes. During the model learning, all HRRP echoes from all targets are considered to be independent and fed into the initialized TACNN at the same time. Then the outputted result of each HRRP echo can be obtained via the forward propagation of network, and the cross entropy can further be individually calculated for each HRRP echo. Taking the summation of cross entropies from all HRRP echoes as the objective function, we can perform model optimization via the SGD algorithm until the model convergence.

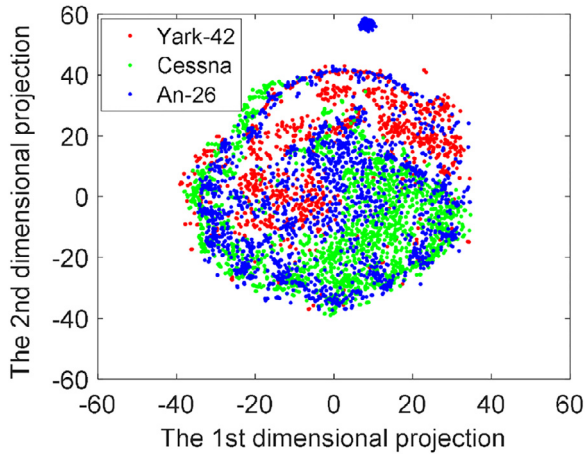

Fig. 9. Two-dimensional t-SNE visualization of HRRP data from three types of airplane targets.

In the test stage, each test HRRP echo is separately input to the learned TACNN to perform class prediction. After the testing of all HRRP echoes from all targets is completed, the recognition rate of each target (denoted as R_c with c being the class index, $c = 1, \cdots, C$, and C the number of targets) can be calculated as $R_c = n_c^{correct}/N_c$, with $n_c^{correct}$ and N_c representing the numbers of correctly classified HRRPs and HRRPs in the test HRRP set from target c, respectively. Finally, the average recognition rate (AR) of all targets can be expressed as $AR = \frac{1}{C}\sum_{c=1}^{C} R_c$.

The similarity of different targets is an important factor in assessing recognition performance, and the high recognition rate on dissimilar targets cannot give an objective evaluation to the model performance. Since the proposed TACNN directly uses HRRP to perform target recognition, we give the two-dimensional projections of test HRRPs from all targets via the t-SNE method in Fig. 9, to assess the difficulty degree of recognition task on airplane targets used in the paper. Here the t-SNE can map high-dimensional data to the low-dimensional space with the correlation of any two samples not being destroyed. As shown Fig. 9, except for a small number of samples from An-26, the HRRPs of all airplanes group into a cluster, in which HRRP samples from different targets overlap together. Consequently, three airplanes used in this paper are very similar to each other, resulting in the great difficulty to recognize different airplanes via their HRRPs. Accordingly, three airplanes in the experiment are suitable to validate the recognition performance of the proposed TACNN.

We compare our TACNN with some related radar HRRP target recognition methods, covering PCA [16], K-SVD [17], FA [18], SCAE [21], DBN [23], FCN [24], RNN [31], 1-D CNN, and TARAN [28], of which the first three methods are shallow models and others are deep models. In the experiment, we make multiple random samplings from the Gaussian distribution of $\mathcal{N}(\mathbf{0}, 0.1\mathbf{I})$ to initialize the TACNN's parameters, except for the biases being initialized as 0. During the optimization of loss function, we exploit Adam [41] optimizer to accelerate the model learning and set the learning rate to 10^{-4}.

For the implementation details, the 1-D CNN in the feature extraction module is composed of three convolutional layers, followed by each of which is a max-pooling based subsampling layer. The size of convolution kernel in each channel is 9×1, with the numbers of convolution kernels from all convolutional layers being set to 32, 32, and 64, respectively, i.e., $L = 1, 2, 3$, $h^1 = h^2 = h^3 = 9$, $Q_1 = Q_2 = 32$, and $Q_3 = 64$. In addition, the pooling rate is 2×1, and thus the output dimensionalities for each channel of three subsampling layers (i.e., H^1, H^2, and H^3) are 128, 64, and 32, respectively. In the attention module of our TACNN, the CNN features are segmented by sliding window on each feature channel with the window and stride lengths both being 2×1, i.e., $d = b = 2$; therefore, there are 16 feature segments ($T = 16$), and after flattening, the 16 feature vectors of size 128×1 are input into a single-layered FCN with its output dimensionality being 64 (i.e., $S = 64$). For the Bi-GRU, it generates a 100-dimensional hidden state for each timestep, which is the concatenation of two 50-dimensional hidden states from the unidirectional GRUs in opposite directions.

It should be pointed out that the HRRPs in this paper are obtained via some cooperative experiments and of high signal-to-noise ratios (SNRs) with SNR $\approx$ 30dB. However, the radar system is often faced with the non-cooperative environment in the test stage, where the high SNR of test HRRP samples cannot be guaranteed. To evaluate the noise-robust performance of our TACNN, we add simulated white noise to test data to obtain noisy test HRRPs. For each HRRP, the SNR is denoted as the ratio of the average power of target signal to that of the noise and can be mathematically expressed as

$$\mathrm{SNR} = 10 \times \log_{10}\frac{\bar{\mathrm{P}}_x}{\mathrm{P}_{Noise}} = 10 \times \log_{10}\frac{\sum_{i=1}^{D} E_{x_i}}{D \times \mathrm{P}_{Noise}}, \tag{13}$$

where $\bar{\mathrm{P}}_x$ denotes the average power of original HRRP, P_{Noise} denotes the noise power, E_{x_i} denotes the energy of target signal in the ith range cell, and D denotes the dimension of HRRP. Then the noise power added to the HRRP can be calculated as $\sum_{i=1}^{D} E_{x_i}/(D \times 10^{\frac{\mathrm{SNR}}{10}})$.

4.2. Model analysis

In the TACNN, the 1-D CNN is constructed to extract abundant features from HRRP. To examine what the network learns at different convolutional layers, we randomly select an HRRP test sample from each target and plot their learned features at three convolutional layers in Fig. 10. For each convolutional layer of each HRRP sample, the corresponding feature contains 64 channels (i.e., 64 features vectors), and the high-energy feature vectors from three channels are only presented for clarity. We see from Fig. 10 that the wave-forms and amplitudes of feature vectors in the first layer are more similar to those of the corresponding HRRP samples, whereas the wave-forms and amplitudes of feature vectors in the second and third layers gradually deviate from those of the HRRPs, especially for the third-layered features which are mostly of single peak and large amplitude. This demonstrates that the CNN feature presents the variation characteristic from intuition to abstraction with the increase of layer number.

From Fig. 10 we also see that the CNN features from non-target areas (i.e., the elements on either side of feature vectors) in the third convolutional layer in TACNN have been suppressed. This owes to the introduction of attention mechanism. Due to the label difference constraint, the attention module tends to only "select" the local CNN features from target areas with the most discrimination, in the way of assigning large attention coefficients to the features of discriminative target areas and small attention coefficients to those of the indiscriminative target areas and non-target areas. Furthermore, the attention coefficients from different local regions are calculated in the basis of corresponding local CNN features via the Bi-GRU in our TACNN. The larger the local CNN feature is, the greater the attention coefficient is. Therefore, to obtain the small attention coefficients for non-target areas, the corresponding learned local CNN features must be small, i.e., the local features from non-target areas at the third convolutional layer is suppressed. When there is no attention mechanism, i.e., the TACNN degenerate into 1-D CNN, the inter-class separability of CNN features in the support of HRRP weakens, due to the existence of features from indiscriminative target areas. With the purpose of en-

