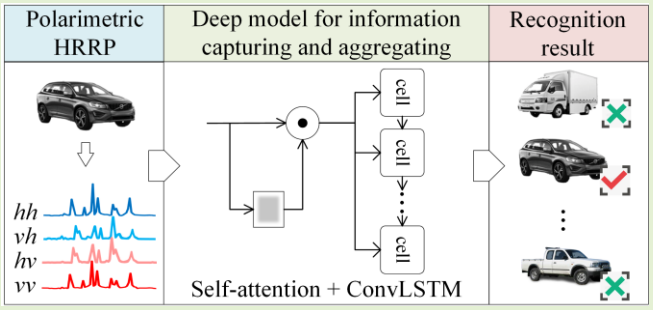
Polarimetric HRRP Recognition Based on ConvLSTM With Self-Attention

Liang Zhang, *Member, IEEE*, Yang Li, *Member, IEEE*, Yanhua Wang, *Member, IEEE*, Junfu Wang, and Teng Long, *Fellow, IEEE*

***Abstract*—Polarimetric high resolution range profile (HRRP) holds great potential for radar automatic target recognition (RATR) owing to its capability of providing both polarimetric and spatial scattering information. Conventional polarimetric HRRP recognition methods generally extract a set of features and put them into a well-trained classifier, which heavily rely on expertise. On the contrary, deep learning can automatically learn deep features of the training data and has obtained state-of-the-art results in many classification tasks. In this article, a novel deep model based on convolutional long short-term memory (ConvLSTM) network and self-attention mechanism is proposed for polarimetric HRRP recognition. In the proposed model, ConvLSTM employs the convolutional operator and recurrent LSTM structure to capture and aggregate target polarimetric and spatial scattering information from different dimensions simultaneously. Self-attention mechanism is introduced before ConvLSTM to make the model focus on the discriminative range cells and help improve the learning capacity of ConvLSTM. Experiment results on the simulated and measured datasets demonstrate the effectiveness of the proposed model for multi-class targets recognition. The results of two expanded experiments also show that this model performs well in noise and small training sample cases, showcasing good potentials in practical applications.**



***Index Terms*—High resolution range profile (HRRP), radar polarimetry, radar automatic target recognition (RATR), convolutional long short-term memory (ConvLSTM), self-attention mechanism.**

I. Introduction

HIGH resolution range profile (HRRP) represents the distribution of target scattering centers along radar line-of-sight (LOS), which provides target geometrical and structural characteristics in spatial dimension. Because of its low requirement on the acquisition, storage and processing, it plays an important role in radar automatic target recognition (RATR) community [1]–[6]. However, HRRP-based RATR is still challenging because its information is limited to range direction only. To improve the recognition performance, one effective approach is to integrate polarimetric information to reveal detailed target signatures, such as surface roughness and structure symmetry [7], [8].

Conventional polarimetric HRRP recognition methods focus on extracting highly discriminated features and designing an effective classifier. Numerous polarimetric feature extraction methods have been proposed in the past few decades. The most direct way is to utilize the backscattering coefficients or the parameters of the coherency/covariance matrix as a feature vector [9], [10]. To characterize target physical signatures, a series of descriptive features are extracted from the polarization scattering matrix (PSM) and the coherency matrix, such as polarimetric intercorrelation parameters [11] and roll-invariant features [12]. Another popular feature extraction way is based on the target polarimetric decomposition theory. Various polarimetric decomposition methods, including the well-known coherent [13]–[15] and incoherent decomposi-

Manuscript received November 15, 2020; accepted December 6, 2020. Date of publication December 14, 2020; date of current version February 17, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFE0202101 and Grant 2018YFE0202102, in part by the National Natural Science Foundation of China under Grant 61701026, and in part by the Youth Science and Technology Innovation Leader of National Innovation Talent Promotion Program under Grant 2013RA2034. The associate editor coordinating the review of this article and approving it for publication was Dr. Kagan Topalli. *(Corresponding author: Yang Li.)*

Liang Zhang and Teng Long are with the Radar Research Laboratory, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Beijing Key Laboratory of Embedded Real-Time Information Processing Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: zhangliang-bit@foxmail.com; longteng@bit.edu.cn).

Yang Li and Yanhua Wang are with the Radar Research Laboratory, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, also with the Beijing Key Laboratory of Embedded Real-Time Information Processing Technology, Beijing Institute of Technology, Beijing 100081, China, and also with the Chongqing Innovation Center, Beijing Institute of Technology, Chongqing 401120, China (e-mail: bit_liyang@bit.edu.cn; wyhlucky@bit.edu.cn).

Junfu Wang is with Beijing Racobit Electric Information Technology Company Ltd., Beijing 100081, China (e-mail: iam_wjf@126.com).

Digital Object Identifier 10.1109/JSEN.2020.3044314

tions [16]–[21], are proposed to interpret target scattering mechanisms. The weights and coefficients of the decomposition components are regarded as features and perform well in classification. Besides, similarity measurement [22], [23], dimensionality reduction [24], [25], statistical analysis [26] and many other techniques have also been explored for polarimetric feature extraction. For classifier design, with the development of machine learning, various classifiers, such as maximum likelihood classifier [27], support vector machine (SVM) classifier [28], random forest (RF) classifier [29] and neural networks classifier [30], [31] have been established or adopted for polarimetric feature classification and achieved impressive results. Nevertheless, conventional methods always require large effort from human experts, and the features and classifier designed for one operation scenario may be not still effective in another one [32].

In recent years, as a fast-developing technique, deep learning has achieved great success in many classification tasks [33]–[35]. In contrast with conventional methods, deep learning is an "end-to-end" learning approach that can automatically extract deep features and produce final classification results, and a well-designed deep model can be adopted for different operation scenarios by retraining itself. Many classic deep learning frameworks, such as autoencoder (AE), convolutional neural network (CNN) and recurrent neural network (RNN), have been studied for HRRP recognition. Since AE offers a powerful capability to extract high-level features, it is generally utilized as the feature extractor and cooperates with a classifier to complete recognition [36]–[38]. For CNN, HRRP data can be directly input into the model with one-dimensional convolution kernels for recognition [39]–[44]. Meanwhile, many developed techniques can be learned from computer vision community, making it the most widely studied framework at present. As for RNN, it received much attention in HRRP recognition because of its unique advantages for processing sequential data [45]–[47]. HRRP is regarded as a time series and fed into the network sequentially for modelling. In addition, the frameworks studied for HRRP recognition also include deep belief network [48], concatenated deep neural network [49], and so on. These above frameworks show impressive results in accuracy and flexibility. Naturally, it can be expected that deep learning can distinguish itself in polarimetric HRRP recognition. For multi-polarization HRRP, target high resolution information spreads along the range direction in spatial dimension, while target polarization information is distributed across multiple channels in polarimetric dimension. Therefore, the key to design a deep model is to make it feasible to capture and aggregate both high resolution and polarization information from different dimensions simultaneously.

In this article, a deep model based on convolutional long short-term memory (ConvLSTM) network [50] and self- attention mechanism [51] is proposed for polarimetric HRRP recognition. As a modification of RNN by integrating the long short-term memory (LSTM) and convolutional operator, ConvLSTM is employed to extract target feature representation from different dimensions simultaneously. Specifically, when the raw data is fed into the ConvLSTM, convolutional operator is applied in each time step to capture the scattering information of each range cell along polarimetric dimension. The LSTM network structure is used to aggregate such information together along spatial dimension to form target final feature representation. Considering that the range cells in the practical HRRP slice may have different discriminative abilities, self-attention mechanism is introduced before ConvLSTM to make the model focus on more discriminative range cells and help improve the learning capacity of ConvLSTM. Experiments are conducted on both simulated and measured polarimetric HRRP data, and the results verify the effectiveness of the proposed model for multi-class targets recognition. The results of two expanded experiments also show that our model performs well in noise and small training sample cases, showcasing good potentials in practical applications.

The reminder of this article is organized as follows. Section II briefly reviews the background of target polarimetric HRRP recognition, and introduces the self-attention mechanism and ConvLSTM network. In section III, the proposed deep model for polarimetric HRRP recognition are described in detail. The experiment results based on the simulated and measured data are provided in Section IV. Finally, Section V concludes the paper and gives a prospect of future works.

II. Background

A. Target Polarimetric HRRP Recognition

Scattering of a complex target is approximated to a sum of responses from local scattering sources at high frequency, which are known as scattering centers [52]. The scattering characteristic of a complex target can be described by the scattering center model approximately as follows [53]

$$x_l(t) = \sum_{i=1}^{I} S_l^i h\left(t - \frac{2r_i}{c}\right) \exp\left(-j\frac{4\pi r_i}{\lambda}\right), \tag{1}$$

where c denotes the light speed, λ is the signal wavelength, I is the scattering center numbers, $h(t)$ represents the high resolution impulse response function after pulse compression, r_i is the radial range of the i-th scattering center to the radar and S_l^i is its complex scattering coefficient in l polarization measurement mode ($l = hh$, vh, hv, vv, where h and v represent horizontally and vertically linear polarizations, respectively). The polarimetric scattering information of each scattering center can be described by the PSM, which is known as [54]

$$\mathbf{S}^i = \begin{bmatrix} S_{hh}^i & S_{hv}^i \\ S_{vh}^i & S_{vv}^i \end{bmatrix}. \tag{2}$$

A practical polarimetric HRRP sample from a vehicle target is illustrated in Fig. 1. The polarimetric projection of target scattering centers contain a wealth of scattering information in both spatial and polarimetric dimensions. Specifically, in spatial dimension, HRRP precisely reveals the amplitudes, phases and relative positions of target scattering centers along the radar LOS, which is called high resolution information. In polarimetric dimension, target scattering centers present different characteristics, which is known as polarimetric information. This rich information provides strong support for RATR but how to accurately capture and fully aggregate such information gives a challenge to recognition task. In addition,

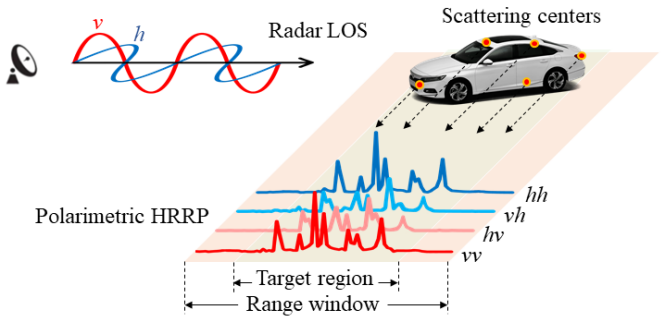


Fig. 1. Illustration of a practical polarimetric HRRP sample from a vehicle target.

