A New Method of Robust Ground Moving Target Detection Under Different Backgrounds of Airborne SAR Based on Spatial Deformable Module

He Yan , Hui Liu , Jialin Hao, Wenshuo Xu, Jingdong Zhang , *Member, IEEE*, Di Wu , Xudong Wang, Gong Zhang , *Member, IEEE*, and Daiyin Zhu , *Member, IEEE*

Abstract—Combined with the synthetic aperture radar (SAR) imaging algorithm, diverse simulation sample sets of ground moving targets are constructed to tackle the problem of insufficient measured data in the SAR ground moving target indication algorithm based on deep learning. In view of this, a overall scheme of realizing robust detection of ground moving targets under varying backgrounds is conducted on the integration of the adaptive spatial location extraction network based on deformable module and the multichannel clutter suppression technology. In particular, a spatial deformable module is incorporated into the network to enhance its modeling capacity of the input targets with different shapes. Furthermore, the multichannel clutter suppression technology of airborne SAR is adopted to significantly mitigate the interference of complex background clutter. The effectiveness of the proposed method is verified on the simulation sample sets, and comparison with other detection methods is given simultaneously.

Index Terms—Adaptive spatial location extraction network based on deformable module (ASLE-DM), clutter suppression, ground moving target detection, synthetic aperture radar (SAR) imaging algorithm.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) is an active microwave imaging sensor, which realizes high-resolution imaging and recognition of region of interest (ROI) and plays an irreplaceable role in military surveillance and civilian remote sensing [1], [2], [3], [4], [5]. As one of the key applications of SAR, ground moving target indication (GMTI) has aroused the interests of many researchers, aiming to quickly find and locate the ROI from noise and complex backgrounds [6], [7].

Traditional SAR-GMTI algorithms are mainly based on echo such as constant false alarm rate (CFAR) algorithm [8]. Since the CFAR detector is conducted on the statistical characteristics of clutter, the detection performance is easily disturbed by background clutter, and the false alarm rate increases significantly, especially under the complex urban background. Guo et al. [9] proposed a modified SAR moving target detection method based on

robust principal component analysis, which separated the sparse matrix of moving targets from the low-rank matrix of static background. In [10], based on the clutter suppression and multitarget tracking algorithm, the target trajectory was reconstructed by using the parameter estimation method on the foundation of Doppler feature and circular SAR geometry. In [11], moving targets with different parameters were classified according to the relative positions of moving target spectrum and clutter spectrum. On this basis, a two-stage GMTI method was proposed to detect multiple moving targets and especially those submerged by clutter. However, the above mentioned traditional SAR-GMTI algorithms rely heavily on the output signal-to-clutter-and-noise ratio [12] and signal processing technology. The clutter disturbance under complex backgrounds can easily affect the detection results, and the operation complexity is rather high.

In recent years, the intelligent SAR-GMTI algorithm based on deep learning has been continuously improved and developed, which realizes the detection of the ground moving targets by learning the morphological characteristics of 2-D SAR images. In [13], a new anchor-free moving target detection algorithm was proposed to improve the performance of the model by enhancing the strong scattering properties of the targets, and the effectiveness of the algorithm was verified on the published relevant datasets. Wang et al. [15] took the SAR-GMTI task as a blind inverse problem, and it was tackled by using the deep complex-valued convolutional network, which significantly improved the detection and refocusing accuracy. Aiming at the problems of insufficient feature extraction and high false alarm rate in the current SAR-GMTI algorithm based on the convolutional neural network (CNN), Mu et al. [14] proposed a network model based on the improved GoogLeNet architecture to enhance the dependencies between channels, obtain more context information, and improve the detection performance under a complex background. Inspired by transfer learning, Zhang et al. [15] proposed a subdomain adaptive residual network for moving target detection in the multichannel SAR system and was verified through three-channel SAR data.

However, these algorithms mainly focus on optimizing the detection performance of moving targets under the same background, but most of them lack flexibility and universality for the targets under varying backgrounds. In practical applications, the background where the moving target is located often changes, so it is of great research significance to realize the robust detection

of the moving targets under new backgrounds. The performance of the intelligent detection algorithm depends largely on the quantity and quality of sample datasets, but it is difficult to obtain the measured SAR data due to the operating cost and complexity of the SAR system. The performance and generalization capability of the neural network will be extremely constrained when only a small number of SAR image samples are used for training.

In this article, the SAR imaging algorithm is adopted to construct the SAR sample datasets for network training and testing. On this basis, we propose a new method that enables the network to realize the robust detection of ground moving targets under different backgrounds. The contributions of this article are given as follows.