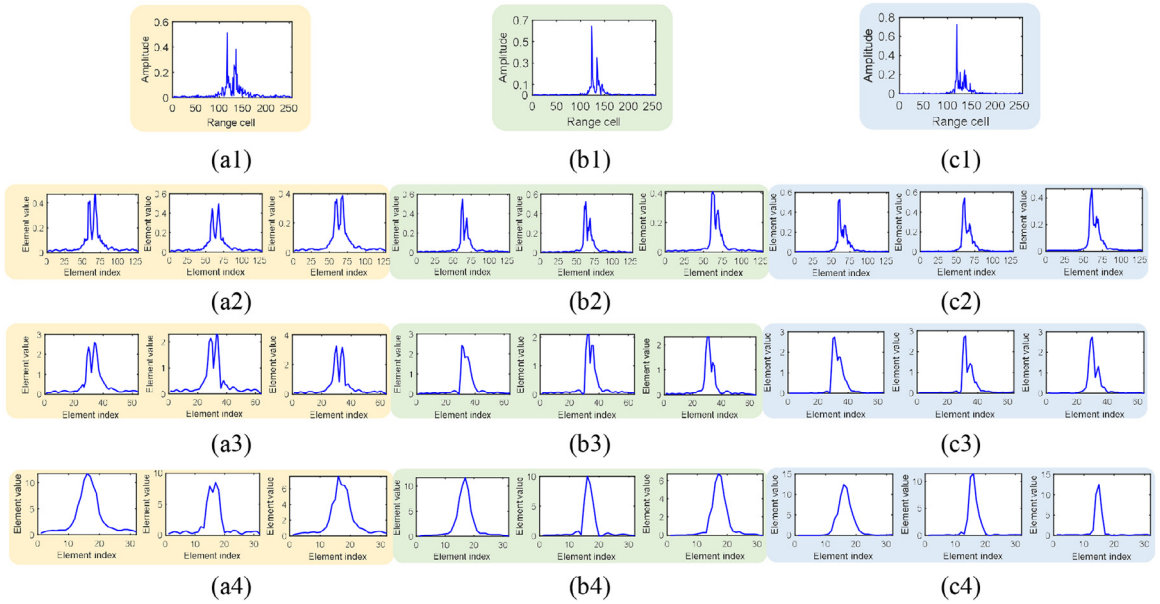

Fig. 10. CNN features of three test HRRP samples at different convolutional layers in our TACNN. Here the high-energy feature vectors from three channels are presented for each layer of each sample. (a1), (b1), and (c1) in the first row are HRRPs from Yark-42, Cessna Citation S/II, and An-26, respectively, and their features at the first convolutional layer are shown in (a2), (b2), and (c2) in the second row, similarly for the other rows.

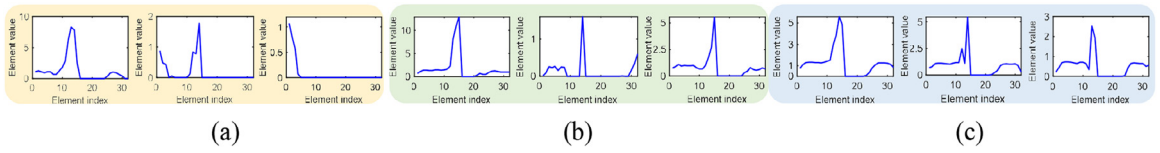

Fig. 11. CNN features of three test HRRP samples in Fig. 10 at the third convolutional layers, learned by 1-D CNN without attention mechanism. Here the high-energy feature vectors from three channels are presented for each sample. (a), (b), and (c) are the features from Yark-42, Cessna Citation S/II, and An-26, respectively.

hancing the feature discrepancies from different classes, the 1-D CNN tends to seek and strengthen non-target regional CNN features with certain inter-class differences, as shown in Fig. 11. However, the information in non-target areas, i.e., noise, is of great randomness, and the discriminative non-target areas found in the training stage via the label constraint is probably not suitable to test stage, which leads to the poor generalization ability, ultimately the worse recognition performance. Therefore, the attention mechanism in the proposed TACNN not only provides access to find target features with more separability and discard features with poor separability, but also can suppress the learning of features from non-target areas, both of which are beneficial to the recognition performance improvement.

To explore the separability of CNN features at different convolutional layers, we map all features of test HRRPs extracted by the 1-D CNN in our model to 2-D space via the t-SNE method. The detailed projection results of features from all convolutional layers are shown in Fig. 12. Moreover, the ratio of inter-class distance to intra-class distance (abbreviated as r) for the HRRPs/features is calculated to quantitatively validate the advantages of 1-D CNN in separable feature extraction. Here r can be expressed as

$$r = \frac{trace(S_b)}{trace(S_w)}, \quad S_b = \frac{1}{N}\sum_{c=1}^{C} N_c(\bar{\mathbf{z}}_c - \bar{\mathbf{z}})(\bar{\mathbf{z}}_c - \bar{\mathbf{z}})^{\mathrm{T}},$$
$$S_w = \frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N_c}(\mathbf{z}_n^c - \bar{\mathbf{z}}_c)(\mathbf{z}_n^c - \bar{\mathbf{z}}_c)^{\mathrm{T}}, \tag{14}$$

where S_b and S_w denote the between-class and within-class scatter matrices, respectively; $\mathbf{z}_n^c$, $\bar{\mathbf{z}}_c$, and $\bar{\mathbf{z}}$ denote the nth HRRP/feature from the class c, the mean of HRRPs/features from the cth class, and the mean of all HRRPs/features from all classes, respectively; N denotes the total number of HRRP samples, and $trace(\cdot)$ the trace of a matrix. According to the definition, the larger the value of r is, the more separable the HRRPs/features are. The comparison of r for observed HRRPs and learned features from different convolutional layers are summarized in Table 3.

In comparison of original HRRPs in Fig. 9, we can see from Fig. 12 that the CNN feature's dispersion increases and there is the phenomenon that some features separate from other features, even being individually grouped into a cluster, such as the features framed by dotted curves in Fig. 12. Furthermore, the pureness of categories included in each isolated cluster is relatively high. Although one target still cannot be better differentiated from others due to the less clear clustering phenomenon, the 1-D CNN in TACNN indeed has the tendency to learn the intra-class aggregated and inter-class dispersed HRRP features, such as the quantitative results shown in Table 3, which makes sense to recognition task. From Table 3, we can also see that the inter-class distinction of high-level features is greater than that of the low-level features, demonstrating the significance of deep structure.

As discussed above, the introduction of attention module in the proposed TACNN can make the model discover the valuable target features. To illustrate the effectiveness of attention mechanism, we separately train the 1-D CNN without attention mechanism and our TACNN, and then randomly choose two test HRRP samples from each target to send them to above models. The CNN and attention-weighted CNN features extracted by the 1-D CNN without attention mechanism and TACNN are shown in Fig. 13. It should be noted that the 1-D CNN without attention mecha-

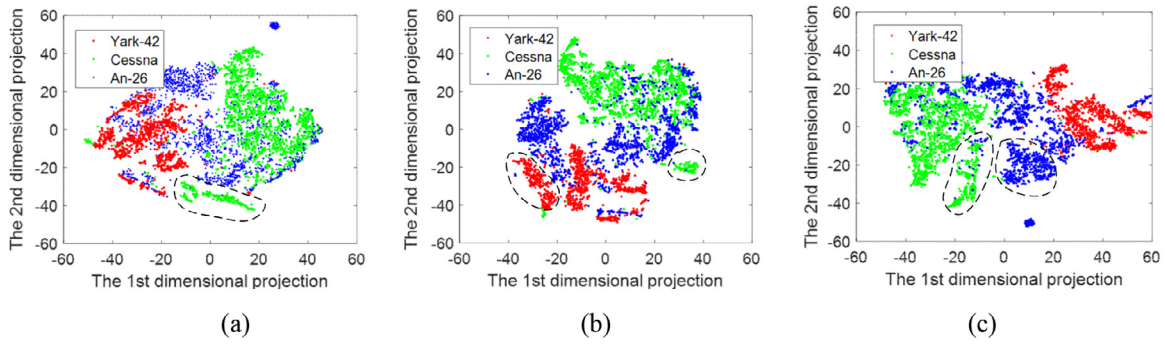

Fig. 12. 2-D t-SNE visualizations of test HRRPs' CNN features at different convolutional layers in the TACNN. The subfigure in (a), (b), and (c) are the visualizations of features from the first, second and third convolutional layers, respectively.

Table 3
The ratios of inter-class distance to intra distance (r) for HRRP samples and features from different convolutional layers.