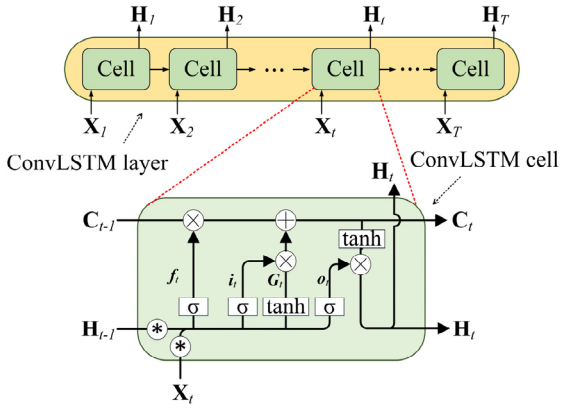


Fig. 2. The structure of ConvLSTM. "σ" and "tanh" are the sigmoid and TanHyperbolic function, respectively. Symbols "$\otimes$", "$\oplus$" and "$*$" represents Hadamard product, matrix addition, and convolutional operator, respectively.

practical HRRP slice generally depicts the radar echo from a certain range window around target region. Different range cells in range window contain different amounts of target scattering information, resulting in different discriminative abilities, which should be considered in HRRP recognition.

B. Self-Attention Mechanism

The essence of attention mechanism comes from human visual attention mechanism. When people perceive things, they often pay more attention to specific parts according to their needs. Inspired by this phenomenon, attentive neural networks were proposed by Mnih *et al.* [55] in computer vision for image classification.

With the widespread and successful application of attention mechanism in many fields, Vaswani *et al.* [51] further proposed a novel attention mechanism named self-attention for sequential data representation. The input to the self-attention mechanism is first mapped to a query matrix $\mathbf{Q}$, a key matrix $\mathbf{K}$ and a value matrix $\mathbf{V}$. These matrixes have the same dimension, and $\mathbf{K}$ and $\mathbf{V}$ have the same length. Then, an attention matrix is generated by computing the similarity between $\mathbf{Q}$ and $\mathbf{K}$ using the scaled dot product. Finally, using the attention matrix as weights, the output is a weighted sum over the value matrix $\mathbf{V}$. The output of the self-attention can be described as

$$\text{Attention}\left(\mathbf{Q},\mathbf{K},\mathbf{V}\right)=\text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{d}}\right)\mathbf{V}, \tag{3}$$

where "softmax" is softmax function which is performed to normalize the attention matrix to probability distribution, $\sqrt{d}$ is the scaling factor and d equals the length of the input data, controlling the dot product of $\mathbf{Q}$ and $\mathbf{K}$ not too large. For a sequential input, self-attention helps to provide different weights for different parts of the input by capturing relational indication information between two related elements. Recently, it has become a popular research topic in speech recognition [56] and remote sensing data classification [57], [58].

C. ConvLSTM Network

RNN has unique advantages for modelling sequential data because it employs recurrent nodes to connect the neuron to itself across time [59]. However, due to gradient vanishing, it is difficult for RNN to model long-range dependencies [60]. To address this issue, LSTM is proposed as an extension of the RNN by additionally utilizing a memory cell $\mathbf{C}_t$ and three gates: the forget gate $\boldsymbol{f}_t$, the input gate $\boldsymbol{i}_t$, and the output gate $\boldsymbol{o}_t$ [61]. The memory cell $\mathbf{C}_t$ is used to store foregone information and three gates are designed to control the updating of the memory cell. However, a shortcoming of LSTM is that it captures the relational information between all time steps while ignoring the internal information of each individual time step.

Considering this shortcoming, a modification of LSTM is developed, i.e., ConvLSTM [50]. The structure of ConvLSTM is depicted in Fig. 2. Same as LSTM, ConvLSTM layer holds a recurrent network structure and ConvLSTM cell employs a memory cell $\mathbf{C}_t$ and three gates $\boldsymbol{f}_t$, $\boldsymbol{i}_t$ and $\boldsymbol{o}_t$ to model long-term dependencies. But different from LSTM, the fully connected operators used for input-to-state and state-to-state transitions are replaced by convolutional operators in ConvLSTM cell. Due to this modification, ConvLSTM can learn the internal information of each individual time step. For the t-th time step in ConvLSTM, the value of the memory cell $\mathbf{C}_t$, three gates $\boldsymbol{i}_t, \boldsymbol{f}_t, \boldsymbol{o}_t$, and output $\mathbf{H}_t$ can be formally formulated as

$$\begin{aligned}
\boldsymbol{i}_t &= \sigma\left(\mathbf{W}_{xi} * \mathbf{X}_t + \mathbf{W}_{hi} * \mathbf{H}_{t-1} + \boldsymbol{b}_i\right)\\
\boldsymbol{f}_t &= \sigma\left(\mathbf{W}_{xf} * \mathbf{X}_t + \mathbf{W}_{hf} * \mathbf{H}_{t-1} + \boldsymbol{b}_f\right)\\
\boldsymbol{o}_t &= \sigma\left(\mathbf{W}_{xo} * \mathbf{X}_t + \mathbf{W}_{ho} * \mathbf{H}_{t-1} + \boldsymbol{b}_o\right)\\
\mathbf{G}_t &= \tanh\left(\mathbf{W}_{xg} * \mathbf{X}_t + \mathbf{W}_{hg} * \mathbf{H}_{t-1} + \boldsymbol{b}_g\right)\\
\mathbf{C}_t &= \boldsymbol{f}_t \circ \mathbf{C}_{t-1} + \boldsymbol{i}_t \circ \mathbf{G}_t\\
\mathbf{H}_t &= \boldsymbol{o}_t \circ \tanh\left(\mathbf{C}_t\right),
\end{aligned} \tag{4}$$

where "σ" is the sigmoid function and "tanh" is TanHyperbolic function, "$*$" and "$\circ$" represent convolutional operator and Hadamard product, respectively. $\mathbf{X}_t$ is the input of the current cell and $\mathbf{G}_t$ is a candidate memory cell for information transmission. In addition, $\mathbf{W}_{\sim}$ and $\boldsymbol{b}_{\sim}$ denote convolution kernels and bias terms, respectively. The subscripts of $\mathbf{W}$ and $\boldsymbol{b}$ have obvious meanings. For instance, $\mathbf{W}_{xo}$ is the input-output

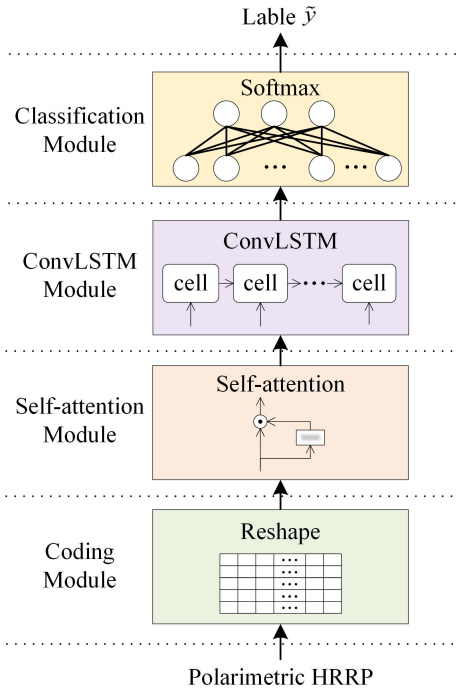


Fig. 3. Flowchart of the proposed model.

gate convolution kernel, and $\boldsymbol{b}_i$ is the input gate bias term etc. Ever since ConvLSTM network was first proposed for precipitation nowcasting [50], the idea of it has been employed in many other classification studies, such as gesture recognition [62] and hyperspectral image classification [63], [64].

III. Methodology

In this section, we design a deep model for polarimetric HRRP recognition based on ConvLSTM with self-attention. The framework of the proposed model is first illustrated. Then, the four basic modules contained in the model are demonstrated in detail. The working principle of ConvLSTM with self-attention for polarimetric HRRP data is analyzed combined with theoretic deduction. Finally, we give the training method.

A. Framework of Proposed Model

Fig. 3 shows the flowchart of the proposed model. The whole model consists of four modules: the coding module, the self-attention module, the ConvLSTM module and the classification module. Specifically, for a given polarimetric HRRP, it will be first coded as a real matrix with original amplitude and phase information. Then, the coded matrix is fed into the self-attention module to make the model focus on more discriminative range cells and generate a weighted represen- tation of raw polarimetric HRRP. Next, in the ConvLSTM module, a dual layer ConvLSTM network is designed to learn target feature representation from both spatial and polarimetric dimensions. Finally, a classification module is employed to predict the target label. The detailed description of each module is given as follows.

B. Coding Module

Suppose a fully polarimetric HRRP that can be expressed as

$$\left\{(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_i, \cdots, \mathbf{S}_n) \,|\mathbf{S}_i \in \mathbb{R}^{2\times 2}, 1 \le i \le n\right\}, \quad (5)$$

where n represents the length of HRRP, i.e. the range cell numbers, and $\mathbf{S}_i$ indicates the PSM of the i-th range cell. From (2), it can be known that the PSM $\mathbf{S}_i$ is a 2×2 complex matrix that can be further written as

$$\mathbf{S}_i = \begin{bmatrix} S^i_{hh} & S^i_{hv} \\ S^i_{vh} & S^i_{vv} \end{bmatrix} = \begin{bmatrix} |S^i_{hh}|\, e^{j\phi^i_{hh}} & |S^i_{hv}|\, e^{j\phi^i_{hv}} \\ |S^i_{vh}|\, e^{j\phi^i_{vh}} & |S^i_{vv}|\, e^{j\phi^i_{vv}} \end{bmatrix}, \quad (6)$$

where $|S^i_{\sim}|$ and $\phi^i_{\sim}$ represents the amplitude and phase of complex scattering coefficient $S^i_{\sim}$, respectively. The subscript "$\sim$" means polarization mode. Earlier works demonstrate that phase information helps improve classification results in radar RATR applications [65]–[67]. However, complex data cannot be directly used by the deep model. Thus, we code the complex matrix$\mathbf{S}_i$ as a real-vector $\boldsymbol{x}_i$ while preserving phase information as follows

$$\boldsymbol{x}_i = \left[\left|S^i_{hh}\right|,\ \phi^i_{hh},\ \left|S^i_{vh}\right|,\ \phi^i_{vh}, \left|S^i_{hv}\right|,\ \phi^i_{hv}, \left|S^i_{vv}\right|,\ \phi^i_{vv}\right]^{\mathrm{T}}, \quad (7)$$

where the superscript "T" means transposition operator. In this way, the fully polarimetric HRRP can be further coded as a real matrix $\mathbf{X}$ as follows

$$\begin{aligned}\mathbf{X} &= [\boldsymbol{x}_1 \boldsymbol{x}_2, \cdots, \boldsymbol{x}_i, \cdots, \boldsymbol{x}_n] \\ &= \begin{bmatrix} |S^1_{hh}| & |S^2_{hh}| & \cdots & |S^i_{hh}| & \cdots & |S^n_{hh}| \\ \phi^1_{hh} & \phi^2_{hh} & \cdots & \phi^i_{hh} & \cdots & \phi^n_{hh} \\ |S^1_{vh}| & |S^2_{vh}| & \cdots & |S^i_{vh}| & \cdots & |S^n_{vh}| \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi^1_{vv} & \phi^2_{vv} & \cdots & \phi^i_{vv} & \cdots & \phi^n_{vv} \end{bmatrix}. \end{aligned} \quad (8)$$

This coding method will not loss any information by preserving the original amplitude and phase information. We employ a reshape layer to implement it in this module. To a more general case, for a given polarimetric HRRP with l polarization channels and n range cells, the raw data will be coded into a real matrix whose size is $2l \times n$, and be fed into the self-attention module for weight assignment.