1) To tackle the problem of insufficient SAR image datasets when applying intelligent detection algorithms to SAR-GMTI, a new method of experimental data construction is proposed in this article to simulate multiple moving targets under varying backgrounds combined with the SAR imaging algorithm, taking the backprojection (BP) algorithm [16] as an example. Particularly, the characteristics of SAR moving targets and scenes are fully considered in the experiment to ensure proper and reliable simulation results. By traversing the location parameters of the moving targets (azimuth–range coordinates) within a reasonable range, it is easy to obtain a large number of SAR image datasets required for network training and testing accordingly.

2) In this article, the adaptive spatial location extraction network based on deformable module (ASLE-DM) network structure based on the spatial deformable module is proposed to adaptively model the input targets with different shapes, which realizes robust detection of polymorphous moving targets in the SAR scene. The network can adjust the positions of spatial sampling points and the size of receptive field according to the geometry transformation, thus optimizing the feature extraction mode of moving targets and improving the detection performance of the model. Especially for the SAR system where the imaging scene rotates due to the change of observation angle (such as circular spotlight SAR), ground moving targets can still be detected correctly with no need for image registration [17].

3) The proposed method possesses flexibility and universality for the moving targets under varying backgrounds in practical applications. For the intelligent detection of moving targets under the complex background, the interference of artificial buildings, rocks, and vegetation in the scene will bring about many false alarms in the detection results. In this regard, the multichannel clutter suppression algorithm of airborne SAR is applied to significantly suppress the background clutter, while the moving target information is fully reserved to further mitigate the false alarms in the detection results. Even when there exists channel error in practical applications, the network can still realize the robust detection of moving targets in different SAR scenes.

II. CONSTRUCTION OF MULTIPLE MOVING TARGET DATASETS

At present, most of the SAR-GMTI algorithms based on deep learning focus on the optimization of moving target detection performance under a single background. However, in practical applications, the scene where the target is located often varies, so it is of great significance to explore how to robustly detect the moving targets under a new background.

For the target detection algorithm based on deep learning, whether the network can realize the flexible and robust detection of ROI depends largely on the quality and quantity of the training sample datasets. However, it is extremely difficult to obtain the measured SAR data for network training and testing, and it is even impossible to obtain the datasets of moving targets under different backgrounds required in the experiment. Therefore, the construction of multiple high-fidelity simulation sample sets is the key to improving the performance of the algorithm.

A. Signal Model

Assuming that the airborne SAR works in the spotlight imaging mode, the linear frequency modulation continuous wave (LFMCW) transmitted by SAR can be expressed as

$$S_0(\tau) = A\text{rect}\left(\frac{\tau}{T_r}\right)\exp(j2\pi f_0\tau)\exp(j\pi K_r\tau^2) \quad (1)$$

where A represents the signal amplitude, τ represents the range time, T_r is the bandwidth, K_r is the range chirp rate, and f_0 is the carrier frequency. The received signal is the delay of the transmitted signal $S_0(\tau)$ and can be expressed as

$$r_0(\tau, t) = A\text{rect}\left(\frac{\tau - \tau_i}{T_r}\right)\exp(j2\pi f_0(\tau - \tau_i))$$
$$\exp(j\pi K_r(\tau - \tau_i)^2) \quad (2)$$

where τ_i is the time delay given by $\tau_i = \frac{2R(t)}{c}$, t is the time index in slow time, and $R(t)$ is the instantaneous range between the radar platform and the moving target. The echo signal after demodulation can be expressed as

$$S(\tau, t) = A_0\text{rect}\left(\tau - \frac{2R(t)}{c}\right)\exp\left(-j4\pi f_0\frac{R(t)}{c}\right)$$
$$\exp\left(j\pi K_r(\tau - \frac{2R(t)}{c})^2\right) \quad (3)$$

where A_0 is an arbitrary complex constant. Then, the received echo signal after range compression is

$$S_{rc}(\tau, t) = A'\rho_r\left(\tau - \frac{2R(t)}{c}\right)\exp\left(-j4\pi f_0\frac{R(t)}{c}\right) \quad (4)$$

where A' is the gain including scattering coefficient, and ρ_r is the envelope of the signal echo after pulse compression.