HRRPs/Features	HRRP test samples	Features at the first convolutional layer	Features at the second convolutional layer	Features at the third convolutional layer
r	3.0886	4.1790	5.0183	9.4139

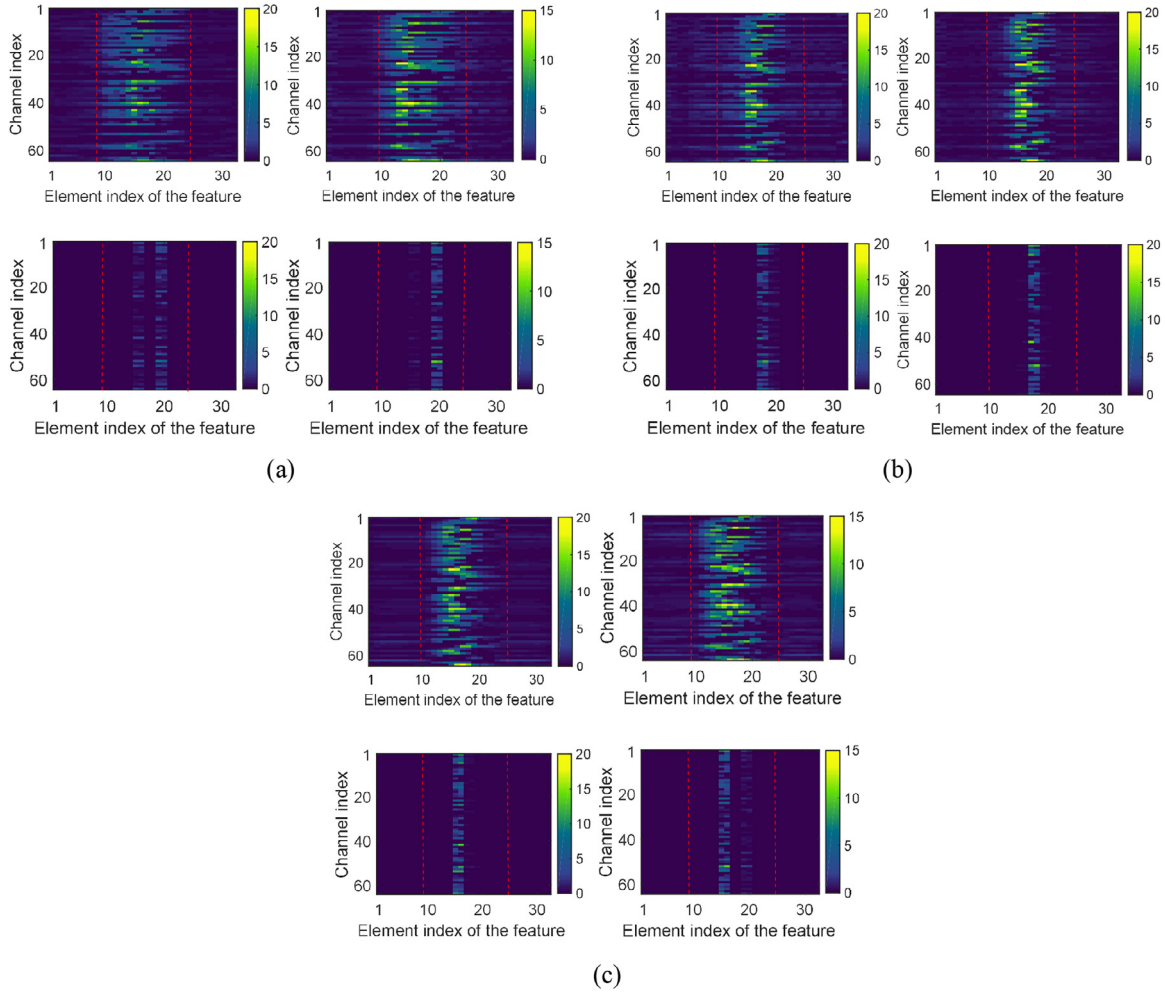

Fig. 13. The CNN features and attention-weighted CNN features from two test HRRPs of each target, learned by the 1-D CNN without attention mechanism and TACNN, respectively. (a), (b), and (c) are the extracted CNN features (the first row) and attention-weighted CNN features (the second row) of two test HRRP samples from Yark-42, Cessna Citation S/II, and An-26, respectively. For an arbitrary HRRP, its feature contains 64 channels, each of which is a 32-dimensional vector. Here the elements between two red dotted lines shown in each subfigure represent the features extracted from the target-regional areas of HRRP.

nism directly projects features at the last convolutional layer, i.e., the called CNN features shown in Fig. 13, to the corresponding labels, and thus CNN features shown here are immune to attention mechanism and different from those in Fig. 12, which are learned via the 1-D CNN of TACNN and affected by the attention mechanism. In Fig. 13, the first and second rows in each subfigure represent the CNN and attention-weighted CNN features, respectively; different columns correspond to different HRRP samples. Here the 1-D CNN without attention mechanism has the same architecture as the feature extraction module in TACNN. Thus, on the basis of discussion in Section 4.1, two types of features for an HRRP both have 64 vector channels and each channel is embodied as a 32-dimensional vector. In addition, an element in each vector forming the CNN feature/attention-weighted CNN feature is learned from 8 range cells of the HRRP sample due to the HRRP dimension, convolutional layer number, and pooling rate being 256, 3, and 2, respectively, in the experiment, and different elements correspond to different range cell groups. By combining above analysis and the length of HRRP support area, we can roughly determine the locations of elements from target areas for each vector of each CNN feature/attention-weighted CNN feature, which are marked in red dotted lines in Fig. 13.

From Fig. 13, we note that the CNN features learned by 1-D CNN without attention mechanism have the large values in the positions of HRRP non-target areas, where the values of attention-weighted CNN features learned by our TACNN are nearly 0, thus having no effect on the subsequent recognition. Moreover, in comparison of features learned via the 1-D CNN without attention mechanism, the large feature values in the TACNN only appear in the partial target areas, rather than all target areas, and the feature values in some target regions even also equal to 0. There is a broad consensus that the non-target areas of HRRP is full of noise, which has adverse influence on the target recognition. Furthermore, since only a part of the target structures may possess the characteristic to differentiate itself from others, not all features from all target areas are available to recognition task, and there should be a selection for the features from all target areas. Therefore, the introduction of attention mechanism in our TACNN plays the desired role in HRRP feature extraction. That is, the attention mechanism in the proposed TACNN can promote to locate the features from target areas of HRRP, by suppressing the non-target regional features unfavorable to recognition task, and meanwhile has the capability to find more discriminative features from all target-regional features. Consequently, the attention mechanism in the proposed TACNN has the potential to enhance the inter-class separability of extracted features, ultimately can facilitate the improvement of model recognition ability.

In principle, the feature selection ability of our TACNN is owing to the attention coefficients produced by Bi-GRU. The learned attention coefficients of features for all test HRRPs are exhibited in Fig. 14. In addition, to qualitatively demonstrate the superiority of Bi-GRU based attention mechanism compared to the conventional and unidirectional GRU based attention mechanism, we substitute Bi-GRU in the TACNN with conventional attention approach used in TARAN and unidirectional GRU, respectively, and then construct a conventionally attentional CNN (CACNN) and a unidirectional GRU-based attentional CNN (UACNN), respectively, with the detailed calculation of attention coefficient in CACNN being provided in Appendix C for clarity. The learned attention coefficients of features for all test HRRPs via the CACNN and UACNN are also shown in Fig. 14.

From Fig. 14 we also note the attention weights corresponding to the HRRP features from non-target areas are significantly smaller than those from the target areas in our TACNN. Therefore, the Bi-GRU based attention mechanism in our TACNN can effectively strengthen the features from target signals and suppress the features from useless noises, which is advantageous to the acquisition of satisfactory recognition accuracy. However, the attention weight differences corresponding to HRRP local features from non-target and target areas in the CACNN and UACNN are much less than those in the TACNN, which demonstrates that the abilities of CACNN and UACNN to highlight/weaken target/non-target regional features degenerate. As a result, the Bi-GRU based attention mechanism can achieve the more accurate attention weights than the conventional and unidirectional GRU based attention mechanisms.