C. Self-Attention Module

Considering that different range cells in practical HRRP have different discriminative abilities, we build a self-attention module to make the model focus on more discriminative range cells while suppress unnecessary ones. The architecture of self-attention module is shown in Fig. 4. It shows that when the coded matrix $\mathbf{X} \in \mathbb{R}^{2l\times n}$ is fed into the module, three weight matrixes are first initialized to project $\mathbf{X}$ into three different matrixes: the query matrix $\mathbf{Q} \in \mathbb{R}^{n\times m}$, the key matrix $\mathbf{K} \in \mathbb{R}^{n\times m}$, and the value matrix $\mathbf{V} \in \mathbb{R}^{n\times m}$, which can be expressed as

$$\mathbf{Q} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_Q, \mathbf{K} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_K, \mathbf{V} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_V, \quad (9)$$

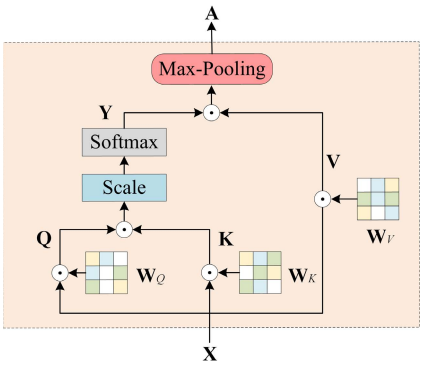


Fig. 4. Architecture of the self-attention module.

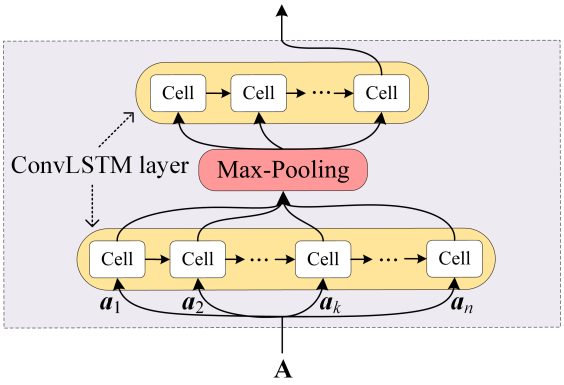


Fig. 5. Architecture of the ConvLSTM module.

where $\mathbf{W}_Q$, $\mathbf{W}_K$, $\mathbf{W}_V \in \mathbb{R}^{2l\times m}$ are three learnable weight matrixes and m represents the length of weight matrix which is a settable parameter when building the network.

Then, the self-attention weight matrix $\mathbf{Y} \in \mathbb{R}^{n\times n}$ is computed by the scaled dot product of matrix $\mathbf{Q}$ and $\mathbf{K}$, which can be formulated as follows

$$\mathbf{Y} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{n}}\right), \tag{10}$$

where $\sqrt{n}$ plays a regulating role to scale the dot product result and softmax function is performed row-wised. The elements in $\mathbf{Y}$ are the learned weights that stress the importance difference of range cells by increasing the values of significant ones.

Finally, we apply the self-attention weight matrix $\mathbf{Y}$ to refine the value matrix $\mathbf{V}$ as follows

$$\mathbf{A} = \mathbf{Y}\mathbf{V}, \tag{11}$$

where the values in $\mathbf{V}$ are weighted by corresponding weights in $\mathbf{Y}$, generating the weighted representation $\mathbf{A} \in \mathbb{R}^{n\times m}$ of polarimetric HRRP. In addition, a max-pooling layer is employed at the end of this module to reduce the data dimension. The length of $\mathbf{A}$ will be reduced from m to $2l$ to match the length of input matrix $\mathbf{X}$. With the weighted representation $\mathbf{A}$ obtained, a ConvLSTM module is designed to learn target feature representation from it.

D. ConvLSTM Module

The architecture of ConvLSTM module is demonstrated in Fig. 5. For polarimetric HRRP data, there generally only exists a limited number of available training samples, so it is difficult to train very deep networks. Thus, we apply two ConvLSTM layers to construct the ConvLSTM module and employ a max-pooling layer between the two ConvLSTM layers to reduce the data dimension and computation complexity. We adopt tanh function in each layer to learn the nonlinear representation. The kernel size of each ConvLSTM layer is $1 \times l$ where l represents the polarization channel numbers of the raw data, and the filter numbers of the first and second layer are 16 and 32, respectively.

When the weighted representation $\mathbf{A} \in \mathbb{R}^{n\times 2\ l}$ is fed into the ConvLSTM module, it will be split into an n-length sequence $\{(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n) |\ \boldsymbol{a}_k \in \mathbb{R}^{1\times 2l},\ 1 \le k \le n\}$ across the spatial dimension. The k-th element $\boldsymbol{a}_k$ in the sequence represents the weighted representation of the k-th range cell. These elements are fed into the ConvLSTM cells in the first layer one by one to capture and aggregate both the high resolution and polarization information. Specifically, for the k-th element $\boldsymbol{a}_k$, ConvLSTM cell accomplishes this in three steps.

Step 1: ConvLSTM cell calculates the values of three gates $\boldsymbol{i}_k$, $\boldsymbol{f}_k$ and $\boldsymbol{o}_k$ based on $\mathbf{H}_{k-1}$ and $\boldsymbol{a}_k$. $\mathbf{H}_{k-1}$ is the output of the previous ConvLSTM cell, containing the high resolution information captured from $\boldsymbol{a}_1$ to $\boldsymbol{a}_{k-1}$. The polarization information of $\boldsymbol{a}_k$ will be captured by the three gates via convolutional operators. Take the input gate $\boldsymbol{i}_k$ as an example, the polarization information representation $\mathbf{F}_i$ can be formulated as follows

$$\mathrm{F}_i = f\left(\mathbf{W}_i * \boldsymbol{a}_k + \boldsymbol{b}_i\right), \tag{12}$$

where $\mathbf{W}_i$ and $\boldsymbol{b}_i$ are convolution kernel and bias term of the input gate $\boldsymbol{i}_k$, "*" is the convolutional operator, and f is the activation function.

Step 2: ConvLSTM updates the current memory cell $\mathbf{C}_k$. This includes two parts: first, ConvLSTM decides what information to throw away from the previous memory cell $\mathbf{C}_{k-1}$ via the forget gate $\boldsymbol{f}_k$. Second, ConvLSTM decides what new information to store via a candidate memory cell$\mathbf{G}_k$ and the input gate $\boldsymbol{i}_k$. After finishing these two parts, ConvLSTM adds them together to update $\mathbf{C}_k$. The purpose of this step is to remove the useless information of previous range cells and add the new polarimetric information of current range cell to it.

Step 3: ConvLSTM decides what information to output via the memory cell $\mathbf{C}_k$ and output gate $\boldsymbol{o}_k$.

The above three steps can be formulated as the following equations:

$$\begin{aligned}
&\text{input gate}: \boldsymbol{i}_k = \sigma\left(\mathbf{F}_i + \mathbf{W}_{hi} * \mathbf{H}_{k-1} + \boldsymbol{b}_i\right)\\
&\text{forget gate}: \boldsymbol{f}_k = \sigma\left(\mathbf{F}_f + \mathbf{W}_{hf} * \mathbf{H}_{k-1} + \boldsymbol{b}_f\right)\\
&\text{output gate}: \boldsymbol{o}_k = \sigma\left(\mathbf{F}_o + \mathbf{W}_{ho} * \mathbf{H}_{k-1} + \boldsymbol{b}_o\right)\\
&\text{candidate memory cell}:\\
&\quad \mathbf{G}_k = \tanh\left(\mathbf{W}_{ag} * \mathbf{a}_k + \mathbf{W}_{hg} * \mathbf{H}_{k-1} + \boldsymbol{b}_g\right)\\
&\text{memory cell}: \mathbf{C}_k = \boldsymbol{f}_k \circ \mathbf{C}_{k-1} + \boldsymbol{i}_k \circ \mathbf{G}_k\\
&\text{output}: \mathbf{H}_k = \boldsymbol{o}_k \circ \tanh\left(\mathbf{C}_k\right),
\end{aligned} \tag{13}$$

where $\mathbf{F}_{\sim}$ represents the polarization information representation captured in three gates, $\mathbf{W}_{\sim}$ and $\boldsymbol{b}_{\sim}$ denote convolution kernels and bias terms. The subscripts of $\mathbf{F}$, $\mathbf{W}$ and $\boldsymbol{b}$ indicate where they are used.

After flowing though the two ConvLSTM layers, the output of the last ConvLSTM cell is adopted as the final feature representation to be fed into the next module for recognition.

E. Classification Module

In the classification module, we employ a fully connected layer and a softmax layer to obtain recognition result. Assuming that vector $\boldsymbol{h}$ is the output of fully connected layer, the probability of each class that the polarimetric HRRP belongs to can be formulated as

$$\boldsymbol{p}\left(y|\boldsymbol{h}\right) = \text{softmax}\left(W_s\boldsymbol{h} + \boldsymbol{b}_s\right), \tag{14}$$

where $\mathbf{W}_s$ and $\boldsymbol{b}_s$ are the weight matrix and bias term of the softmax layer, and y denotes an arbitrary label vector. The predicted target label $\tilde{y}$ with highest probability value is identified as follows

$$\tilde{y} = \arg\max_{y} \boldsymbol{p}\left(y|\boldsymbol{h}\right). \tag{15}$$

F. Network Training

In the training stage, we use cross entropy as our loss function. Given a training set consisting of N samples, the objective function is given by

$$\mathcal{J} = \sum_{i=1}^{N} y^i \log \tilde{y}^i \tag{16}$$

where y^i is the ground-truth label of the i-th training sample. The model is trained by minimizing (16), and we optimize the model through adaptive moment estimation (Adam) algorithm [68] and back-propagation method over the training set.