B. BP Algorithm

With the continuous development of SAR technology, high-resolution imaging algorithms [18], [19] have emerged one after another. Taking the BP imaging algorithm as an example,
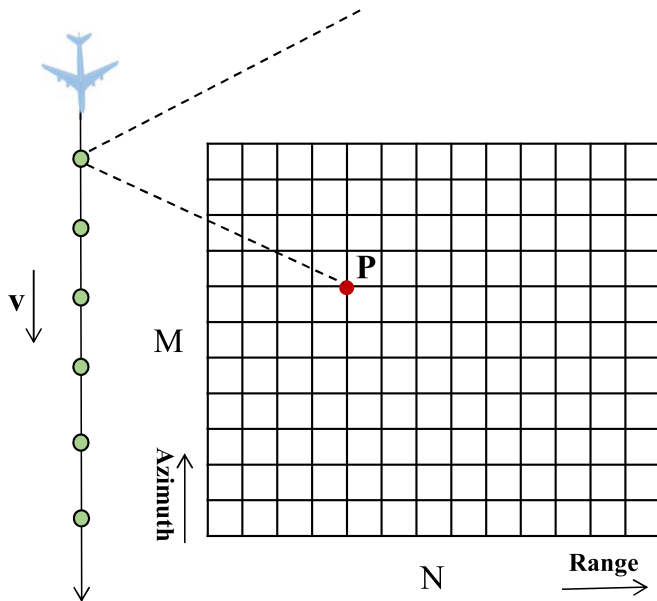
Fig. 1.	Diagram of the BP imaging algorithm.

Fig. 2.	Key steps for the construction of simulation sample sets.

the construction of diverse SAR image sample sets for neural network training and testing is completed in this article. The BP algorithm is a typical time-domain imaging algorithm. Compared with the imaging algorithms conducted in the frequency domain, such as polar format algorithm, range–Doppler, and chirp scaling, it can accurately eliminate the coupling between range and azimuth directions without any assumption of slant range approximation, and it is applicable to the radar system in any working mode [20].

The imaging diagram of the BP algorithm is shown in Fig. 1. The imaging scene is divided into $M \times N$ grids along azimuth and range directions. M is the number of points in the azimuth direction, and N is the number of points in the range direction. The high-resolution imaging result of the BP algorithm is realized through the relative position between the grid and the phase center of the antenna. The echo data are projected back to the imaging grid, and the amplitude of each pixel at different azimuth moments is coherently accumulated to obtain the objective function [21].

For the pixels in the imaging grid, the instantaneous slant range between each pixel and the platform at the current azimuth moment is calculated to acquire the two-way propagation delay and obtain the phase compensation factor. By interpolating the radar echo signal, the amplitude value is correspondingly assigned to each pixel and then coherently accumulated along the azimuth time to obtain the imaging result of this pixel. The pulse at each azimuth moment is coherently accumulated to obtain the imaging result of each pixel point

$$I(x,y) = \sum_{t=t_1}^{t=t_2} S_{rc}(\tau_{xy}, t) \exp(j2\pi f_0 \tau_{xy}) \qquad (5)$$

where (x,y) represents the coordinate of the pixel point in the imaging grid, τ_{xy} is the two-way propagation delay from the
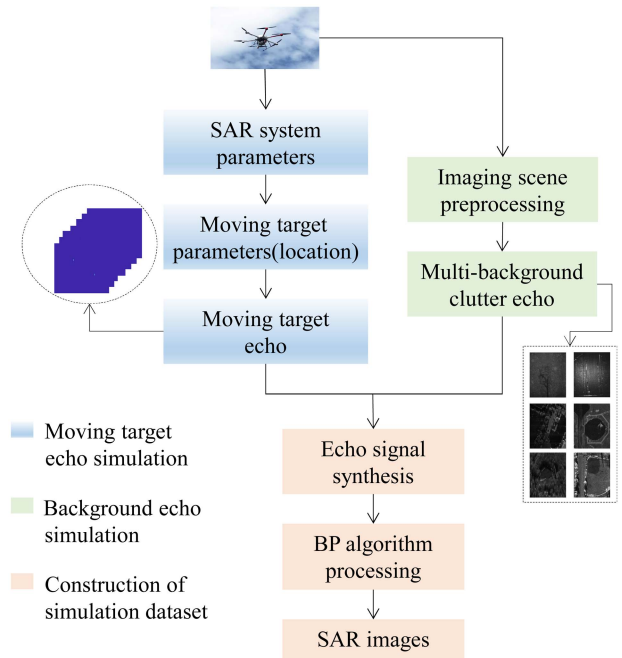
pixel point to the radar platform, and t_1 and t_2 represent the starting and ending time of the current azimuth pulse, respectively.

C. Construction of Multiple Moving Target Datasets Under Different Backgrounds

As the SAR scene and position state where the moving target is located always vary in practical applications, the simulation dataset should possibly comprise the moving targets under different backgrounds and at different positions. In the process of simulation, the matching between the moving target and the scene (such as scale ratio and contrast) also needs to be considered reasonably.