To more intuitively compare the selections of discriminative target areas, we project the attention coefficients learned by the CACNN, UACNN and TACNN to the observation space, respectively, and randomly select two HRRP samples from each target to show in Fig. 15. Moreover, we also present the attention coefficients learned via the TARAN in Fig. 15 as the baseline. As we see from Fig. 15 that the worst attention coefficient learning occurs in the TARAN, where the large attention coefficients almost locate in the non-target areas. This is because the TARAN adopts RNN to extract the sequential features of HRRP. Due to the memory property of the RNN, the target information is preserved into the non-target regions, thus affecting the accurate learning of attention coefficients. And meanwhile, the attention module in TARAN is realized via a simple nonlinear transformation, which cannot rectify the inaccuracy caused by the unsatisfactory feature learning. The drawbacks of TARAN both in the feature extraction module and attention module lead to the worst results. For CACNN and UACNN, they take the 1-D CNN as the feature extraction module, with which the feature extractions of different HRRP local regions are independent to each other, thus avoiding the mixture of features from non-target and target areas appearing in the TARAN. Consequently, the learning of attention coefficients in the CACNN and UACNN is better than that of the TARAN. However, because of the unreasonable attention mechanism, as discussed in Section 3.2, there are still large attention coefficients locating in the non-target areas of HRRP in the CACNN and UACNN. Fortunately, attention coefficients corresponding to the non-target areas in our TACNN are almost zero due to the better constructions both of the feature extraction and attention modules. Therefore, the proposed TACNN can better remove the adverse effect of information from non-target regions on the recognition task, compared with the TARAN, CACNN and UACNN. Thus, the TACNN with Bi-GRU based attention mechanism should acquire higher recognition accuracy.

4.3. Recognition performance

In this subsection, we evaluate the TACNN's recognition capacity via the comparison of recognition accuracies and the analysis of feature separability. In Fig. 16, the recognition performance of different methods on three airplane targets is displayed. Here the recognition rates of UACNN and CACNN are also presented in Fig. 16 to quantitatively demonstrate the superiority of Bi-GRU based attention mechanism than the conventional and unidirectional GRU based attention mechanism. As shown in Fig. 16, the first two models, i.e., PCA and K-SVD, get the worst recognition performance, since they can only extract shallow features with poor characterization capabilities. Although the FA is also of shallow structure, it constructs a probabilistic model for HRRPs from each frame of each target, within which the HRRPs are collected roughly without the motion through range cells (MTRC) [18] [20], rather than for HRRPs from each class (such as the PCA and K-SVD) or all classes (such as the deep models). The fine modeling of FA makes it possible to acquire the higher recognition rate than the other shallow methods, even several deep models. For the latter nine deep methods (from CFVAE to UACNN), their performance is superior to that of the PCA and K-SVD. Especially for the CFVAE, SCAE, TARAN, 1-D CNN, CACNN, and UACNN, their recog-

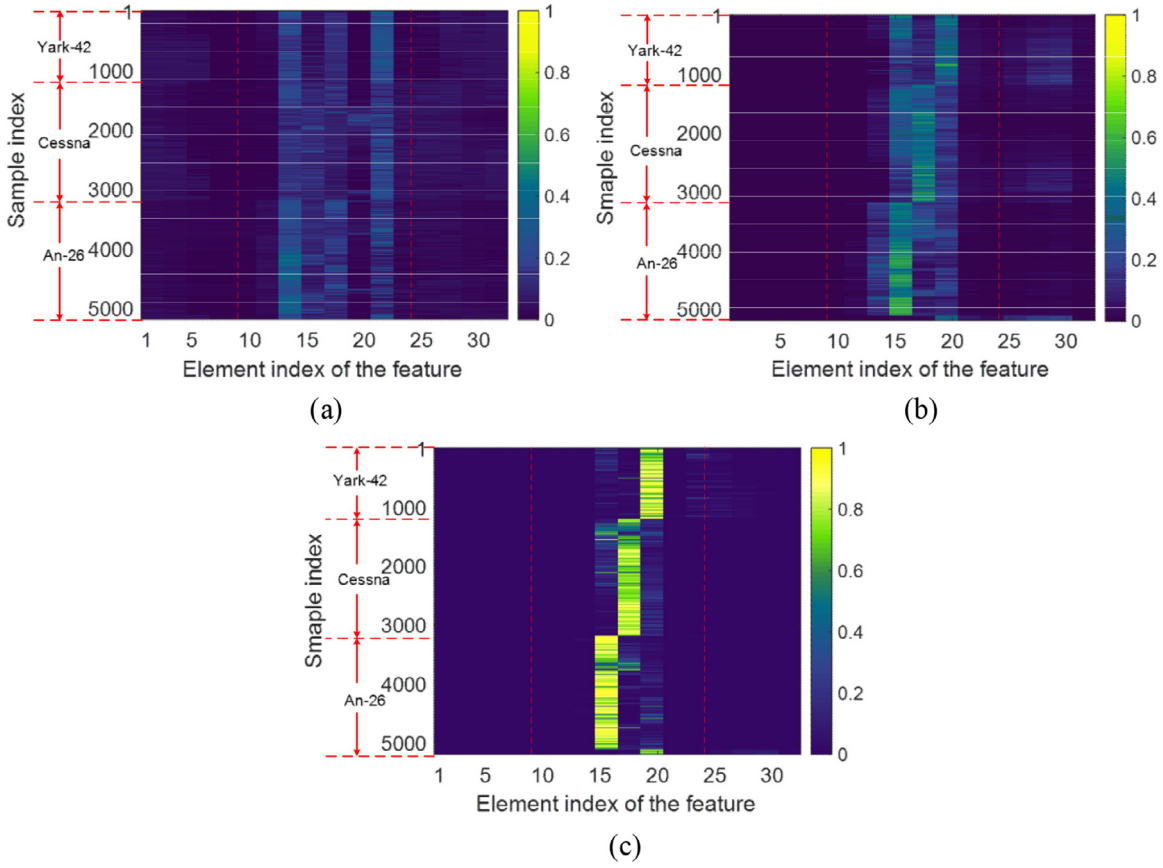

Fig. 14. HRRP features' attention coefficients for all test samples, learned via the CACNN, UACNN and TACNN. The previous 1200 HRRPs are from Yark-42, the last 2000 HRRPs are from An-26, and the rest are from Cessna Citation S/II. (a) The learned attention coefficients via the CACNN. (b) The learned attention coefficients via the UACNN. (c) The learned attention coefficients via the TACNN. Here the values between two red dotted lines shown in each subfigure represent the attention coefficients corresponding to the target areas.

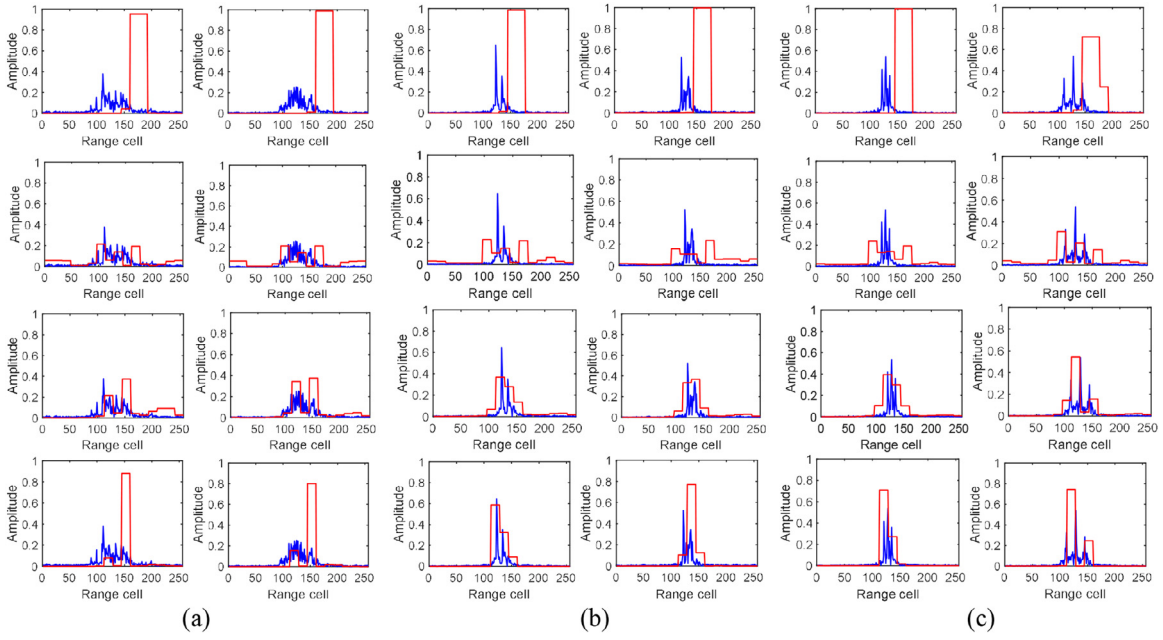

Fig. 15. Examples of the attention coefficients projected to the data domain, learned by the TARAN, CACNN, UACNN and our TACNN, respectively. The subfigures from (a), (b), and (c) show the HRRPs (blue curves) and the corresponding attention coefficients (red polygonal lines) from Yark-42, Cessna Citation S/II, and An-26. For each group of subfigures, the rows from 1 to 4 represent two HRRP samples and their learned attention coefficients via the TARAN, CACNN, UACNN, TACNN, respectively.

