LETTER

Millimeter-Wave InSAR Target Recognition with Deep Convolutional Neural Network

Yilu MA[†a)], ***Nonmember and*** **Yuehua LI**[†b)], ***Member***

SUMMARY Target recognition in Millimeter-wave Interferometric Synthetic Aperture Radiometer (MMW InSAR) imaging is always a crucial task. However, the recognition performance of conventional algorithms degrades when facing unpredictable noise interference in practical scenarios and information-loss caused by inverse imaging processing of InSAR. These difficulties make it very necessary to develop general-purpose denoising techniques and robust feature extractors for InSAR target recognition. In this paper, we propose a denoising convolutional neural network (D-CNN) and demonstrate its advantage on MMW InSAR automatic target recognition problem. Instead of directly feeding the MMW InSAR image to the CNN, the proposed algorithm utilizes the visibility function samples as the input of the fully connected denoising layer and recasts the target recognition as a data-driven supervised learning task, which learns the robust feature representations from the space-frequency domain. Comparing with traditional methods which act on the MMW InSAR output images, the D-CNN will not be affected by information-loss accused by inverse imaging process. Furthermore, experimental results on the simulated MMW InSAR images dataset illustrate that the D-CNN has superior immunity to noise, and achieves an outstanding performance on the recognition task.

key words: *target recognition, MMW InSAR, feature extractor, denoising convolutional neural network*

1. Introduction

Target recognition appears to be key components in MMW InSAR imaging based applications, such as indoor security, earth remote sensing [1], and aircraft navigation and so on, because of the advantage of good concealment performance, all-weather condition, high-resolution, rapid and accurate data collection. Furthermore, InSAR measures the correlation between pairs of various nondirective antennas to realize high-resolution instead of using a large aperture antenna directly, that makes the system convenient to physical applications [2].

A number of sensor modelings have been developed, and some target recognition algorithms deriving from optical image processing have been applied to InSAR image processing. In [3], Yun combined kernel Principle Component Analysis (KPCA) to analyze the phase coherence and backscattering coefficient, producing highly reliable information of interpreting land-cover. This study also clearly proved the effectiveness of InSAR signatures for comprehensive classification. M.E. Engdahl proposed a method which utilized multi-temporal InSAR data to perform land-cover classification, providing volume estimates for the forested areas [4]. Chen proposed an A-CNN framework which only utilized CNN layers to reduce overfitting problem, and produced a high accuracy to SAR image target recognition [5], [6].

However, due to the imaging mechanism and facing numerous objects in complex scenarios, the MMW InSAR image still suffers from artifacts, Gibbs ringing effect of the edges, information-loss caused by the inverse imaging procedure, noise of environment and system, which easily invalidate conventional target recognition algorithms [7]. To solve these problems, as well as to learn high-level robust feature representations, a novel denoising convolutional neural network, called D-CNN, is proposed for MMW InSAR target recognition in this paper. In the proposed framework, data augmentation operations are utilized to overcome the risk of limited training data in advance. Then, different from the conventional recognition algorithms, the D-CNN directly utilizes visibility function samples as inputs. A fully connected denoising structure deriving from the pre-trained DAE is utilized to learn robust projection between visibility function samples and intermediate images. After that, the intermediate images are fed to the following CNNs, which are designed to extract stable features from input and generating high-level representations. Finally the classification result is provided by softmax classifier. As the directly utilizing of visibility function samples, the proposed framework could effectively avoid being affected by information-loss and Gibbs ringing accused by inverse imaging process, thus preserving detailed information of the target. Experimental results on the MMW InSAR image dataset verify the effectiveness of the proposed method.

2. The Related Work

2.1 MMW InSAR Model

The simplified geometric relationship of interferometry is illustrated in Fig. 1. The MMW InSAR system is usually composed of the binary interferometers, and it measures the correlation value between pairs of spatially separated antennas, which is named visibility function. The visibility function is defined as:

Manuscript received July 27, 2018.
Manuscript publicized November 26, 2018.