After setting the parameters such as position coordinate of moving target according to the characteristic of scene, the simulation of moving target is conducted, and the obtained echo data are stored in the matrix for subsequent processing. For the echo simulation of SAR scene, as SAR images are gray images, it is necessary to preprocess the pixel values first, that is, normalize the value of each pixel point to 0–255 pixel by pixel. Then, the obtained echo data of each SAR scene and moving targets are adaptively added, and the noise component is added to make the simulation results closer to practical applications. The SAR images of moving targets under different backgrounds can be obtained by processing the integrated echo data by using the BP algorithm in Section II-C. In addition, the noise component is superimposed to the echo of moving target alone so as to obtain the SAR datasets for network training. The main steps for the construction of simulation datasets are shown in Fig. 2.

As the position of a moving target can be set arbitrarily within a reasonable range, it is easy to obtain a large number of SAR image sample sets by traversing the position parameters
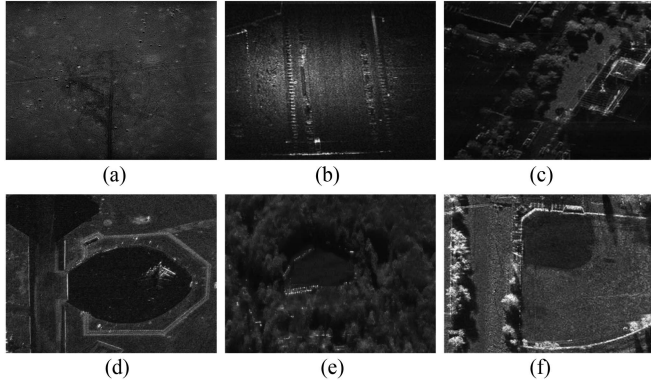
Fig. 3. Simulation results of SAR scenes. (a) Scene 1 (desert). (b) Scene 2 (urban road). (c) Scene 3 (artificial buildings). (d) Scene 4 (apron). (e) Scene 5 (forest). (f) Scene 6 (baseball diamond).
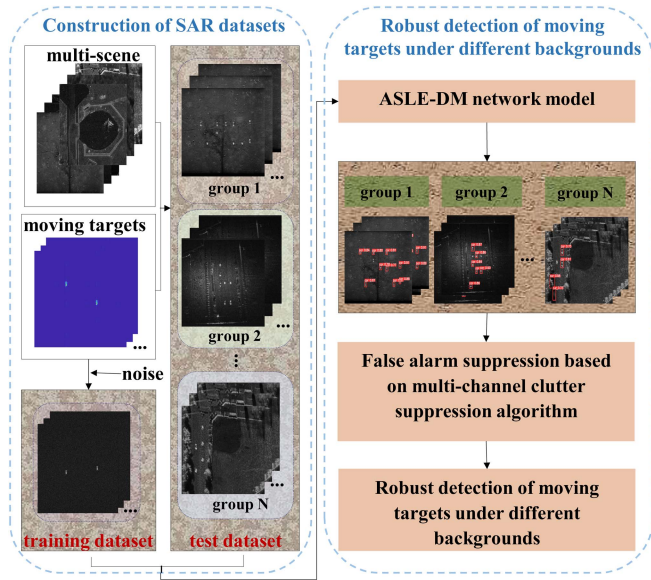

Fig. 4. Overall framework of realizing robust detection of ground moving targets under different backgrounds based on the ASLE-DM network.

(azimuth–range coordinates) of the targets. In this article, typical SAR scenes (desert, complex urban backgrounds, forest, etc.) are selected for simulation, and the imaging results are shown in Fig. 3. Under the desert background in Fig. 3(a), some rocks show strong scattering properties, and the features of gullies and target trajectories are obvious. In addition, the overall characteristic of the desert is similar to the noise background. However, there relatively exists plenty of interference from ground objects such as vegetation, roads, and artificial buildings in other backgrounds shown in Fig. 3(b)–(f), which will make it difficult for the network to accurately detect the moving targets in the scene.

Combined with the constructed SAR image sample sets, it is proved that the proposed method can realize robust detection of moving targets under different backgrounds. The overall framework of this article is shown in Fig. 4. In particular, the simulation dataset of moving targets under the noise background

is used for network training, and the trained model is verified by using the simulation test sets of moving targets under different backgrounds.

III. ADAPTIVE FEATURE EXTRACTION OF POLYMORPHIC GROUND MOVING TARGETS USING ASLE-DM

The CNN [22], [23] is limited to adaptively model geometric transformations such as the variation of target shape and scale due to its fixed model structure, which greatly limits the detection performance of the network. The ASLE-DM network based on the spatial deformable module is proposed in this article, allowing the network to adjust the spatial sampling position and ROI feature mapping mode according to the geometric characteristics of the input feature map [24], [25].