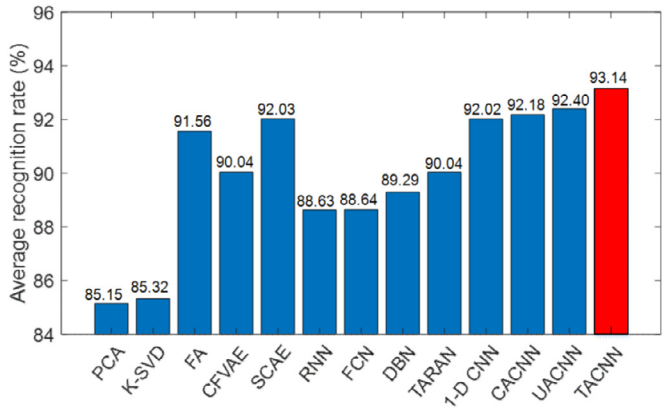

Fig. 16. The comparison of our TACNN with other related methods in average recognition rate.

Table 4
Comparison of ratios of inter-class distance to intra distance (r) with different methods.

Method	TARAN	1-D CNN	TACNN
r	6.7394	12.2472	19.8417

nition accuracies are larger than 90%. Nevertheless, they cannot beat the proposed TACNN, due to its better feature extraction ability by combining the 1-D CNN with attention mechanism. Moreover, because of the unreasonable attention approaches, the recognition performance of CACNN and UACNN is worse than that of the TACNN.

To obtain insights into the recognition performances of different models in each target, we present the recognition rate confusion matrices of our TACNN and other models closely related to the proposed method, including TARAN, 1-D CNN in Fig. 17. Moreover, as the baseline, the confusion matrix of K-SVD is also given for comparison. From Fig. 17, we note that the ability of deep models (TARAN, 1-D CNN and TACNN) to accurately recognize each target is obviously better than that of the traditional model with shallow structure (K-SVD). This is because that the deep models can excavate fundamentally different information among all targets via the constant feature extraction layer-by-layer, and thus they can obtain the deep representations with stronger inter-class separability and more effectiveness than the traditional shallow models, by which the extracted features are vulnerable to external factors (such as the change of target attitude) and has limited ability in the representation of target essence. As a result, the overall recognition rates of deep models are higher than those of the traditional shallow methods. For deep models, we also see from Fig. 17 that the recognition ability of our TACNN to each target is stronger and more stable than those of other two models. Especially for Cessna Citation S/II target, our TACNN can correctly classify 90.10% of the samples, nearly 8% and 2% higher than the TARAN and 1-D CNN. Consequently, the TACNN model can more accurately describe the diversities among different targets, further acquiring the superior recognition performance.

Similarly to Section 4.2, we also give the t-SNE visualizations and ratios of inter-class distance to intra-class distance for the test HRRPs' features extracted by the TARAN, 1-D CNN, and TACNN, in Fig. 18 and Table 4, to intuitively exhibit the feature separability. Different from the features in Fig. 12, here the features correspond to the TACNN mean the attention-weighted CNN features, i.e., $\tilde{\boldsymbol{F}}$ in Fig. 5. We can clearly see that our TACNN has the best feature separability, and then the 1-D CNN and TARAN in turn, which is consistent with the conclusion drawn from Fig. 16. By comparing Fig. 12(c) and Fig. 18(c), we can also obviously see that the attention-weighted CNN features are grouped into three clusters, as framed by dotted curves in Fig. 18(c), and moreover, the category purity of attention-weighted CNN features is extremely high for each cluster. Thus, we can conclude that the separability of weighted CNN features is stronger than those of the unweighted CNN features for the proposed TACNN, further validating the effectiveness of the attention mechanism. Because of the strong separability, a quite good recognition performance can be obtained when the attention-weighted CNN features extracted by the TACNN are inputted into the classifier. Therefore, the proposed TACNN have the strong capability in the acquisition of targets' discriminative information that can be used to easily recognize different targets.

In the experiment, the target attitudes in the training stage almost cover all those in the test stage to deal with the sensitivity of HRRP to target attitudes. Nevertheless, the HRRPs with complete target attitudes cannot be guaranteed due to the rapid change of target attitude, especially for the non-cooperative targets. In this case, the majority of learning-based methods will have limited recognition performance on the test HRRPs whose corresponding target attitudes does not appear in the training phase, due to the great distinction of HRRPs with different target attitudes, i.e., the target-attitude sensitivity of HRRP.

As demonstrated in Introduction, the proposed TACNN should be robust to the change of target orientation due to its ability to extract the most discriminative HRRP features. To validate the robustness of our TACNN to target orientation change, we first divide all training HRRPs from each target into several different groups, with each group containing 1024 HRRP samples and the variation range of target attitudes for each group approximately being 3°, and then select HRRP groups at equal interval to train the TACNN. In detail, the selection intervals are set to 0 (i.e., all training HRRPs with nearly complete target attitudes), 1, 2, 3, 4, and 5, expressed as $\Delta = 0$, $\Delta = 1$, $\Delta = 2$, $\Delta = 3$, $\Delta = 4$, and $\Delta = 5$, respectively, for convenience. The recognition accuracies of TACNN and other related methods under different selection interval conditions are shown in Fig. 19. From Fig. 19 we see that the sensitivity of deep models, including TACNN, 1-D CNN, TARAN, FCN), and SCAE, to the target orientation change is weaker than those of PCA and K-SVD methods. Especially in the cases of $\Delta = 3$, $\Delta = 4$, and $\Delta = 5$, the decreases of recognition rates for PCA and K-SVD are far more obvious than those of the deep models. This is because that the deep models tend to mine more essential information hidden in HRRP data, which is largely dependent on the self-structure of target and less affected by other factors, such as changes of target aspect and noise level. That is, the deep models have a certain ability to extract target-attitude robust features. However, the single-layered models, e.g., K-SVD and PCA, can only acquires shallow and simple representations of HRRP, which have the limited capability to characterize target's essence and are sensitive to other factors besides target itself. Therefore, the deep models in Fig. 19 exhibit the more robustness to the change of target attitude than the K-SVD and PCA. For the FA method, it models HRRP data within probabilistic framework to describe the distribution characteristic of observations. Since the probability distribution represents the uncertainty of data and has a certain generalization ability to unseen samples, some HRRPs from adjacent groups may also be correctly recognized after the completion of model learning with one HRRP group. Thus FA has better the robustness to target attitude change than the PCA and K-SVD, even though it also belongs to the shallow model. Among deep learning-based approaches, our TACNN obtains the best robustness to the change of target attitude because the attention mechanism in our TACNN can better make the model focus on the extraction of essential target-regional features, further weakening the effect of other factors outside of target itself. In particular, when $\Delta = 3$, $\Delta = 4$, and $\Delta = 5$, the recognition accu-

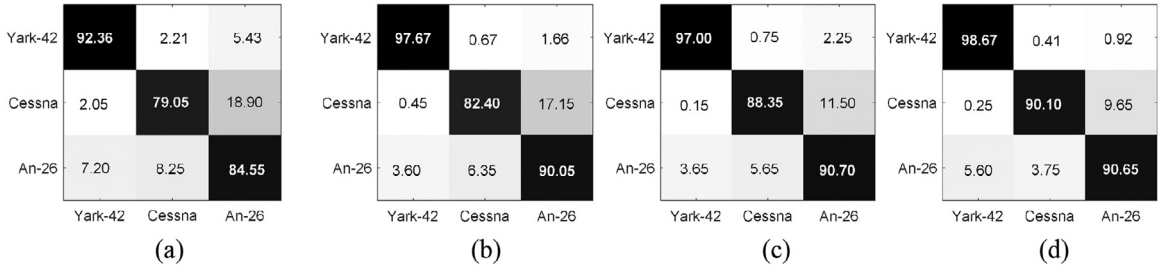

Fig. 17. Confusion matrices of the recognition accuracy (%) via the K-SVD, TARAN, 1-D CNN, and our TACNN, respectively. (a) K-SVD; (b) TARAN; (c) 1-D CNN; (d) TACNN.