IV. Experiment

In this section, we build two polarimetric HRRP datasets for experiment by simulation and measurement. The experimental settings and results are given following the datasets. Two expend experiments are conducted at the end of this section to verify the effectiveness of the proposed model in noise and small training sample cases.

A. Datasets Preparation

Unlike RGB images in the computer vision community, which can be easily collected, polarimetric HRRP is much more difficult to obtain due to its big costs in both manpower and material. Since there are non-existent public polarimetric HRRP datasets in recognition community at present, we build two datasets including a simulated and a real dataset for experiments in this part. The simulated dataset considers a complex recognition scenario where HRRPs are collected from 10 civilian vehicles with 360° azimuth and 4 elevation angles. It is generated from the full polarized scattering data released by the Air Force Research Laboratory (AFRL) [69]. The real dataset considers a relatively simple recognition scenario where HRRPs are collected from 5 civilian vehicles with 7 different azimuth angles. It is collected by our own dual polarimetric radar platform. The detailed descriptions of these two datasets are given in the following parts.

TABLE I
Training and Testing Data Used in the Simulated Data Experiment

	Train	Test
Target number	10	10
Azimuth angle	1~359° (interval of 2°)	2~360° (interval of 2°)
Elevation angle	30°/40°/50°/60°	30°/40°/50°/60°
Sample number	7200 [a]	7200

[a]7200 = 10×180×4, i.e. target number × azimuth angle number × elevation angle number.

1) Full Polarized Simulated Dataset: The X-band scattering data released by the AFRL contains ten different civilian vehicles: Camry, Honda Civic 4dr, Jeep93, Jeep99, Maxima, Mazda MPV, Mitsubishi, Sentra, Toyota Avalon and Toyota Tacoma. The simulation is performed using the CAD models of targets as well as ground plane with 360 degrees azimuth and four common elevation angles of 30°, 40°, 50° and 60°.

In our experiments, we select the data with the azimuth angle of [1°, 360°] equally spaced by 1° and the elevation angle of 30°, 40°, 50° and 60° to generate HRRP samples, resulting a total of 1440 samples per target. The center frequency and synthetic bandwidth of the selected data are 9.6GHz and 1.5GHz, respectively. The samples of generated polarimetric HRRP for each target are shown in Fig. 6. Further, in the simulated data experiment, these samples are partitioned into training and testing data according to whether their azimuth angle is odd or even. The purpose of this partition is to ensure the comprehensiveness of training and testing data while reserving their differences for generalization performance evaluation. As a result, both of training and testing data contain 10 targets of a total of 7200 polarimetric HRRP samples as shown in Table I.

2) Dual Polarized Real Dataset: The dual polarized real dataset contains five classes of civilian vehicles which are common in our daily life: truck, pick-up, sedan, van and minibus. The dataset is collected by a dual polarimetric wideband radar, which transmits a H linear polarization and receives H and V orthogonal polarizations simultaneously. The radar system operates at W-band and uses stepped-frequency linear frequency modulated signal with a synthetic bandwidth of 1250MHz to generate target HRRP.

The five targets are measured in two outfield scenarios where the radar platform is set at 50m above the ground and about 800m from the target. Specifically, in the first scenario, the polarimetric HRRP samples of each vehicle are collected at 7 azimuth angles with the interval of 45° in static state as shown in Fig. 7(a). The number of polarimetric HRRP samples at each azimuth angle is 500. In the second scenario, the samples are collected when each vehicle moves following a circular path in dynamic state as shown in Fig. 7(b), and a total of approximately 300 samples per target are acquired in this process. The dual polarimetric HRRP samples of five

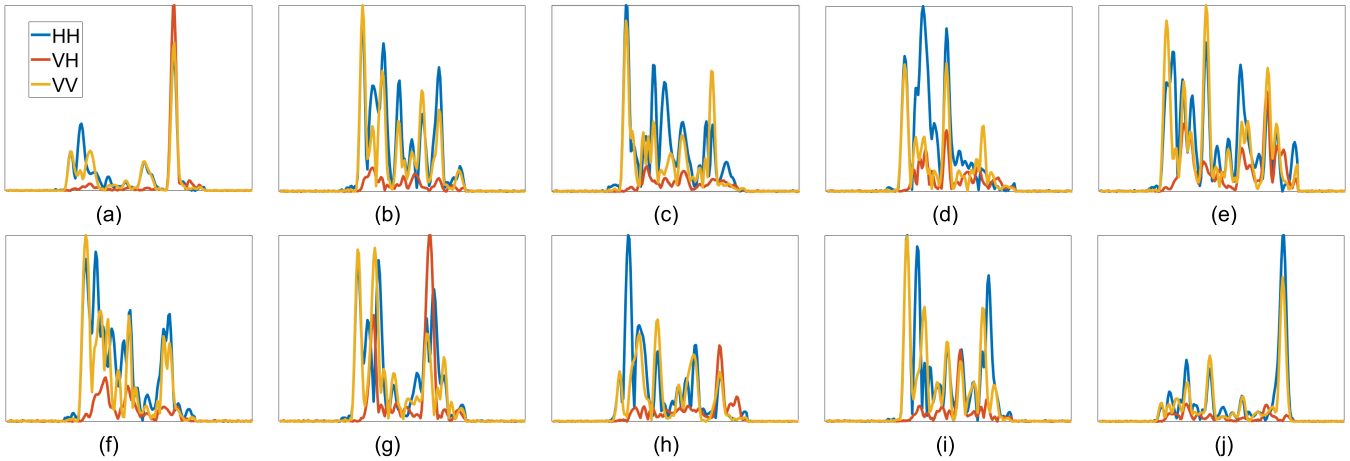


Fig. 6. Ten HRRP samples in full polarized simulated dataset: (a) Camry; (b) Honda Civic 4dr; (c) Jeep93; (d) Jeep99; (e) Maxima; (f) Mazda MPV; (g) Mitsubishi; (h) Sentra; (i) Toyota Avalon; (j) Toyota Tacoma. The blue, red and yellow line represent HH, VH and VV polarization, respectively.

TABLE II
TRAINING AND TESTING DATA USED IN THE SOC AND EOC EXPERIMENTS

Class	Train for SOC and EOC			Test for SOC			Test for EOC		
	Target state	Azimuth angle	Number	Target state	Azimuth angle	Number	Target state	Azimuth angle	Number
1	Static	45°~315° (interval of 45°)	2332	Static	45°~315° (interval of 45°)	1167	Dynamic	0°~360° (uniform distribution)	319
2			2334			1166			314
3			2332			1167			305
4			2334			1166			279
5			2332			1167			309

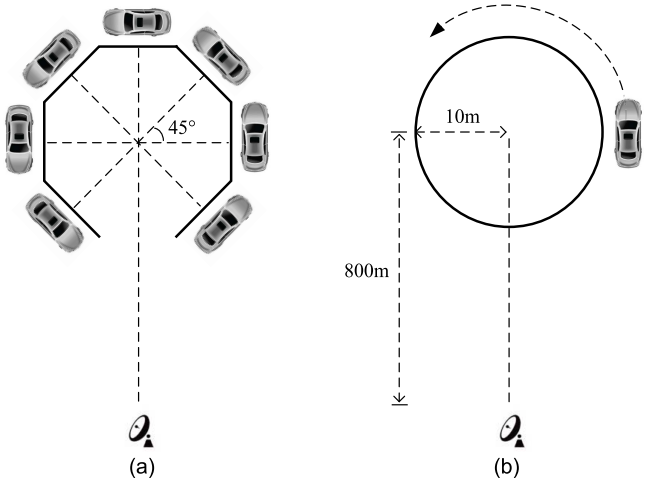


Fig. 7. Two scenarios for real data collection. (a) Scenario for static data collection; (b) Scenario for dynamic data collection.

vehicles and their corresponding optical images are given in Fig. 8.

Based on the collected data, we conduct our experiments under the standard operating condition (SOC) and extended operating condition (EOC) by reference to the MSTAR dataset [70], which is a famous public SAR image dataset used for target recognition. In the SOC experiment, we make the training samples and testing samples have the same observation condition. Thus, the data collected in the first scenario [Fig. 7(a)] are used for both training and testing, where 2/3 of them are used as training data and 1/3 of them are used as testing data. In the case of EOC, we make the training samples and testing samples have a major difference in observation condition. We use the data collected in the first scenario [Fig. 7(a)] for training and the data collected in the second scenario [Fig. 7(b)] for testing, thus azimuth angle and motion state of targets are much different. The variation of observation condition has a huge impact on the performance of HRRP-based recognition algorithm. A slight change of the observation condition may cause a serious decrease in accuracy. Therefore, the EOC experiment is useful for testing the robustness of the proposed model. The detailed configurations of training and testing data of SOC and EOC experiments are shown in Table II.

3) Polarimetric HRRP Preprocessing: Target-aspect, time-shift and amplitude-scale sensitivities are three issues which should be considered in HRRP-based RATR [2]. Thus, we adopt preprocessing operations in our experiment to reduce their effects. Firstly, in the dataset building stage, target polarimetric HRRP data are collected in as many azimuth angles as possible to reduce the effect of target-aspect sensitivity. Secondly, we uniformly cut out 200 range cells from each original HRRP, and align the target region in the range window to tackle the time-shift sensitivity. Finally, each HRRP sample is normalized by dividing its L_2-norm to deal with the amplitude-scale sensitivity [16].

B. Experimental Settings

1) Parameter Settings of Proposed Model: The detailed configurations and parameter settings of the proposed model are listed in Table III, where l represents the polarization channel

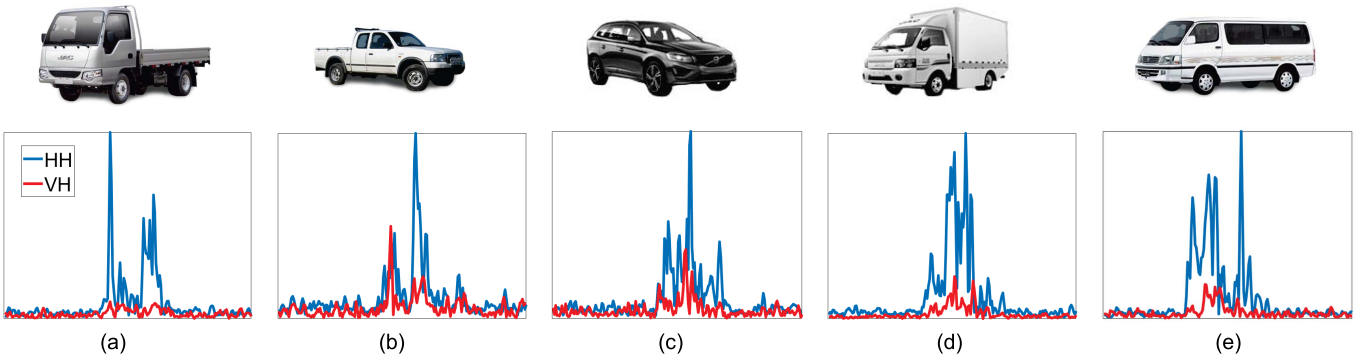


Fig. 8. Optical images and dual polarimetric HRRP samples of five different vehicles: (a) truck; (b) pick-up; (c) sedan; (d) van; (e) minibus. The blue and red line represent HH and VH polarization, respectively.