A. Spatial Deformable Module

The receptive field plays an important role in feature extraction, and it should be neither too big nor too small. In the traditional CNN network structure, all the activation units in the same CNN layer have the same receptive field when processing different input feature maps, which is obviously undesirable for the deep CNN layer to encode high-level semantic information. The deformable convolution and deformable ROI pooling module can adjust the position of spatial sampling points according to the shapes of input feature maps, and the receptive field is transformed accordingly.

1) Deformable Convolution: Based on standard convolution, a 2-D offset is additionally designed for spatial sampling position points, which provides the network with the modeling capacity of the freely transformed grid points. 2-D deformable convolution is conducted in two steps: 1) position sampling is conducted on the input feature map I_{fm} by using regular grid G; and 2) all the sampling values in step 1 are summated by weight, and the weight is expressed by w. Each position g_0 on the output feature map O_{fm} can be expressed as

$$O_{fm}(g_0) = \sum_{g_n \in G} w(g_n) \cdot I_{fm}(g_0 + g_n) \qquad (6)$$

where g_n represents the position in the grid G.

In the deformable convolution, the extra offset $\triangle p_n (n = 1, 2, \ldots, N)$ is added to the grid G, where N is the number of grids in the set G, and then, we can obtain

$$O_{fm}(g_0) = \sum_{g_n \in G} w(g_n) \cdot I_{fm}(g_0 + g_n + \triangle p_n). \qquad (7)$$

Then, the current sampling point is at the irregular offset position $g_n + \Delta p_n$.

The feature map is calculated on the input image by standard convolution operation, and then, the position offset is obtained by another convolution layer. The dimension of the generated channel is $2N$, namely $N \times$ 2-D offsets (corresponding to the offsets in the X and Y axes). The learned offset is then added to the original standard convolution operation to form an irregular convolution, which makes the sampling points closer to the shape of the targets.
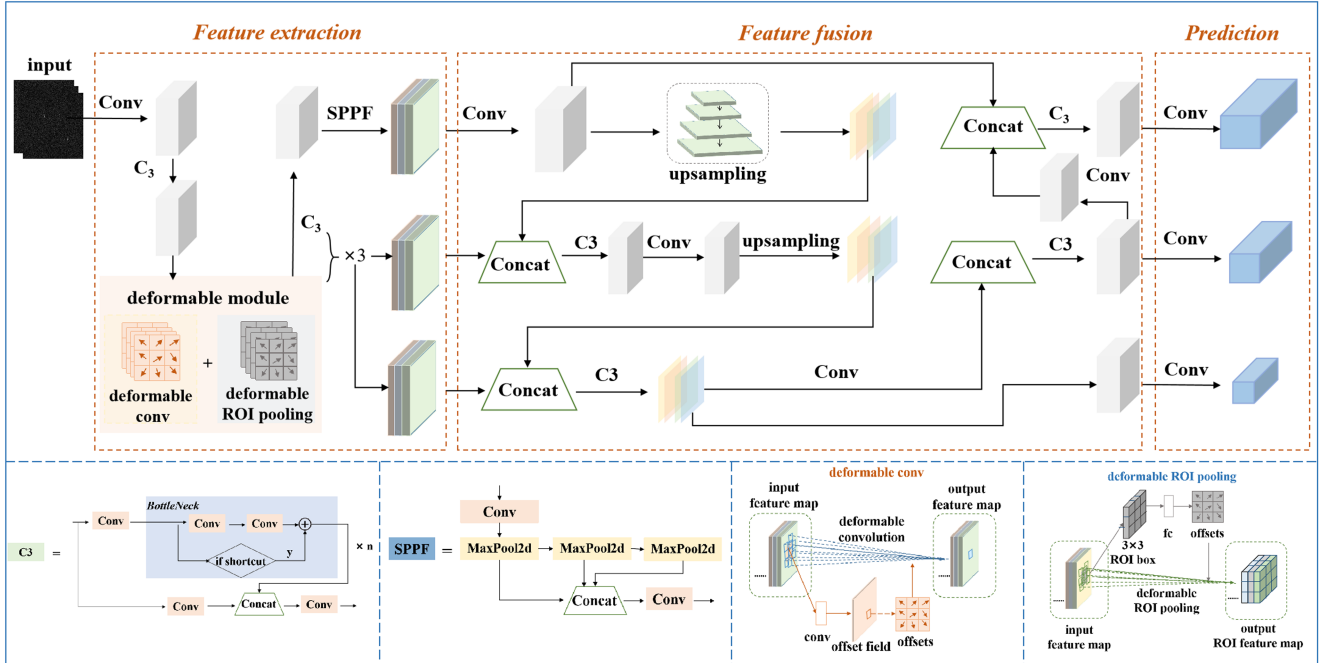
Fig. 5.   Network structure model of ASLE-DM.