[†]The authors are with School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China.
a) E-mail: mayilu1991@njust.edu.cn
b) E-mail: nlglyh2013@sina.cn (Corresponding author)

DOI: 10.1587/transinf.2018EDL8158

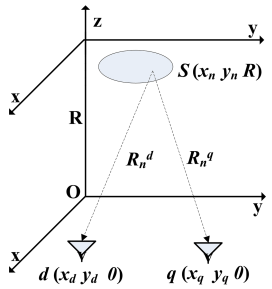

Fig. 1 Interference measurement schematic

$$V(u,v) = \iint\limits_{\xi^2+\eta^2 \leqslant 1} T(\xi,\eta) r\left(-\frac{u\xi+v\eta}{f_0}\right) e^{-j2\pi(u\xi+v\eta)} d\xi d\eta \tag{1}$$

where $(\xi,\eta) = (\sin\theta\cos\phi, \sin\theta\sin\phi)$ is the polar coordinate with respect to the spatial axes (x,y), f_0 is the center frequency, and $r(-\frac{u\xi+v\eta}{f_0})$ is the fringe washing function. In the ideal digital signal processing case, the radiation source S is dispersed into N small parts (ΔS_n), and the fringe washing function could be ignored, thus the visibility function $V_{d,q}$ can be represented as:

$$V_{d,q} = \sum_{n=1}^{N} T(n) F_d(n) F_q^*(n) e^{-jK\left(R_n^d - R_n^q\right)} \Delta S_n \tag{2}$$

where $T(n)$ represents target brightness temperature image, $F_d(n)$ and $F_q(n)$ are the normalized antenna pattern of antenna d and q. R_n^d and R_n^q are the distances between the radiation source S and antennas. K is circular wavenumber, defined as $\frac{2\pi}{\lambda}$, and λ is the center wavelengths of the electromagnetic radiation received by InSAR imaging system. As the phase compensation is need in near-field, the distance will be processed accurately to establish a new G matrix. Thus, Eq. (2) can be expressed as Eq. (3) in the matrix form.

$$V_{M\times 1} = G_{M\times N} T_{N\times 1} \tag{3}$$

$$G(m,n) = F_{md}(n) F_{mq}^*(n) e^{-jK\left(R_n^{md} - R_n^{mq}\right)} \tag{4}$$

$$R_n^{md} = \sqrt{(x_n - x_d)^2 + (y_n - y_d)^2 + R^2} \tag{5}$$

$$R_n^{mq} = \sqrt{\left(x_n - x_q\right)^2 + \left(y_n - y_q\right)^2 + R^2} \tag{6}$$

where md and mq represent the antennas' position when generating the mth sample of visibility function. The obtained visibility function V is a complex-valued matrix.

2.2 Neural Networks

The definitions of DAE and CNN are briefly introduced in this section.

1. DAE: A typical DAE is composed of an encoder and a decoder. Let x be the given input vector, the operation of the DAE can be defined as follows.

Corruption: it adds binary mask noise n to x.

$$\hat{x} = x + n \tag{7}$$

Encoder: it maps the input vector to the hidden representation h.

$$h = f(W_1\hat{x} + b_1) \tag{8}$$

Decoder: it reconstructs y from hidden representation.

$$y = g(W_2 h + b_2) \tag{9}$$

where W_1 and W_2 denote the weight matrix, b_1 and b_2 are bias vectors, f and g are nonlinear activation function.

2. CNN: Let the $O_i^{l-1}(x,y)(i = 1,\ldots,I)$ represents the unit at the position (x,y) of $i{-}th$ input feature map in the previous layer, and $O_j^l(x,y)(j = 1,\ldots,J)$ represents the unit at the position (x,y) of $j{-}th$ output feature map in this layer. Each step of CNN can be represented as follows.

Convolution: it computes the convolution of input with a bank of convolution kernels (filters) k_{ji}^l.

$$\begin{aligned} O_j^l(x,y) &= f(V_j^l(x,y)) \\ &= f\left(\sum_{i=1}^{I}\sum_{u,v=0}^{F-1} k_{ji}^l(u,v)\cdot O_i^{l-1}(x-u,y-v) + b_j^l\right) \end{aligned} \tag{10}$$

where f is the nonlinear activation function, F denotes the size of filter, and b_j denotes the bias, I is the number of input feature maps, and l represents the present layer.