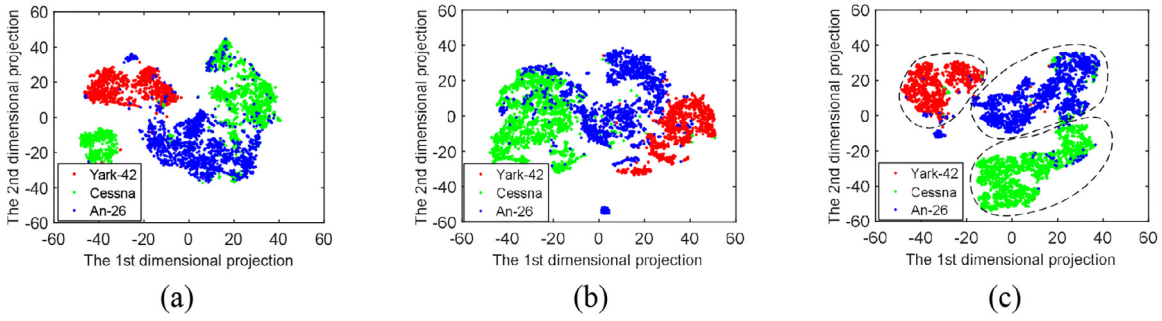

Fig. 18. 2-D t-SNE visualizations of the test HRRPs' features learned by the (a) TARAN, (b) 1-D CNN, and (c) TACNN, respectively. (d) is the no manual coloring version of (c). Here the features correspond to the TACNN mean the attention-weighted CNN features.

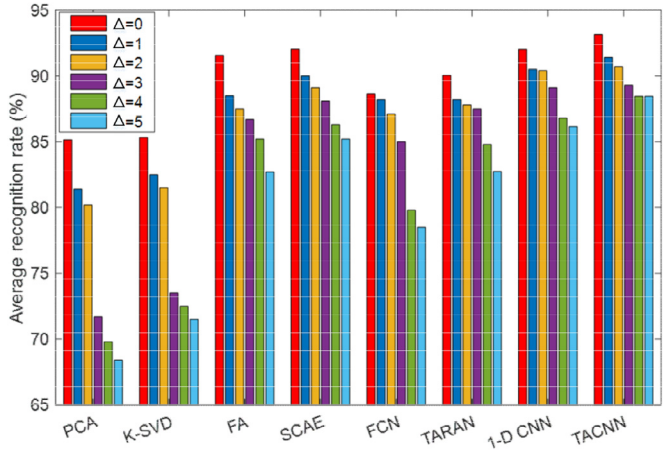

Fig. 19. Performance comparison of different methods on the robustness to target orientation change.

racies of TACNN are above 88%, which are far larger than those of other related models. For the TARAN method, its attention mechanism may make the model focus on some non-target areas, thus reducing the capability of model to extract essential features from target, and ultimately having the poor robustness to target attitude change, as shown in Fig. 19. In summary, the proposed TACNN is more robust to the target orientation change than other related methods.

As discussed in Section 2.1, the SNRs of training HRRPs are high in the case of cooperative measurement circumstances, whereas those of test HRRP samples are often low due to the non-cooperative test scenes. That is, the test HRRPs are interfered by noise, especially for the non-target areas. To evaluate the noise-robust performance of our TACNN, we add white noise with different power to test data and obtain 7 groups of noisy test HRRP sets, with SNRs for different test sets being 30 dB (i.e., the original test HRRPs), 25 dB, 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, respectively. In Table 5, we present the recognition rates of test HRRPs with different SNRs, via the TACNN and other related methods. For

Table 5
Recognition rates (%) of test HRRPs with different SNRs via the TACNN and other related models.

Methods	SNR						
	0dB	5dB	10dB	15dB	20dB	25dB	30dB
TACNN	***57.18***	***62.31***	***71.47***	***74.69***	***86.30***	***92.27***	***93.14***
1-D CNN	56.32	61.98	68.49	71.85	86.23	91.38	92.02
TARAN	56.12	61.13	63.81	67.15	72.96	88.34	90.04
RNN	53.39	59.35	63.88	65.21	73.17	87.13	88.82
DBN	52.65	55.87	62.97	66.57	74.80	86.64	89.29
FCN	47.19	58.31	63.12	65.29	71.04	87.67	88.64
FA[15]	51.28	61.37	65.01	67.75	76.32	86.21	91.65
K-SVD	56.23	57.39	60.96	63.47	72.31	84.17	85.32
PCA[22]	35.29	44.28	51.37	58.44	68.43	83.20	85.15

each group of experiment, the noise adding operation is repeated 5 times and the final result for each method reported in Table 5 is the average over 5 recognition accuracies, to deal with the uncertainty caused by the randomness of noise. Moreover, it should be pointed out that the recognition rates of FA and PCA are directly from references [15] and [22], respectively.

From Table 5, we see that the overall recognition performance of deep models, including TACNN, 1-D CNN, TARAN, RNN, DBN, and FCN, is higher than that of the K-SVD and PCA methods. The reason is similar to that in the robustness experiment on target attitude. That is, the deep models tend to mine more essential information hidden in HRRP data, which be of a certain robustness to the external environment change, such as the variation of noise level. However, the single-layered models, e.g., K-SVD and PCA, can only acquires shallow and simple representations of HRRP, having the limited capability to characterize target's essence and being sensitive to other factors besides target itself. The FA is a probabilistic model and can reflect the data uncertainty, which is robust to the change of HRRP data caused by the introduction of random noise, to some extent. Therefore, the FA has the higher recognition rates than the K-SVD and PCA under different SNR conditions. For the deep learning-based approaches, the recognition accuracies of the proposed TACNN are significantly higher than those of other related models because the attention mechanism in our TACNN can

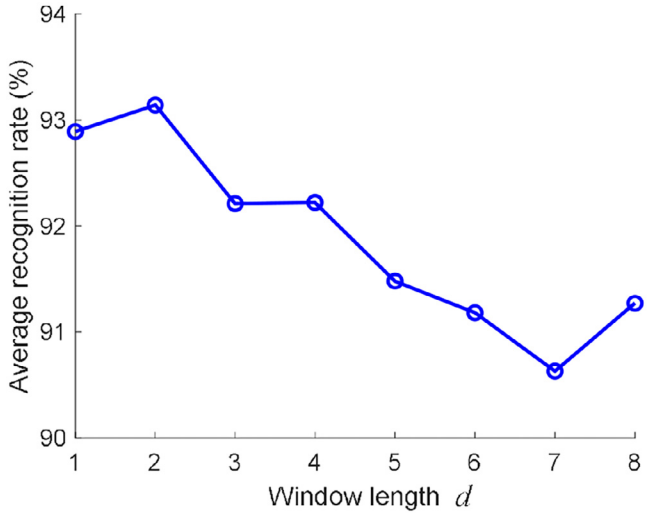

Fig. 20. Variation of the recognition performance with window length d in the attention module of our TACNN.

Fig. 21. The effect of GRU hidden state dimension on the recognition performance.

better suppress the noise in non-target areas of HRRP and only utilize the target-regional information to perform recognition, weakening the interference from noise. Especially in the case of SNR being 10 dB, our TACNN can still work (if we take 70% as the recognition threshold), whereas other methods can not. In summary, the proposed TACNN has the superior noise-robust recognition performance than other related methods.

4.4. Effect of model parameters

The setting of window length d used in the attention module of our TACNN has a great effect on the model recognition performance. Fig. 20 presents the variation of average recognition rate with d, from which we know that the recognition accuracy is the highest when $d = 2$, and too large or too small value of d will cause the performance degradation. The small d value brings extensive timestep inputs for the Bi-GRU, which increases the training difficulty of the attention module; on the contrary, the large d value may make some timestep inputs contain both the target information and useless information from the non-targets, affecting the accurate judgment of attention mechanism to the feature importance. Thus, the window length d should be adjusted with data in the experiment.

The attention coefficients are generated via a Bi-GRU network, which consists of two GRUs in opposite directions. Besides window length d, the dimension of hidden state outputted by the GRU is also a key to obtain the good recognition performance. As shown in Fig. 21 where the average recognition rate versus the hidden state dimension is presented, the recognition accuracy of the TACNN increases firstly and then decreases. The reason is that the low-dimensional hidden state has a serious information loss during the dimensionality reduction and cannot comprehensively express the input; the high-dimensional hidden state will make the model more complicated, causing the overfitting. In our experiment, 50 is a better choice for the hidden state dimension, as shown in Fig. 21.