2) Deformable ROI Pooling: ROI pooling is commonly applied in the object detection network with region proposal structure to convert the input feature map with arbitrary size to the fixed-size output so as to realize feature dimension reduction and data compression. Deformable pooling adds an offset for each box position and learns the offset from the prior feature map and ROI to implement the adaptive positioning of input with different shapes.

Given the input feature map X, the size of ROI is $w \times h$, and the upper left corner is p_0. After pooling operation, ROI will be divided into several $k \times k$ boxes, and a $k \times k$ feature map will be output, in which k is the kernel size. For the $(i, j)(i \geq 0, j < k)$ box in standard ROI pooling, the output $y(i, j)$ can be expressed as

$$y(i, j) = \sum_{p \in \text{box}(i,j)} X(p_0 + p)/n_{ij} \qquad (8)$$

where n_{ij} is the total number of pixels in the box. Similarly, the offset Δp_{ij} is added to the position of the spatial group box to obtain

$$y(i, j) = \sum_{p \in \text{box}(i,j)} X(p_0 + p + \triangle p_{ij})/n_{ij}. \qquad (9)$$

B. Basic Structure of ASLE-DM

The network is required to adaptively adjust the size of receptive field to learn more effective feature information when confronted with input targets with different shapes. As a consequence, the ASLE-DM network model structure based on the spatial deformable module is proposed in this article to offer a more flexible and effective ROI feature extraction mode and

improve the adaptive modeling capability of geometric transformations. The network structure of ASLE-DM is shown in Fig. 5. The ASLE-DM network structure mainly consists of three modules.

1) *Multireceptive field feature extraction:* Original fixed-size convolution and pooling operation are substituted into the deformable module described in Section III-A. The size of the receptive field is adaptively adjusted according to the input geometric transformations. The spatial sampling positions of convolution and pooling layer are strengthened by using additional offset to obtain more diverse feature information. This module extracts the high-level semantic information of the input feature map (downsampling) and generates a feature map, which is more suitable for subsequent intensive prediction. It has strong adaptability to model the input targets with different shapes.

2) *Shallow feature fusion:* Inspired by the feature pyramid network [26], shallow feature information (such as edges and textures) and high-level features are fully fused on multiscale feature maps through splicing and addition modes. In this way, the feature map has not only rich semantic information but also accurate position information of the target, which helps the network understand the position and shape of the monitored targets in subsequent prediction better.

3) *Multiobject prediction:* The prediction head is stacked by a series of convolution layers and fully connected layers. Based on the preset prior anchors, multidimensional arrays that record the coordinate, width, and height, category and confidence of the detection boxes on the input feature maps are calculated by convolution layers. Afterward,
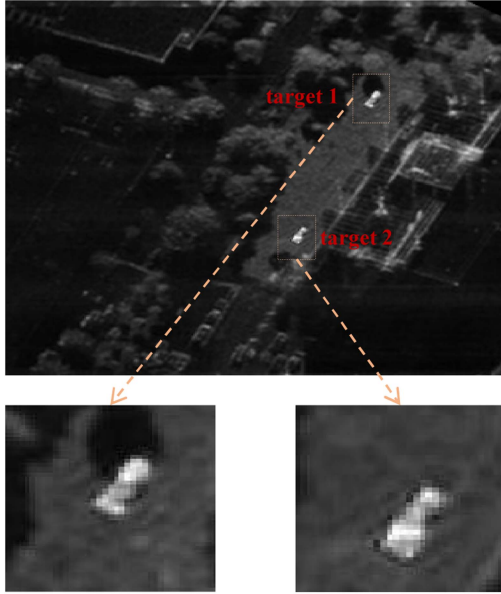
Fig. 6. Rotated SAR scene and moving targets.

Fig. 7. Receptive field and sampling position of standard and deformable module. (a) Standard convolution. (b) Deformable module.

each bounding box will be further classified via fully connected layers and softmax function, and the coordinate, width, and height of the detection box will be regressed by linear transformation, fully connected layers, and sigmoid function simultaneously. Ultimately, the parameters of the detection boxes are spliced and output as arrays after filtered by several postprocessing steps to obtain the final detection results.