Max-Pooling: It outputs the maximum value on a group of units located within a local patch.

$$O_j^{l+1}(x,y) = \max_{u,v=0,\ldots,P-1} O_j^l(x\cdot s + u, y\cdot s + v) \tag{11}$$

where P is the pooling size and s denotes the stride of pooling windows.

3. Dataset and Learning for D-CNN

In this section, we first give a brief introduction to the simulated MMW InSAR image dataset. Then we present the details of specific configurations and leaning of D-CNN.

3.1 Dataset Simulation

In this paper, all the simulated images are generated from 43 models in 3 classes, which are composed of 16 types of planes, 14 types of tanks and 13 types of ships. Five specified view angles are manually selected to cover the surface and capture major features of the models, generating 215 different images. For the purpose of augmenting the dataset, each posture is anticlockwise rotated from 0° to 180°, and the interval of the angle we set is 15°. Therefore, we obtain 2580 images in total. The near field based imaging algorithm is applied to each image. Some thumbnails of the simulated MMW InSAR images are presented in Fig. 2.

3.2 Learning for D-CNN

Inspired by the perceptual learning archetype, a data-driven

target recognition algorithm is designed to learn robust feature representations from visibility function samples, which is called D-CNN. The D-CNN consists of two parts: the fully connected denoising layers deriving from pre-trained DAE and the CNNs. The structure of D-CNN is depicted in Fig. 3.

The size of the complex-valued input is 50×50. Since the neural networks we used operate on real-valued inputs and parameters, complex data are separated in to real components in the input vector. Thus, the 50×50 complex-valued matrix is reshaped to a 5000×1 real-valued vector.

For the first part, a DAE is firstly trained to depress the noise interference. The hidden structure ($5000 \times 2500 \times 1000 \times 2500$) of the DAE is utilized to approximate the projection between visibility function samples and images, generating 2500×1 real-valued vectors. After that, these real-valued vectors are reshaped to the 50×50 intermediate images.

For the second part, the reshaped intermediate image is fed to the CNNs. The first convolutional layer ($C1$) convolves 8 filters of 13×13. The second convolutional layer ($C2$) convolves 16 filters of 3×3, followed by a max-pooling layer ($P1$). The third convolutional layer ($C3$) convolves 32 filters of 15×15 again followed by a max-pooling layer ($P2$). The stride of the filters is 1 and the pooling size is 2×2. Then the reshaped output is fed to the softmax classifier layer which contains 128 neurons, followed by an output layer contains 3 neurons.

After the separately pre-training, we integrate the denoising layers and the CNNs to build a new network (D-CNN) between visibility function samples and corresponding labels. Labeled visibility function samples are again utilized to further fine-tune the parameters of the whole network.

Let L be the recognition error, the trainable parameters is fine-tuned by the gradient decent of subset, and the derivatives can be deduced according to the chain rule.

$$\frac{\partial L}{\partial w_{ij}^l} = \frac{\partial L}{\partial V_i^l}\frac{\partial V_i^l}{\partial w_{ij}^l} = \delta_i^l \frac{\partial V_i^l}{\partial w_{ij}^l} = \delta_i^l O_i^{l-1} \tag{12}$$

The δ_i^l is defined as following in fully connected layers.

$$\delta_i^l = \frac{\partial L}{\partial V_i^l} = \frac{\partial L}{\partial O_i^l}\frac{\partial O_i^l}{\partial V_i^l} \tag{13}$$

However, if the $l{-}th$ layer is convolutional layer, the error term δ_i^l is related to the error term δ_j^{l+1} of the pooling layer. The δ_i^l can be represented as:

$$\delta_i^l(x,y) = \frac{\partial O_i^l}{\partial V_i^l} \cdot \delta_j^{l+1}(x,y) \tag{14}$$