TABLE III
CONFIGURATIONS AND PARAMETER SETTINGS OF THE PROPOSED MODEL

Module	Layer		Filter number @size	Output size
Coding	L1	Reshape		200×(2*l*)
Self-attention	L2	Self-attention		200×(60*l*)
	L3	Max-pooling	@1×30	200×(2*l*)
ConvLSTM	L4	ConvLSTM	16@1×*l*	200×(2*l*)×16
	L5	Norm		200×(2*l*)×16
	L6	Max-pooling	@2×1×1	100×(2*l*)×16
	L7	ConvLSTM	32@1×*l*	1×(2*l*)×32
	L8	Norm		1×(2*l*)×32
Classification	L9	Fully connected		1×128
	L10	Softmax		1

numbers of the input data. Specifically, l equals 2 for dual polarization and 3 for full polarization under the assumption of scattering reciprocity ($S_{hv} = S_{vh}$). In the self-attention module, the parameter m which represents the length of weight matrixes $\mathbf{W}_Q$, $\mathbf{W}_K$, $\mathbf{W}_V$ as mentioned in (9) is set to $60l$, and a max-pooling layer of size 1×30 is adopted to reduce the output size to $200 \times 2l$. The ConvLSTM module has two ConvLSTM layers. The kernel size of each ConvLSTM layer is $1 \times l$, and the kernel numbers of the two layers are set to 16 and 32, respectively. In addition, each ConvLSTM layer is followed by a batch normalization layer for fast convergence in training phase, and a max-pooling layer of size $2 \times 1 \times 1$ is set between the two ConvLSTM layers to reduce data dimension and computational complexity. For the fully connected layer in the classification module, the output node numbers are 128.

2) Comparative Methods: In order to effectively analyze the performance of the proposed model, we implement two different kinds of comparative methods in experiment. The first one is to evaluate the effectiveness of ConvLSTM network and self-attention mechanism. Five models with different architectures are configured as comparative methods which are shown in Table IV. Specifically, model A (abbreviated as SA-CLSTM) is our proposed model, and models B to F (abbreviated as CLSTM, SA-LSTM, LSTM, SA-CNN and CNN, respectively) are five comparative models. In model C and model E, we keep the self-attention module but replace the ConvLSTM layer with fully connected LSTM layer and CNN layer respectively to compare the effectiveness of ConvLSTM for feature extraction. Model B, D and F which are derived from model A, C and E respectively, are configured without self-attention module to compare the effectiveness of self-attention mechanism. The second one is to compare the proposed deep model with conventional recognition method. In conventional method, 16 polarimetric features including polarimetric scattering entropy (H), scattering angle (α), seven Sinclair matrix similarity features, and seven Mueller similarity features [17], [23], and 6 high resolution features including target length, sparsity, energy, variance, scattering center number and the first central moment [71]–[73] are extracted as a feature vector. An SVM classifier and a RF classifier are established for feature classification. Since these polarimetric features are designed for full polarized data, we only perform conventional method on the full polarized simulated dataset for comparison.

To quantitatively evaluate the performance of each model, we adopt the overall accuracy (OA), the per-class accuracy (PA), training time and the confusion matrix as indicators. In addition, all the experiments are performed with TensorFlow codes on a computer with 12 GB RAM and a NVIDIA GeForce GTX1080Ti graphic card.

C. Results and Analysis

1) Results of Simulated Data Experiment: Table V shows the confusion matrix, PA and OA of the proposed model on full polarized simulated dataset. Each column in the confusion matrix denotes the actual target class, and each row represents the class predicted by our proposed model. The PA and OA values are given in the last column and last row of the table, respectively. It can be observed that the proposed model obtains an OA of 72.14% over ten targets which shows a good performance on multi-class classification. The PA of Tacoma achieves the maximum of 85.42% among the ten targets, and the PAs of Jeep93 and Jeep99 are the two minimum values: 66.39% and 65.28%. This may be caused by the fact that Tacoma is a pick-up whose scattering characteristic is much different from the other vehicles, but Jeep93 and Jeep99 are both jeeps with no significant difference that are different to distinguish.

The comparison results with different methods are given in Table VI. From this table, we can observe that most deep models perform better than the conventional method

TABLE IV
CONFIGURATIONS OF THE PROPOSED AND FIVE COMPARATIVE MODELS

Layer	A	B	C	D	E	F
	SA-CLSTM	CLSTM	SA-LSTM	LSTM	SA-CNN	CNN
L1	Reshape					
L2	Self-attention		Self-attention		Self-attention	
L3	Max-pooling		Max-pooling		Max-pooling	
L4	ConvLSTM-16@1×l		LSTM-16@1×l		Conv-16@l×l	
L5	Normalization					
L6	Max-pooling					
L7	ConvLSTM-32@1×l		LSTM-32@1×l		Conv-32@l×l	
L8	Normalization					
L9	FC-128					
L10	Softmax					

Fully connected layer is represented as "FC-(Number of units)", and the ConvLSTM, LSTM and convolution layers are represented as "ConvLSTM, LSTM and Conv-(number of filters)@(filter size)".

TABLE V
RECOGNITION RESULT FOR THE SIMULATED DATA EXPERIMENT

	Camry	Civic 4dr	Jeep93	Jeep99	Maxima	MPV	Mitsubishi	Sentra	Avalon	Tacoma	PA (%)
Camry	496	3	11	13	101	5	14	60	5	12	68.89
Civic 4dr	2	488	10	58	3	37	19	8	75	20	67.78
Jeep93	11	17	478	95	8	54	17	5	23	12	66.39
Jeep99	5	38	73	470	3	67	23	11	21	9	65.28
Maxima	106	10	1	1	515	8	3	39	11	26	71.53
MPV	6	37	36	56	2	515	17	16	20	15	71.53
Mitsubishi	4	41	15	18	1	17	598	4	12	10	83.06
Sentra	40	7	20	19	76	15	6	515	1	21	71.53
Avalon	6	95	18	41	9	23	7	3	504	14	70.00
Tacoma	6	23	17	6	14	14	11	12	2	615	85.42
OA (%)											72.14

TABLE VI
COMPARISON OF DIFFERENT METHODS ON SIMULATED DATASET

Method		OA (%)	Training time (h)
A	SA-CLSTM	72.14	2.12
B	CLSTM	63.39	1.81
C	SA-LSTM	52.49	2.85
D	LSTM	50.99	2.47
E	SA-CNN	63.99	0.60
F	CNN	45.43	0.45
Features + SVM		42.23	0.35
Features + RF		50.34	0.14

TABLE VII
RECOGNITION RESULT FOR THE SOC EXPERIMENT

Target	1	2	3	4	5	PA (%)
1	1166	0	1	0	0	99.91
2	0	1165	0	0	1	99.83
3	0	0	1148	0	19	98.37
4	0	0	0	1166	0	99.91
5	1	2	27	0	1137	97.43
OA (%)						99.13

based on feature extraction and classifier. This result indicates that the high resolution and polarimetric features designed in [17], [23], [71]–[73] are not distinguishable enough for ten targets recognition on this dataset. For the deep models, the proposed SA-CLSTM obtains the highest OA. The models with ConvLSTM achieve better performance than the models with LSTM or CNN, which certifies the effectiveness of the ConvLSTM for capturing and aggregating target polarization and high resolution information from different dimensions simultaneously. In addition, the models with self-attention module (SA-CLSTM, SA-LSTM and SA-CNN) can further improve the recognition result. It shows that self-attention can help enhance the learning capacity of ConvLSTM.

We also compare the training time of these models on simulated dataset in Table VI. The conventional recognition methods (SVM and RF) have the shortest training time because their input is a set of extracted feature vectors but not raw HRRP data. For the deep models, it can be seen that the models with CNN (CNN and SA-CNN) cost relatively less time for training, but the models with LSTM (SA-LSTM and LSTM) cost the longest time. This is because LSTM structure needs more time for feature learning than CNN structure. In addition, it can be seen that training time of CLSTM structure is between them and self-attention module will increase time cost. In general, it is an acceptable efficiency for the proposed model to complete the training process within a few hours.

2) Results of Real Data Experiment: The real data experiment includes two sub-experiments: the SOC experiment and the EOC experiment, and the recognition results of the proposed model are demonstrated in Table VII and Table VIII. From Table VII, we can observe that the OA of the SOC experiment reaches 99.13% and only 51 test samples are

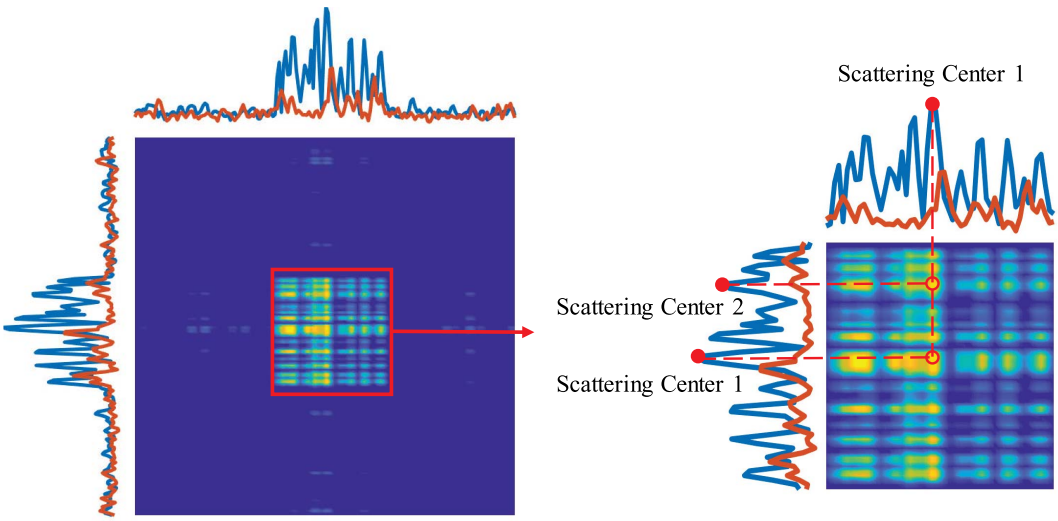


Fig. 9. Visualization result of the self-attention weight matrix **A** obtained by SA-CLSTM. Values of the weights are normalized for a better visualization.