C. Effective Receptive Field

Due to the change of observation angle, the imaging results of SAR scenes may rotate at different angles, as shown in Fig. 6. For the neural network, the shapes of the moving targets under surveillance are also transformed, and the morphological characteristics of such targets are not completely the same as those of standard rectangular targets. Whether the network can adjust to the geometry transformations is the key to realizing robust detection of the moving targets under different backgrounds.

Taking two CNN layers with 3×3 convolution kernel as an example, the outcome of substituting the original module with the spatial deformable module is shown in Fig. 7. In the process of standard convolution operation, the receptive field and spatial sampling positions on the feature maps are entirely fixed [see Fig. 7(a)], so it is impossible to accurately model and learn the features of input with different shapes. In comparison, the size of receptive field and the spatial sampling positions are adaptively adjusted to the input features in the deformable module [see Fig. 7(b)], which makes the spatial sampling points closer to the real shapes of the targets and optimizes the detection accuracy of the network. When encountering the input with different shapes, sizes, or observation angles, the network can still robustly detect the targets without image registration or other operations.
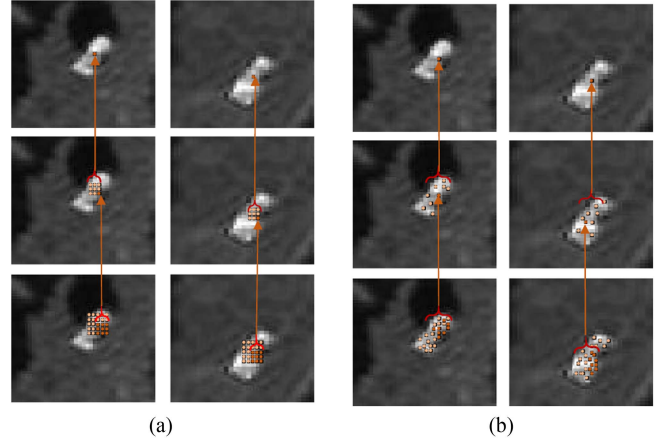
IV. ROBUST DETECTION OF MOVING TARGETS UNDER DIFFERENT BACKGROUNDS BASED ON CLUTTER SUPPRESSION

The intelligent detection of moving targets under a complex background is easily disturbed by background clutter. The scattering coefficients of artificial buildings, rocks, vegetation, and other ground objects are not much different from those of the monitored vehicles, and they show similar characteristics in imaging results, which will bring a lot of false alarms to the detection results. In order to improve the performance of the network under different backgrounds, the imaging results need to be further processed by the clutter suppression technology to alleviate the interference from background clutter [27], [28], [29].

Compared with traditional single-channel moving target detection methods, the multichannel SAR-GMTI system can effectively suppress background clutter, reserve the information of moving targets, and effectively detect ground moving targets. At present, the multichannel SAR-GMTI methods, such as displaced phase center antenna (DPCA) [30], along-track interferometry, and space-time adaptive processing [31], [32], can be flexibly selected for different SAR systems. Based on clutter suppression technology, the false alarms in the detection results can be significantly mitigated to realize the robust detection of ground moving targets under different backgrounds.

The geometric observation model of the multichannel airborne SAR system is shown in Fig. 8 (number of channels: $N \geq 2$). The radar platform moves along the x-axis (azimuth direction) at a constant speed of v, and the flight height is H. R_0 is the slant range of the scene center. N antennas are placed along the track, in which antenna 1 is a duplexer, antennas $2, 3, \ldots, N$ are receiving antennas, and the distance between two adjacent antennas is d. Antenna 1 transmits LFMCW signals, and then, all antennas receive the echo signal simultaneously. Suppose that there is a point P in the imaging scene. At the azimuth slow time $t = 0$, the coordinate of the nth antenna is $(-(n-1)d, 0, H)(n = 1, 2, \ldots, N)$, and the initial coordinate of the point P is $(x_0, y_0, 0)$. The azimuth velocity of moving target is v_a, and the radial velocity is v_r. After a certain time, the coordinate of moving target is changed to $(x', y', 0)$, and $R_n(t)$
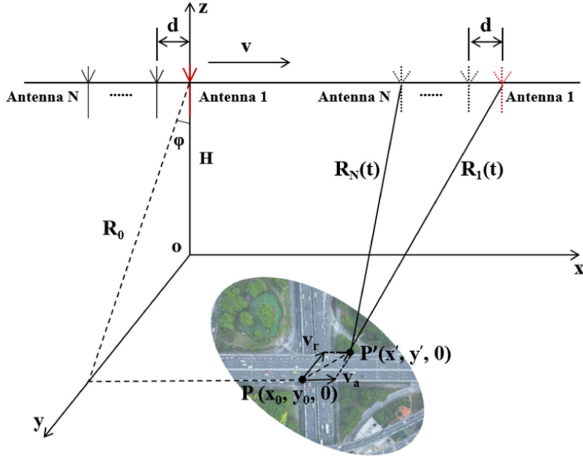
Fig. 8. Geometric observation model of multichannel SAR.

represents the instantaneous slant range ($n = 1, 2, \ldots, N$) between the nth antenna and the moving target.