If the $l{-}th$ layer is a pooling layer, the δ_i^l can be represented as:

$$\delta_i^l(x,y) = \sum_{j=1}^{J}\sum_{u,v=0}^{F-1} k_{ji}^{l+1}(u,v) \cdot \delta_j^{l+1}(x+u, y+v) \tag{15}$$

After computing the error term in each layer, the updated rule of weight and bias is represented as:

$$W^l{}_{(k)} = W^l{}_{(k-1)} + \eta\delta_i^l \cdot O_i^{l-1} \tag{16}$$

$$b^l{}_{(k)} = b^l{}_{(k-1)} + \eta\delta_i^l \tag{17}$$

where k is the number of iterations, η is the learning rate. We initially fix the learning rate to 0.01 and reduce it during the training, with the reduction of 10^{-5} after each epoch.

4. Experiments and Result Analysis

In this section, we demonstrate the performance of the D-CNN and give the comparison with three traditional target recognition algorithms: Stacked Auto-encoder (SAE), Local Binary Patterns (LBP) and CNN.

In the first experiment, we evaluate the immunity to noise for all algorithms. The Gaussian noise whose intensity varies from 0 to 0.30 is added to the MMW InSAR images as well as the visibility function samples. The recognition accuracy across the noise level are shown in Fig. 4,

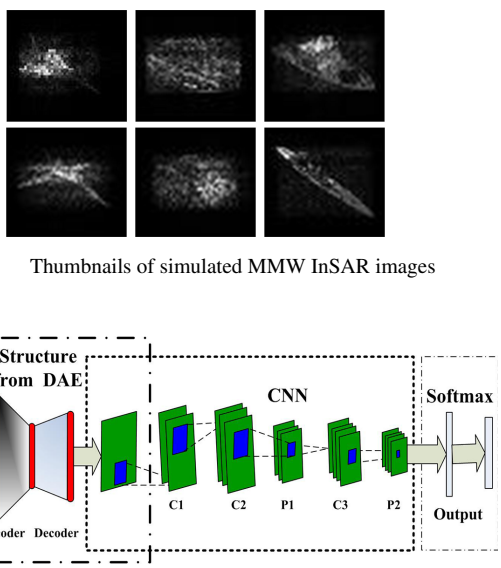

Fig. 3 The illustration of networks for manuscript

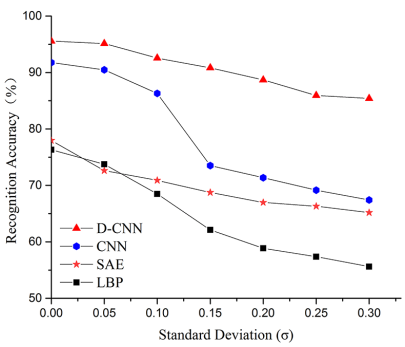

Fig. 4 Recognition accuracy at variable noise levels

Table 1 Recognition accuracy with different ratios

(a) Noise free

Ratio	60%	70%	80%	90%
D-CNN	88.12	92.07	95.63	96.87
CNN	80.26	85.93	90.96	91.74
SAE	61.81	69.74	74.85	77.53
LBP	59.28	68.45	73.47	76.10

(b) Noise intensity is 0.3

Ratio	60%	70%	80%	90%
D-CNN	73.52	79.16	85.43	90.48
CNN	60.42	66.36	69.05	79.76
SAE	56.25	64.69	66.32	73.66
LBP	49.41	52.67	57.44	67.25

Table 2 Confusion matrix (σ=0.05)

Method	D-CNN			CNN			SAE			LBP		
Type	Plane	Tank	Ship	Plane	Tank	Ship	Plane	Tank	Ship	Plane	Tank	Ship
Plane	93.47	3.05	3.48	88.29	5.39	5.32	72.34	12.06	15.60	73.76	12.06	14.18
Tank	1.57	95.10	3.33	3.82	90.98	5.2	9.81	71.56	18.63	11.76	73.53	14.71
Ship	0.75	1.40	97.85	2.47	4.09	93.44	10.68	14.13	75.19	9.78	15.95	74.27

which demonstrates that D-CNN has a superior immunity to noise over other methods. First of all, it can be observed that the D-CNN consistently achieves higher recognition accuracy than other three methods for all noise intensities, and reaches 96.87% in the case of noise free. Furthermore, it should be noted that the recognition accuracy of the other three algorithms decrease rapidly with the increasing noise intensity, whilst the accuracy of the D-CNN is always over 90%, that obviously shows its strong robustness to noise. The gap between D-CNN and other two deep learning models is about 20% when the noise intensity reaches 0.3.