TABLE VIII
RECOGNITION RESULT FOR THE EOC EXPERIMENT

Target	1	2	3	4	5	PA (%)
1	276	0	35	0	8	86.52
2	2	230	29	6	47	72.10
3	8	0	274	0	23	85.89
4	63	19	10	182	5	57.05
5	3	0	24	0	282	88.40
OA (%)						81.52

TABLE IX
COMPARISON OF DIFFERENT MODELS ON REAL DATASET

Model		OA of SOC experiment (%)	OA of EOC experiment (%)	Training time (h)
A	SA-CLSTM	99.13	81.52	1.82
B	CLSTM	99.07	61.21	1.57
C	SA-LSTM	98.30	70.45	2.66
D	LSTM	97.62	64.29	2.31
E	SA-CNN	94.02	54.00	0.58
F	CNN	87.35	38.60	0.39

misclassified. Meanwhile, the PAs are all more than 97%, and the PAs of target 1, 2 and 4 are even more than 99%. These results suggest that the proposed model can achieve very good recognition performance on SOC. From Table VIII, it can be observed that the OA of the EOC experiment is 81.52%, which is lower than that of the SOC experiment, and the PA of each class also decreases compared to SOC. It is in line with expectations that the change of the observation condition between the training and testing data in EOC causes a serious decrease in accuracy.

Table IX reports the recognition results acquired by six deep models for comparison. We can observe that the proposed model SA-CLSTM obtains the best performance in both SOC and EOC. Although all these six models perform well in SOC, they show a significant performance difference in EOC. The models which adopt ConvLSTM for feature extraction are better than the models which adopt LSTM or CNN. In addition, the self-attention module enhances the recognition capability once again. For example, the OA of CLSTM is improved from 63.39% to 72.14% by using self-attention mechanism. These results indicate that the proposed model based on ConvLSTM with self-attention can not only perform well in SOC, but also has a strong generalization ability for recognition in EOC, showcasing good potentials in practical applications. The training time of these models on real dataset is also provided in Table IX. The training time of these models follows the same trend but costs relatively less compared with simulated dataset.

3) Visualization of Self-Attention Mechanism: In order to intuitively view the impact of self-attention mechanism on recognition, we take a dual polarimetric HRRP sample of target 1 as an example to visualize the self-attention weight matrix **Y**. The matrix **Y** is obtained by SA-CLSTM and plotted in Fig. 9. Form the figure, it can be observed that the weights of target region are much bigger than the others, which indicates that the self-attention mechanism "thinks" this part is more important for recognition and pays more attention to it. For a clearer observation, we truncate target region and show it on the right. We can further observe that the weights of strong scattering centers have relatively larger values, such as the two weights between strong scattering center 1 and 2. This result means that the range cells with strong scattering centers in target region have more discriminative abilities for recognition.

4) Visualization of Feature Extraction: In this part, we visualize the deep feature representation acquired by these six different models to compare their learning capability. The output of fully connected layer (L9), which is a feature vector of 128 dimensions, is extracted from the testing data in the EOC experiment. We use t-distributed stochastic neighbor embedding (t-SNE) [74] to reduce its dimensionality from 128 dimensions to two dimensions in order to depict it in a 2-D graph. The results are shown in Fig. 10, where different colors denote different classes of target. From this figure, it can be observed that the feature maps extracted by

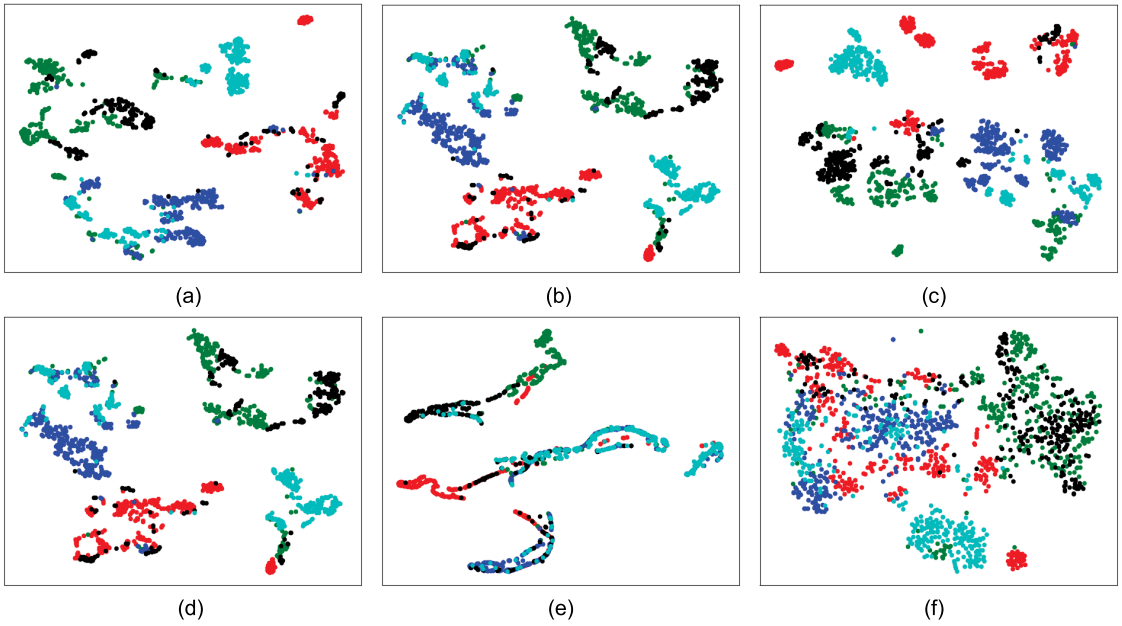


Fig. 10. Two-dimensional t-SNE projection of the extracted feature vectors. Different colors denote different classes of target. Features are extracted by: (a) SA-CLSTM; (b) SA-LSTM; (c) SA-CNN; (d) CLSTM; (e) LSTM; (f) CNN.

the models which have self-attention module [Fig. 10(a)-(c)] are more distinguishable than those of the models without self-attention module [Fig. 10(d)-(f)]. Furthermore, the feature extracted by SA-CLSTM [Fig. 10(a)] for each class are more widely separated from other classes, which indicates it has stronger learning capability. This is in line with the experiment results that the SA-CLSTM achieves higher OA than the others.

D. Performance With Different SNRs

Practical HRRP data often contain background noise and clutter which have negative effects on recognition performance. In this part, we evaluate their effects on the proposed model based on the both simulated and real datasets. Since deep learning is an "end-to-end" learning approach without requiring noise and clutter statistical distributions, without loss of generality, we take simulated Gaussian white noise as a representative for evaluation. The SNR is defined as [2]

$$\begin{aligned} \mathrm{SNR} &= 10 \times \log_{10}\left(\frac{\bar{P}_x}{P_{Noise}}\right) \\ &= 10 \times \log_{10}\left(\frac{\sum_{l=1}^{L} P_{x_l}}{L \times P_{Noise}}\right), \end{aligned} \tag{17}$$

where $\bar{P}_x$ and P_{Noise} denote the average power of target and noise respectively, x denotes range cells which are taken up by target, and L denotes the number of x. In the experiment, Gaussian white noise within the range 0-30dB with 5dB step is added to the both training and testing data of the simulated dataset. Because the real data are measured under high SNR (SNR>30dB), the noise within the range 0-25dB with 5dB step is added to the real dataset for evaluation. The proposed model (SA-CLSTM) is compared with other five models (CLSTM, SA-LSTM, LSTM, SA-CNN and CNN). Particularly, we additionally compared our model with a state-of-the-art noise-robust HRRP recognition method (sparse-low-rank representation method, SLRR) [75] to verify its effectiveness and advantage. For SLRR method, Polarization Whitening Filter (PWF) [76], [77] is used to synthesize multi-polarimetric information together before recognition.

Fig. 11 depicts the overall accuracy versus SNR. It can be observed that the overall accuracy decreases with increasing noise level of all tested methods, and the proposed method shows good performance on the both datasets. Specifically, under the complex simulated dataset [Fig. 11(a)], the overall accuracy of SA-CLSTM is close to 70% when SNR=10dB, and this value subsequently exceeds 70% as the SNR increases. The proposed method performs even better on the relatively simple real dataset [Fig. 11(b)]. When SNR=5dB, the overall accuracy of the proposed method is close to 90%, and this value increases to 95% for SNR=10dB. Furthermore, the proposed method exhibits better recognition performance than the noise-robust SLRR method and achieves the highest overall accuracy when SNR>5dB. The results show that the proposed method can achieve satisfactory recognition performance under low SNR conditions and is valuable for practical applications. In addition, the models which adopt self-attention mechanism perform better than the models without it in the same SNR condition. This result indicates that self-attention mechanism is helpful to improve the anti-noise performance. Combining the visualization of the self-attention weight matrix in previous part, we can reasonably infer that self-attention mechanism makes the model pay more attention to the range cells with strong scattering centers, and strong scattering centers remain stable even in the presence of background noise

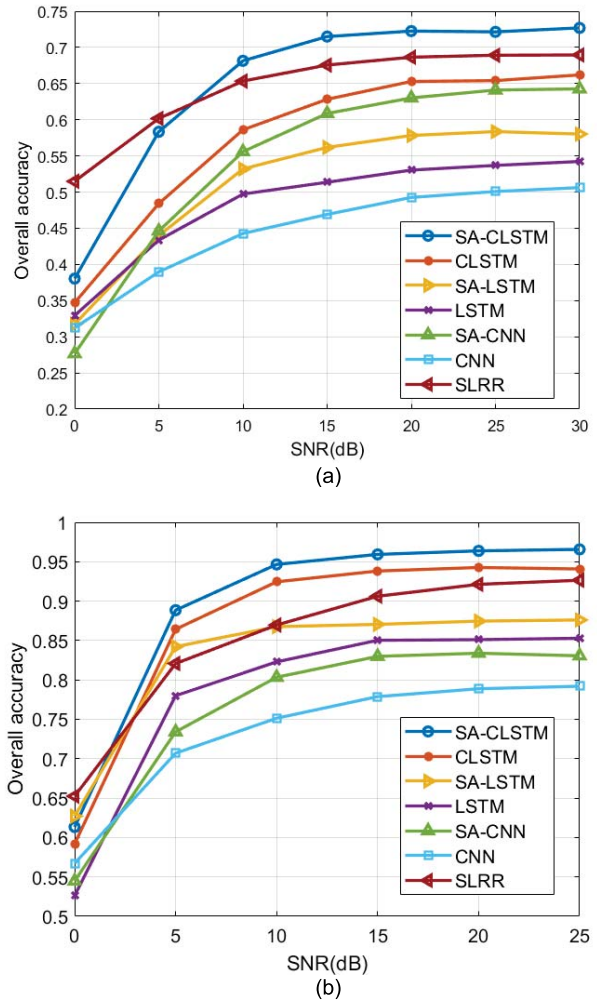


Fig. 11. Relationship between SNR and overall accuracy. (a) Results on the simulated dataset; (b) Results on the real dataset.

and clutter, thus reducing the effects of noise and clutter on recognition performance.