5. Conclusion

The work reported here is the combination of radar signal processing and pattern recognition techniques. In detail, based on HRRP echo characteristic analysis, i.e., the structural and temporal characteristic of HRRP, our work aims to design a recognition model in allusion to radar HRRP signals, by introducing and improving the learning-based recognition algorithm in the field of pattern recognition. The model we construct, abbreviated as TACNN, integrates the 1-D CNN and attention mechanism. On the one hand, the 1-D CNN is capable of separately learning the abundant structure features from different HRRP regions; on the other hand, the attention module implemented via the Bi-GRU can data-adaptively assign a weigh to each local CNN feature by fully considering the temporal relations among all local features. The interplay betweemodule offers the potential for our TACNN to accurately discover the information from discriminative target areas and suppress the information with no contributions to the target recognition, thus promoting the extraction of HRRP features with more inter-class differences and ultimately the improvement in recognition performance. The experiments on three classes of measured HRRP data show that the recognition rate of the TACNN is up to 93.14%, which is significantly better than those of state-of-the-art methods. (Eqs.–1–6, 9, 13–18)

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Jian Chen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft. **Lan Du:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Guanbo Guo:** Formal analysis, Software, Validation, Visualization. **Linwei Yin:** Software, Validation. **Di Wei:** Conceptualization, Formal analysis.

Acknowledgments

This work was partially supported by the National Science Foundation of China (U21B2039), 111 Project, and Fundamental Research Funds for the Central Universities (XJS210219).

Appendix A

In this appendix, we show the calculation approach of multi-channel convolution $\boldsymbol{F}^{l-1} * \mathbf{W}_q^l$ in (7). In detail, the result of ex-

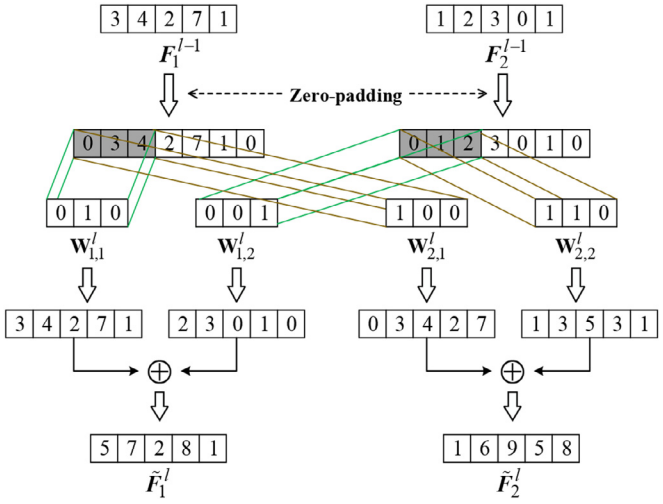

Fig. 22. Diagram of the multi-channel convolution. Here $\boldsymbol{F}_1^{l-1}$ and $\boldsymbol{F}_2^{l-1}$ represent the inputs of the 1st and 2nd channels in the convolutional layer l, respectively. The two convolution kernels, i.e., $\mathbf{W}_1^l = \{\mathbf{W}_{1,1}^l, \mathbf{W}_{1,2}^l\}$ and $\mathbf{W}_2^l = \{\mathbf{W}_{2,1}^l, \mathbf{W}_{2,2}^l\}$, are both of double-channel. For each channel of each kernel, the convolution operation applies it to the input from the corresponding channel, followed by the summation of two convolutional results. Thus, the channel number of output is equal to that of the kernel, such as the $\tilde{\boldsymbol{F}}_1^l$ and $\tilde{\boldsymbol{F}}_2^l$ in the figure.

pression $\boldsymbol{F}^{l-1} * \mathbf{W}_q^l$ is the convolution summation of all channels and has the expression of

$$\boldsymbol{F}^{l-1} * \mathbf{W}_q^l = \sum_{i=1}^{Q_{l-1}} \boldsymbol{F}_i^{l-1} * \mathbf{W}_{q,i}^l, \tag{15}$$

where $\boldsymbol{F}_i^{l-1} \in \mathbb{R}^{H^{l-1}}$ and $\mathbf{W}_{q,i}^l \in \mathbb{R}^{h^l \times 1}$ denote the ith channel of $\boldsymbol{F}^{l-1}$ and $\mathbf{W}_q^l$, respectively. Moreover, in the process of convolution, we take the central element of $\mathbf{W}_{q,i}^l$ (its dimension is often set to an odd number) as reference point and move $\mathbf{W}_{q,i}^l$ point by point on zero-padded $\boldsymbol{F}_i^{l-1}$, in order to avoid the problem of edge information loss caused by the size reduction of outputted feature. For ease of understanding, we take a numerical example to illustrate the multi-channel convolution operation in Fig. 22. From Eq. (7) and Eq. (15) we know that the convolutional result of the input and a kernel at one layer is a one-dimensional vector, referred to as the feature vector, and the output of a convolutional layer is the collection of multiple feature vectors, called a feature map, represented as a vector block shown in Fig. 22.

Appendix B

This appendix shows the calculation of hidden state in each unidirectional GRU of Bi-GRU. The GRU contains two gates, called reset and update gates. The reset gate controls the amount of previous information written into the current input, and update gate determines that the model mainly remembers the past knowledge or present information. Taking the forward GRU as an example, the reset gate $\vec{\mathbf{r}}_t$ and update gate $\vec{\mathbf{z}}_t$ for the tth timestep input $\boldsymbol{FC}_t$ are calculated as

$$\vec{\mathbf{r}}_t = \sigma\left(\vec{\mathbf{W}}_r\left[\vec{\mathbf{h}}_{t-1}, \boldsymbol{FC}_t\right] + \vec{\mathbf{b}}_r\right) \quad , \quad \vec{\mathbf{z}}_t = \sigma\left(\vec{\mathbf{W}}_z\left[\vec{\mathbf{h}}_{t-1}, \boldsymbol{FC}_t\right] + \vec{\mathbf{b}}_z\right), \tag{16}$$

where $\vec{\mathbf{W}}_r \in \mathbb{R}^{M\times(M+S)}$ and $\vec{\mathbf{W}}_z \in \mathbb{R}^{M\times(M+S)}$ are the projection matrices of the reset and update gates, respectively, with M and S being the dimensions of hidden state and $\boldsymbol{FC}_t$, respectively; $\vec{\mathbf{b}}_r$, $\vec{\mathbf{b}}_z \in \mathbb{R}^M$ are the bias terms, and $\sigma(\cdot)$ is the sigmoid function. Then the hidden state $\vec{\mathbf{h}}_t$ has the form of

$$\vec{\mathbf{h}}_t = (1 - \vec{\mathbf{z}}_t) \circ \vec{\mathbf{h}}_{t-1} + \vec{\mathbf{z}}_t \circ \tilde{\mathbf{h}}_t \quad ,$$
$$\tilde{\mathbf{h}}_t = \tanh\left(\vec{\mathbf{W}}_h\left[\vec{\mathbf{r}}_t \circ \vec{\mathbf{h}}_{t-1}, \boldsymbol{FC}_t\right] + \vec{\mathbf{b}}_h\right), \tag{17}$$

where $\vec{\mathbf{W}}_h \in \mathbb{R}^{M\times(M+S)}$ denotes the projection matrix, $\vec{\mathbf{b}}_h \in \mathbb{R}^M$ denotes the bias, and $\circ$ represents the element-wise product.

Appendix C

To demonstrate the superiority of Bi-GRU based attention mechanism in our TACNN, we substitute Bi-GRU with conventional attention approach and construct a CACNN for comparison. This appendix mainly shows the calculation of attention mechanism in the CACNN. For the attention coefficient a_t of each local HRRP feature $\boldsymbol{F}_t$ $(t = 1, \cdots, T)$ in the CACNN, it can be calculated as

$$a'_t = \mathbf{U}_a \tanh\left(\mathbf{W}_a \boldsymbol{FC}_t\right), a_t = \frac{a'_t}{\sum_{t=1}^{T} a'_t} \tag{18}$$

where $\{\boldsymbol{FC}_t | \boldsymbol{FC}_t \in \mathbb{R}^S\}_{t=1}^T$ denote the transformed local feature obtained by successively flattening and nonlinearly transforming $\boldsymbol{F}_t$ learned with the 1-D CNN and are the same as those in the TACNN; $\mathbf{W}_a \in \mathbb{R}^{V\times S}$ and $\mathbf{U}_a \in \mathbb{R}^{1\times V}$ are the transformation matrices in the conventional attention mechanism; $\tanh(\cdot)$ represents the hyperbolic tangent function; a_t is the attention coefficient corresponding to $\boldsymbol{F}_t$ in the CACNN.