In the second experiment, we evaluate the performance of D-CNN with different usage ratios (from 60% to 90%) of training samples. The experiments are performed in two different conditions (noise free and the noise intensity is 0.30), and the results are shown is Table 1. It's clearly observed from Table 1 that the proposed algorithm outperforms than other traditional methods. First, we can find that with the same usage ratio of training samples, the D-CNN performs much better than traditional ones, indicating that the D-CNN can learn much more useful features from visibility function samples whilst the traditional methods suffers from information-loss and artifact caused by InSAR inverse imaging procedure. On the other hand, it can be observed from Table 1 that when utilizing 60% of training samples, the recognition accuracy of D-CNN is 8% and 13% higher than that of other methods in these two conditions. Considering the fact that it's hard to acquire abundant samples for training and the number of testing samples is usually huge, the proposed D-CNN is better for MMW InSAR target recognition task.

The confusion matrix of D-CNN, CNN, SAE, and LBP is listed in Table 2. 80% of MMW InSAR samples are utilized for training, and noise intensity is fixed to 0.05. We can find that the D-CNN outperforms than other three methods for all types target, which indicates that it is more robust than other methods.

5. Conclusions

In this paper, a novel denoising algorithm named D-CNN is designed to recognize the MMW InSAR target. In contrast to conventional algorithms which act on the InSAR output images, the D-CNN directly utilizes the visibility function samples as input, and recasts the target recognition as a data-driven supervised learning task. Therefore, the D-CNN automatically learns the robust feature representations from the visibility function samples, and will not be affected by information-loss or Gibbs ringing accused by inverse imaging process. Furthermore, the denoising structure deriving from DAE can effectively suppress the noise interference from the measurement process. Experimental results demonstrate that D-CNN is able to provide higher recognition accuracy and has superior immunity to noise than other recognition algorithms, especially when the training samples are limited.

References

[1] C. Bentes, D. Velotto, and B. Tings, "Ship Classification in TerraSAR-X Images With Convolutional Neural Networks," IEEE J. Ocean. Eng., vol.43, no.1, pp.258–266, 2018.
[2] J. Chen, Y. Li, J. Wang, Y. Li, and Y. Zhang, "An accurate imaging algorithm for millimeter wave synthetic aperture imaging radiometer in near-field," Progress In Electromagnetics Research, vol.141, pp.517–535, 2013.
[3] H.W. Yun, J.R. Kim, C.Y. Soo, et al., "The application of InSAR time series for landcover classification," IEEE Synthetic Aperture Radar, pp.308–311, 2013.
[4] M.E. Engdahl, J. Pulliainen, and M. Hallikainen, "Combined land-cover classification and stem volume estimation using multitemporal ERS tandem INSAR data," 2003 IEEE International Geoscience and Remote Sensing Symposium, pp.1936–1938, 2003.
[5] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," Proc. 2010 IEEE International Symposium on Circuits and Systems, pp.253–256, 2010.
[6] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target Classification Using the Deep Convolutional Networks for SAR Images," IEEE Trans. Geosci. Remote Sens., vol.54, no.8, pp.4806–4817, 2016.
[7] Y. Zhang, Y. Li, and S. Safavi-naeini, "A Spectrum-Based Saliency Detection Algorithm for Millimeter-Wave InSAR Imaging with Sparse Sensing," IEICE Trans. Inf. & Syst., vol.E100.D, no.2, pp.388–391, 2017.
[8] B. Zhu, J.Z. Liu, S.F. Cauley, B.R. Rosen, and M.S. Rosen, "Image reconstruction by domain-transform manifold learning," Nature., vol.555, no.7697, pp.487–492, 2018.