E. Performance With Different Numbers of Training Samples

In general, the training samples are limited in practical RATR applications. Many algorithms which needs a large number of training samples will be significantly affected in this situation. In this part, we evaluate the proposed model's performance when trained with different amounts of samples. The experiment is conducted based on the dual polarized real dataset, where the testing data we used are the same as SOC testing data, but the training samples are randomly selected from the SOC training data. The number of training samples per class is in the 50-300 range.

The relationship between error recognition rate and the number of training HRRP samples per class is plotted in Fig. 12. It can be observed that SA-CLATM's error rate curve is lower than that of the other five comparative models, and generates the lowest error rate in every training sample number. SA-CLATM achieves greater accuracies faster as training samples increasing. When the number of training samples per

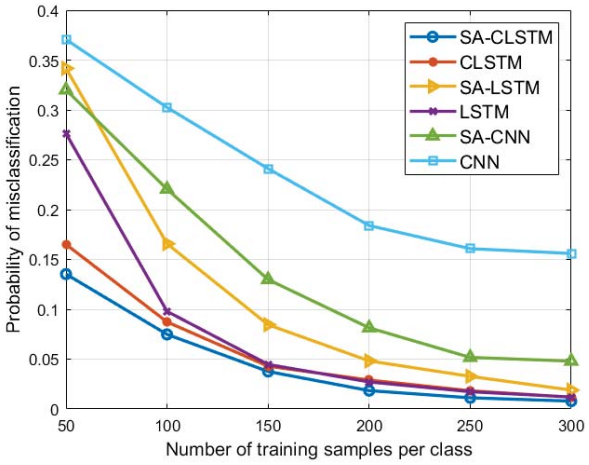


Fig. 12. Relationship between training sample number and error recognition rate.

target becomes greater than 100, its curve becomes relatively flat and smooth, which means that the model has already been well trained. In addition, by comparing with these curves, it can be found that the models with LSTM (SA-CLATM, CLSATM, SA-LSTM and LSTM) perform significantly better than those with CNN (SA-CNN and CNN) when the number of training samples is reduced. It indicates that the recurrent LSTM structure makes it easier for the model to learn target feature representation from small training samples.

V. Conclusion

In this article, we apply the deep learning approach to the challenging polarimetric HRRP recognition problem which so far still widely relies on feature extraction and classifier design. A novel deep model based on ConvLSTM with self-attention is proposed. In the model, ConvLSTM is used to capture and aggregate target scattering information from both polarimetric and spatial dimensions simultaneously. In addition, self- attention mechanism is introduced before ConvLSTM to make it focus on more discriminative range cells and further enhance its learning capacity. Experiment results on both simulated and measured datasets demonstrate the effectiveness of the proposed model for multi-class target recognition. The results of two expanded experiments also show that our model is robust to noise and small training sample cases and has good potentials for practical applications.

This work is a preliminary study on developing ConvLSTM with self-attention for polarimetric HRRP recognition. To fully unlock their potentials, further efforts will be made on optimizing framework and parameters of the proposed model. Moreover, we consider introducing prior knowledge to the model to further improve the recognition performance.

References

[1] R. Williams, J. Westerkamp, D. Gross, and A. Palomino, "Automatic target recognition of time critical moving targets using 1D high range resolution (HRR) radar," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 15, no. 4, pp. 37–43, Apr. 2000, doi: 10.1109/62.839633.

[2] L. Du, H. He, L. Zhao, and P. Wang, "Noise robust radar HRRP target recognition based on scatterer matching algorithm," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1743–1753, Mar. 2016, doi: 10.1109/JSEN.2015.2501850.

[3] X.-Y. Pan, W. Wang, D.-J. Feng, Y.-C. Liu, and G.-Y. Wang, "Signature extraction from rotating targets based on a fraction of HRRPs," *IEEE Trans. Antennas Propag.*, vol. 63, no. 2, pp. 585–592, Feb. 2015, doi: 10.1109/TAP.2014.2379955.

[4] L. Du, P. Wang, H. Liu, M. Pan, F. Chen, and Z. Bao, "Bayesian spatiotemporal multitask learning for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3182–3196, Jul. 2011, doi: 10.1109/TSP.2011.2141664.

[5] P.-L. Shui, S.-W. Xu, and H.-W. Liu, "Range-spread target detection using consecutive HRRPs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 1, pp. 647–665, Jan. 2011, doi: 10.1109/TAES.2011.5705697.

[6] S. P. Jacobs and J. A. O'Sullivan, "Automatic target recognition using sequences of high resolution radar range-profiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 364–381, Apr. 2000, doi: 10.1109/7.845214.

[7] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Advanced Techniques*. Raleigh, NC, USA: SciTech, 2010, doi: 10.1109/TGRS.2011.2164804.

[8] E. Giusti, M. Martorella, and A. Capria, "Polarimetrically-persistent-scatterer-based automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4588–4599, Nov. 2011.

[9] Q. Yin, W. Hong, F. Zhang, and E. Pottier, "Optimal combination of polarimetric features for vegetation classification in PolSAR image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3919–3931, Oct. 2019, doi: 10.1109/JSTARS.2019.2940973.

[10] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999, doi: 10.1109/36.789621.

[11] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2009.

[12] S.-W. Chen, X.-S. Wang, and M. Sato, "Uniform polarimetric matrix rotation theory and its applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4756–4770, Aug. 2014, doi: 10.1109/TGRS.2013.2284359.

[13] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, no. 18, pp. 1525–1527, Aug. 1990, doi: 10.1049/el:19900979.

[14] W. L. Cameron, N. N. Youssef, and L. K. Leung, "Simulated polarimetric signatures of primitive geometrical shapes," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 3, pp. 793–803, May 1996, doi: 10.1109/36.499784.

[15] R. Touzi and F. Charbonneau, "Characterization of target symmetric scattering using polarimetric SARs," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2507–2516, Nov. 2002, doi: 10.1109/TGRS.2002.805070.

[16] J. Wang, Z. Liu, L. Ran, and R. Xie, "Feature extraction method for DCP HRRP-based radar target recognition via $m-\chi$ decomposition and sparsity-preserving discriminant correlation analysis," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4321–4332, Apr. 2020, doi: 10.1109/JSEN.2019.2962573.

[17] F. Berizzi, M. Martorella, A. Capria, R. Paladini, and D. Calugi, "H/α polarimetric features for man-made target classification," in *Proc. IEEE Radar Conf.*, Rome, Italy, May 2008, pp. 1–6, doi: 10.1109/RADAR.2008.4721003.

[18] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, Mar. 1996, doi: 10.1109/36.485127.

[19] J. R. Huynen, "Phenomenological theory of radar targets," Ph.D. dissertation, Faculty Elect. Eng., Math. Comput. Sci., Tech. Univ. Delft, Delft, The Netherlands, 1970, doi: 10.1016/B978-0-12-709650-.6.50020-1.

[20] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1699–1706, Aug. 2005, doi: 10.1109/TGRS.2005.852084.

[21] A. Freeman and S. L. Durden, "Three-component scattering model to describe polarimetric SAR data," *Proc. SPIE*, vol. 1748, pp. 213–225, Feb. 1993, doi: 10.1117/12.140618.

[22] J. Yang, Y.-N. Peng, and S.-M. Lin, "Similarity between two scattering matrices," *Electron. Lett.*, vol. 37, no. 3, pp. 193–194, Feb. 2001, doi: 10.1049/el:20010104.

[23] L. Guo, "Radar target HRRP polarimetric feature extraction and optimal selection," *Progr. Natural Sci.*, vol. 19, no. 7, pp. 784–792, Jul. 2009, doi: 10.3321/j.issn:1002-008X.2009.07.013.

[24] M. Tao, F. Zhou, Y. Liu, and Z. Zhang, "Tensorial independent component analysis-based feature extraction for polarimetric SAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2481–2495, May 2015, doi: 10.1109/TGRS.2014.2360943.

[25] B. Ren, B. Hou, J. Chanussot, and L. Jiao, "PolSAR feature extraction via tensor embedding framework for land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2337–2351, Apr. 2020, doi: 10.1109/TGRS.2019.2948042.

[26] J. Zheng and H. Zhang, "Urban area classification with polarimetric statistical features of simulated data in PolSAR images," *Electron. Lett.*, vol. 55, no. 13, pp. 761–763, Jun. 2019, doi: 10.1049/el.2019.1153.

[27] O. Harant *et al.*, "Fisher PDF for maximum likelihood texture tracking with high resolution PolSAR data," in *Proc. 8th Eur. Conf. Synth. Aperture Radar*, Aachen, Germany, Jun. 2010, pp. 1–4.

[28] C. Lardeux *et al.*, "Support vector machine for multifrequency SAR polarimetric data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4143–4152, Dec. 2009, doi: 10.1109/TGRS.2009.2023908.

[29] R. Hansch and O. Hellwich, "Evaluation of tree creation methods within random forests for classification of PolSAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Milan, Italy, Jul. 2015, pp. 361–364, doi: 10.1109/IGARSS.2015.7325775.

[30] M. Martorella, E. Giusti, A. Capria, F. Berizzi, and B. Bates, "Automatic target recognition by means of polarimetric ISAR images and neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3786–3794, Nov. 2009, doi: 10.1109/TGRS.2009.2025371.

[31] G. Pajares, C. López-Martínez, F. Sánchez-Lladó, and Ì. Molina, "Improving wishart classification of polarimetric SAR data using the hopfield neural network optimization approach," *Remote Sens.*, vol. 4, no. 11, pp. 3571–3595, Nov. 2012, doi: 10.3390/rs4113571.

[32] R. Shang, J. Wang, L. Jiao, R. Stolkin, B. Hou, and Y. Li, "SAR targets classification based on deep memory convolution neural networks and transfer parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2834–2846, Aug. 2018, doi: 10.1109/JSTARS.2018.2836909.

[33] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[34] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[35] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [review article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018, doi: 10.1109/MCI.2018.2840738.

[36] F. Zhao, Y. Liu, K. Huo, S. Zhang, and Z. Zhang, "Radar HRRP target recognition based on stacked autoencoder and extreme learning machine," *Sensors*, vol. 18, no. 2, p. 173, Jan. 2018, doi: 10.3390/s18010173.

[37] P. Mian, C. Jing, Z. Tao, J. Jie, and L. Zhu, "Radar HRRP recognition based on discriminant deep autoencoders with small training data size," *Electron. Lett.*, vol. 52, no. 20, pp. 1725–1727, Sep. 2016, doi: 10.1049/el.2016.3060.