References

[1] W.G. Carrara, R.S. Goodman, R.M. Majewski, Spotlight Synthetic Aperture Radar-Signal Processing Algorithms, Arthech House, Norwood, MA, 1995.
[2] W. Ye, Ph.D, Xidian Univ, Xi'an, China, 1996.
[3] A.R. Persico, C. Clemente, D. Gaglione, C.V. Ilioudis, J.L. Cao, L. Pallotta, A.D. Maio, I. Proudler, J.J. Soraghan, On model, algorithms, and experiment for micro-doppler-based recognition of ballistic targets, IEEE Trans. Aerosp. Electron. Syst. 52 (3) (2017) 1088–1108.
[4] Y. Luo, Q. Zhang, N. Yuan, F. Zhu, F.F. Gu, Three-dimensional precession feature extraction of space targets, IEEE Trans. Aerosp. Electron. Syst. 50 (2) (2014) 1313–1329.
[5] W.P. Zhang, Y.W. Fu, L. Nie, Zhao G.H, W. Yang, J. Yang, Parameter estimation of micro-motion targets for high-range-resolution radar using high-order difference sequence, IET Signal Process 12 (1) (2018) 1–11.
[6] X. Liu, S.W. Xu, S.Y. Tang, CFAR strategy formulation and evaluation based on fox's H-function in positive alpha-stable sea clutter, Remoting Sens 12 (8) (2020) 1273.
[7] X. Liu, L. Du, S.W. Xu, GLRT-based coherent detection in sub-Gaussian symmetric alpha-stable clutter, IEEE Geosci. Remote Sens. Lett. 19 (99) (2021) 1–5.
[8] L. Du, P.H. Wang, L. Zhang, H. He, H.W. Liu, Robust statistical recognition and reconstruction scheme based on hierarchical Bayesian learning of HRR radar target signal, Expert Syst. Appl. 42 (14) (2015) 5860–5873.
[9] L. Shi, P.H. Wang, H.W. Liu, L. Xu, Z. Bao, Radar HRRP statistical recognition with local factor analysis by automatic Bayesian Ying-Yang harmony learning, IEEE Trans. Signal Process. 59 (2) (2011) 610–617.
[10] M. Christopher, K. Alireza, Maritime ATR using classifier combination and high resolution range profile, IEEE Trans. Aerosp. Electron. Syst. 47 (4) (2011) 2558–2573.
[11] J.S. S.lomka, Features for high resolution radar range profile based ship classification, The fifth International Symposium on Signal Processing and its Applications (ISSPA), Queensland, Australia, 1999.
[12] K.T. Kim, D.K. Seo, H.T. Kim, Efficient radar target recognition using the MUSIC algorithm and invariant feature, IEEE Trans. Antenna. Propag. 50 (3) (2002) 325–337.
[13] X.D. Zhang, Y. Shi, Z. Bao, A new feature vector using selected bispectra for signal classification with application in radar target recognition, IEEE Trans. Signal Process. 49 (9) (2001) 1875–1885.
[14] L. Du, H.W. Liu, Z. Bao, M.D. Xing, Radar HRRP target recognition based on higher-order spectra, IEEE Trans. Signal Process. 53 (7) (2005) 2359–2368.
[15] L. Du, H.W. Liu, P.H. Wang, B. Feng, M. Pan, Z. Bao, Noise robust radar HRRP target recognition based on multitask factor analysis with small training data size, IEEE Trans. Signal Process. 60 (7) (2012) 3546–3559.
[16] L. Du, H.W. Liu, Z. Bao, J.Y. Zhang, Radar automatic target recognition using complex high-resolution range profiles, IET Radar Sonar Navig 1 (1) (2007) 18–26.
[17] B. Feng, L. Du, H.W. Liu, F. Li, Radar HRRP target recognition based on K-SVD algorithm, in: Proceedings of the IEEE CIE International Conference on Radar, China, 2011.
[18] L. Du, H. Liu, Z. Bao, Radar HRRP statistical recognition: parametric model and model selection, IEEE Trans. Signal Process. 56 (5) (2008) 1931–1944.
[19] X.F. Z.hang, B. Chen, H.W. Liu, L. Zuo, B. Feng, Infinite max-margin factor analysis via data augmentation, Pattern Recog 52 (2016) 17–32.

[20] L. Du, J. Chen, J. Hu, Y. Li, H. He, Statistical modeling with label constraint for radar target recognition, IEEE Trans. Aerosp. Electron. Syst. 56 (2) (2020) 1026–1044.
[21] B. Feng, B. Chen, H. Liu, Radar HRRP target recognition with deep networks, Pattern Recognit 61 (2017) 379–393.
[22] L.Y. Liao, L. Du, J. Chen, Class factorized complex variational auto-encoder for HRR radar target recognition, Signal Process 182 (2020) 1–11.
[23] M. Pan, J. Jiang, Q. Kong, J. Shi, Q. Sheng, T. Zhou, Radar HRRP target recognition based on t-SNE segmentation and discriminant deep belief network, IEEE Geosci. Remote Sens. Lett. 14 (9) (2017) 1609–1613.
[24] J. Chen, L. Du, L.Y. L.iao, Discriminative mixture variational autoencoder for semisupervised classification, IEEE Trans. Cybern. PP (2020) 1–15.
[25] J.W. Wan, B. Chen, B. Xu, H.W. Liu, L. Jin, Convolutional neural networks for radar HRRP target recognition and rejection, EURASIP J. Adv. Signal Process. 5 (2019) 1–17.
[26] Y.Q. Wang, M.L. Huang, X.Y. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, USA, 2016.
[27] K. Cho, B.V. Merrienboer, C. Gucehre, D. Bahdanau, F. Bougares, H. Schwenk, Y.S. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, proceedings og the 2014 Conference on Empirical Methods in Nature Language Processing (EMNLP), Qatar, 2014.
[28] B. Xu, B. Chen, J.W. Wan, H.W. Liu, L. Jin, Target-aware recurrent attentional network for radar HRRP target recognition, Signal Process 155 (2019) 268–280.
[29] D.D. Guo, B. Chen, W.C. Chen, C.J. Wang, H.W. Liu, M.Y. Zhou, Variational temporal deep generative model for radar HRRP target recognition, IEEE Trans. Signal Process. 68 (2020) 5795–5809.
[30] C. Du, L. Tian, B. Chen, L. Zhang, W.C. Chen, H.W. Liu, Region-factorized recurrent attentional network with deep clustering for radar HRRP target recognition, Signal Process 183 (2021) 1–10.
[31] A. Graves, A. Mohamed, G.E. Hinton, Speech recognition with deep recurrent neural networks, IEEE International Conference on Acoustics, Speech and Signal Processing, Canada, 2013.
[32] W. Ye, Study of the Inverse Synthetic Aperture Radar Imaging and Motion Compensation, Xidian Univ., Xi'an, China, 1996.
[33] J. Chen, L. Du, H. He, Y.C. Guo, Convolutional factor analysis model with application to radar automatic target recognition, Pattern Recognit 87 (2019) 140–156.
[34] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 3rd International Conference on Learning Representations (ICLR), USA, 2015.
[35] F. Wang, M. Jiang, C. Qian, S. Yang, Residual attention network for image classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA, 2017.
[36] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio, Show, attend and tell: neural image caption generation with visual attention, in: Proceedings of the 32nd International Conference on Machine Learning (ICML), USA, 2015.
[37] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, Advances in Neural Information Processing Systems (NIPS), Australia, 2017.
[38] B. Su, S. Lu, Accurate Scene Text Recognition Based On Recurrent Neural Network, Springer International Publishing, 2014.
[39] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, Proceeding of the 27th International Conference on Machine Learning (ICML), Israel, 2010.
[40] B. Chen, Z.Bao H.W.Liu, Analysis of three kinds of classification based on different absolute alignment methods, Mod. Radar 28 (3) (2006) 58–62.
[41] D. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Proceedings of the International Conference on Machine Learning (ICML), China, 2014.