[38] B. Feng, B. Chen, and H. Liu, "Radar HRRP target recognition with deep networks," *Pattern Recognit.*, vol. 61, pp. 379–393, Jan. 2017, doi: 10.1016/j.patcog.2016.08.012.

[39] O. Karabayir, O. M. Yucedag, M. Z. Kartal, and H. A. Serim, "Convolutional neural networks-based ship target recognition using high resolution range profiles," in *Proc. 18th Int. Radar Symp. (IRS)*, Prague, Czech Republic, Jun. 2017, pp. 1–9, doi: 10.23919/IRS.2017.8008207.

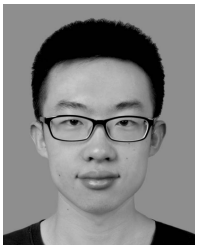
[40] J. Lunden and V. Koivunen, "Deep learning for HRRP-based target recognition in multistatic radar systems," in *Proc. IEEE Radar Conf. (RadarConf)*, Philadelphia, PA, USA, May 2016, pp. 1074–1079, doi: 10.1109/RADAR.2016.7485271.

[41] J. Song, Y. Wang, W. Chen, Y. Li, and J. Wang, "Radar HRRP recognition based on CNN," *J. Eng.*, vol. 2019, no. 21, pp. 7766–7769, Nov. 2019, doi: 10.1049/joe.2019.0725.

[42] Z. Fu, S. Li, X. Li, B. Dan, and X. Wang, "A neural network with convolutional module and residual structure for radar target recognition based on high-resolution range profile," *Sensors*, vol. 20, no. 3, p. 586, Jan. 2020, doi: 10.3390/s20030586.

[43] J. Wan, B. Chen, B. Xu, H. Liu, and L. Jin, "Convolutional neural networks for radar HRRP target recognition and rejection," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, p. 5, Jan. 2019, doi: 10.1186/s13634-019-0603-y.

[44] C. Guo, H. Wang, T. Jian, Y. He, and X. Zhang, "Radar target recognition based on feature pyramid fusion lightweight CNN," *IEEE Access*, vol. 7, pp. 51140–51149, Apr. 2019, doi: 10.1109/ACCESS.2019.2909348.
[45] B. Xu, B. Chen, J. Liu, and C. Du, "Gaussian mixture model-tensor recurrent neural network for HRRP target recognition," in *Proc. Int. Radar Conf. (RADAR)*, Toulon, France, Sep. 2019, pp. 1–6, doi: 10.1109/RADAR41533.2019.171253.
[46] B. Xu, B. Chen, J. Wan, H. Liu, and L. Jin, "Target-aware recurrent attentional network for radar HRRP target recognition," *Signal Process.*, vol. 155, pp. 268–280, Feb. 2019, doi: 10.1016/j.sigpro.2018.09.041.
[47] V. Jithesh, M. J. Sagayaraj, and K. G. Srinivasa, "LSTM recurrent neural networks for high resolution range profile based radar target classification," in *Proc. 3rd Int. Conf. Comput. Intell. Commun. Technol. (CICT)*, Ghaziabad, India, Feb. 2017, pp. 1–6, doi: 10.1109/CIACT.2017.7977298.
[48] M. Pan, J. Jiang, Q. P. Kong, J. G. Shi, Q. H. Sheng, and T. Zhou, "Radar HRRP target recognition based on t-SNE segmentation and discriminant deep belief network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1609–1613, Sep. 2017, doi: 10.1109/LGRS.2017.2726098.
[49] K. Liao, J. Si, F. Zhu, and X. He, "Radar HRRP target recognition based on concatenated deep neural networks," *IEEE Access*, vol. 6, pp. 29211–29218, May 2018, doi: 10.1109/ACCESS.2018.2842687.
[50] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Conf. NIPS*, Montreal, QC, Canada, Dec. 2015, pp. 802–810, doi: 10.1007/978-3-319-21233-3_6.
[51] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Conf. NIPS*, Long Beach, CA, USA, Dec. 2017, pp. 6000–6010.
[52] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 116–130, 1962, doi: 10.1364/josa.52.000116.
[53] W. G. Carrara, R. S. Goodman and R. M. Majewski, *Spotlight Synthetic Aperture Radar-Signal Processing Algorithms*. Norwood, MA, USA: Arthech House, 1995, doi: 10.1016/s1364-6826(97)83336-6.
[54] G. Sinclair, "The transmission and reception of elliptically polarized waves," *Proc. IRE*, vol. 38, no. 2, pp. 148–151, Feb. 1950, doi: 10.1109/JRPROC.1950.230106.
[55] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proc. Conf. NIPS*, Montreal, QC, Canada, Dec. 2014, pp. 2204–2212.
[56] D. Povey, H. Hadian, P. Ghahremani, K. Li, and S. Khudanpur, "A time-restricted self-attention layer for ASR," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 5874–5878, doi: 10.1109/ICASSP.2018.8462497.
[57] R. Cao, L. Fang, T. Lu, and N. He, "Self-attention-based deep feature fusion for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, early access, Feb. 4, 2020, doi: 10.1109/LGRS.2020.2968550.
[58] D. A. Shajahan, V. Nayel, and R. Muthuganapathy, "Roof classification from 3-D LiDAR point clouds using multiview CNN with self-attention," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1465–1469, Aug. 2020, doi: 10.1109/LGRS.2019.2945886.
[59] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, Jun. 1989, doi: 10.1162/neco.1989.1.2.270.
[60] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994, doi: 10.1109/72.279181.
[61] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[62] M. Arsalan and A. Santra, "Character recognition in air-writing based on network of radars for human-machine interface," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8855–8864, Oct. 2019, doi: 10.1109/JSEN.2019.2922395.
[63] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, p. 1330, Dec. 2017, doi: 10.3390/rs9121330.
[64] W.-S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020, doi: 10.1109/TGRS.2019.2961947.
[65] L. J. Moore, B. D. Rigling, and R. P. Penno, "Characterization of phase information of synthetic aperture radar imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 2, pp. 676–688, Apr. 2019, doi: 10.1109/TAES.2018.2864410.
[66] B. Rigling, L. Potter, and R. Moses, "Relative information in phase of radar range profiles," *Proc. SPIE*, vol. 4053, pp. 1–8, Aug. 2000, doi: 10.1117/12.396332.
[67] L. J. Moore, B. D. Rigling, and R. P. Penno, "Characterization of information in phase of radar range profiles," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 2027–2031, doi: 10.1109/ACSSC.2014.7094828.
[68] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, San Diego, CA, USA, May 2015, pp. 1–14.
[69] *The Air Force Moving and Stationary Target Recognition Database*. Accessed: Aug. 20, 2020. [Online]. Available: https://www.sdms.afrl.af.mil/datasets/
[70] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant, "Standard SAR ATR evaluation experiments using the MSTAR public release data set," *Proc. SPIE*, vol. 3370, pp. 566–573, Sep. 1998, doi: 10.1117/12.321859.
[71] Y. Zhou, Z. J. Shu, H. Q. Li, H. Y. Zhou, and X. L. Yu, "A method of radar target length feature extraction based on two-way sliding average," *Mod. Radar*, vol. 38, no. 4, pp. 25–29, Apr. 2016, doi: 10.16592/j.cnki.1004-7859.2016.04.006.
[72] H. W. Liu, L. Du, and Z. Bao, "Progress in radar automatic target recognition based on high range resolution profile," *J. Electr. Inf. Technol.*, vol. 27, no. 8, pp. 1328–1334, Aug. 2005, doi: CNKI:SUN:DZYX.0.2005-08-035.
[73] Q. Xu, X. C. Wang, and Q. Li, "Extraction of target feature using high resolution range profile," *Mod. Radar*, vol. 31, no. 6, pp. 60–63, Jun. 2009, doi: 10.3969/j.issn.1004-7859.2009.06.015.
[74] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008, doi: 10.1007/s10846-008-9235-4.
[75] L. Li and Z. Liu, "Noise-robust HRRP target recognition method via sparse-low-rank representation," *Electron. Lett.*, vol. 53, no. 24, pp. 1602–1604, Nov. 2017, doi: 10.1049/el.2017.2960.
[76] T. Liu, J. Zhang, G. Gao, J. Yang, and A. Marino, "CFAR ship detection in polarimetric synthetic aperture radar images based on whitening filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 58–81, Jan. 2020, doi: 10.1109/TGRS.2019.2931353.
[77] L. M. Novak and M. C. Burl, "Optimal speckle reduction in polarimetric SAR imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 293–305, Mar. 1990, doi: 10.1109/7.53442.



Liang Zhang (Member, IEEE) was born in Nei Mongol, China, in 1993. He received the B.S. degree from the Beijing Institute of Technology, Beijing, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Information and Electronics. His current research interests include high-resolution radar systems, radar polarimetric information processing, and radar automatic target recognition.

Yang Li (Member, IEEE) was born in Liaoning, China. He received the B.S. degree in electrical engineering and the Ph.D. degree in target detection and recognition from the Beijing Institute of Technology, Beijing, China, in 2002 and 2007, respectively.

Since 2007, he has been a Lecturer with the School of Information and Electronics, Beijing Institute of Technology, where he has also been an Associate Professor since October 2010. His research interests include high-resolution radar systems and radar signal processing.

Yanhua Wang (Member, IEEE) was born in Henan, China, in 1984. He received the B.S. degree in electrical engineering and the Ph.D. degree in target detection and recognition from the Beijing Institute of Technology, Beijing, China, in 2006 and 2011, respectively.

From March 2012 to October 2014, he was a Postdoctoral Research Associate with the Department of Electrical Engineering and the Department of Biomedical Engineering, University at Buffalo, State University of New York, USA. Since 2014, he has been a Lecturer with the School of Information and Electronics, Beijing Institute of Technology, where he has also been an Associate Professor since June 2019. His research interests include high-resolution radar systems, signal processing, and radar automatic target recognition.

Junfu Wang was born in Pingliang, China, in 1989. He received the M.S. degree in signal and information processing from the Beijing Institute of Technology, Beijing, China, in 2014. His current research interests include high-resolution radar systems, radar signal processing, radar target detection and recognition, and multi-sensor information fusion.

Teng Long (Fellow, IEEE) was born in Fujian, China, in 1968. He received the M.S. and. Ph.D. degrees in electrical engineering from the Beijing Institute of Technology in 1991 and 1995, respectively.

He was a Visiting Scholar with Stanford University, CA, in 1999, and the University College London, in 2002. He has been a Full Professor with the Department of Electrical Engineering, Beijing Institute of Technology, since 2000. He has authored or coauthored more than 300 articles. He has received many awards for his contributions to research and invention in China. His research interests include synthetic aperture radar systems and real-time digital signal processing, with applications to radar and communication systems. He is a Fellow of the Institute of Electronic and Technology and the Chinese Institute of Electronics.