In this article, take dual-channel DPCA technology [30] as an example to suppress background clutter, and then, the number of channels is $N = 2$. Combined with the signal model described in Section II-A, for the dual-channel airborne SAR system, the instantaneous slant range between the moving target and antennas 1 and 2 can be, respectively, expressed as

$$R_1(t) = \sqrt{(x_0 + v_a t - vt)^2 + (y_0 - v_y t)^2 + H^2} \quad (10)$$

$$R_2(t) = \sqrt{(x_0 + v_a t - vt + d)^2 + (y_0 - v_y t)^2 + H^2} \quad (11)$$

where $v_y = v_r / \sin\varphi$ and φ is the downlooking angle of radar. The echo signal after demodulation can be expressed as

$$S_i(\tau, t) = A_0 \text{rect}\left(\tau - \frac{2R_i(t)}{c}\right) \exp\left(-j4\pi f_0 \frac{R_i(t)}{c}\right)$$
$$\exp\left(j\pi K_r (\tau - \frac{2R_i(t)}{c})^2\right). \quad (12)$$

The obtained echo signal after range pulse compression can be expressed as

$$S_{rc,i}(\tau, t) = A' \rho_r \left(\tau - \frac{2R_i(t)}{c}\right) \exp\left(-j4\pi f_0 \frac{R_i(t)}{c}\right) \quad (13)$$

where $i = 1$ and 2 represent antennas 1 and 2, respectively.

In the actual SAR system, inconsistency of radar antenna pattern, error of antenna phase center, instability of platform velocity, noise, and environmental interference will cause the amplitude and phase error between channels [33], which will greatly limit the performance of clutter suppression. It is particularly important to study how to realize the robust detection of moving targets even if there exists channel error. Therefore, the amplitude and phase error are further added to channel 2, and the echo signal of channel 2 after demodulation is $S'_{rc,2}(\tau, t)$. After adding the channel error, the echo signal of channel 2 can

TABLE I
SYSTEM SIMULATION PARAMETERS

Parameter	Symbol	Value
Pulse repetition frequency	PRF	1872 Hz
Orbit height	H	800 km
Platform velocity	v	7800 m/s
Range bandwidth	B_r	20 MHz
Carrier frequency	f_0	20 GHz
Scene center distance	R_0	850 km
Radial speed of the moving target	v_r	4 m/s
Azimuth speed of the moving target	v_a	4 m/s

be expressed as

$$S'_{rc,2}(\tau, t) = A_e \cdot S_{rc,2}(\tau, t) \cdot \exp(j * P_h) \quad (14)$$

where A_e stands for the factor of amplitude error, and P_h stands for the factor of phase error.

Then, the echo signal of channels 1 and 2 after range compression is, respectively, simulated according to the steps described in Section II-B. The echo signal after phase compensation is coherently accumulated to obtain the imaging results of each pixel in the background

$$I_i(x, y) = \sum_{t=t_1}^{t=t_2} S_{rc,i}(\tau_{xy}, t) \exp(j2\pi f_0 \tau_{xy}) \quad (15)$$

where $i = 1$ and 2 represent antennas 1 and 2, respectively.

Finally, the echo data of the two channels after imaging are further processed by the DPCA technology in the complex image domain, which can be expressed as

$$I_{1,2}(x, y) = I_1(x, y) - I_2(x, y). \quad (16)$$

Consequently, most of the background clutter is significantly suppressed, and the moving target information is sufficiently reserved.

V. EXPERIMENT

Before applying the ASLE-DM network to detect the ground moving targets, it is necessary to complete the construction of moving target simulation sample sets. In this article, the BP imaging algorithm is taken as an example to simulate moving targets and diverse SAR scenes. The simulation parameters of the airborne SAR system are set in Table I.

A. Construction Results of Ground Moving Target Simulation Sample Sets Under Different Backgrounds

The shape of the vehicle to be monitored can be approximately considered as an irregular rectangle, and the number of moving targets in the scene is set within 5–10 according to the characteristics of each scene. In the process of moving target simulation, such parameters as size, shape, number, and position of moving target can be set arbitrarily, so it is easy to obtain a large number of simulation sample sets for the experiment, and some imaging results are shown in Fig. 9(a)–(d). In order to make the imaging results closer to practical applications, noise component is added
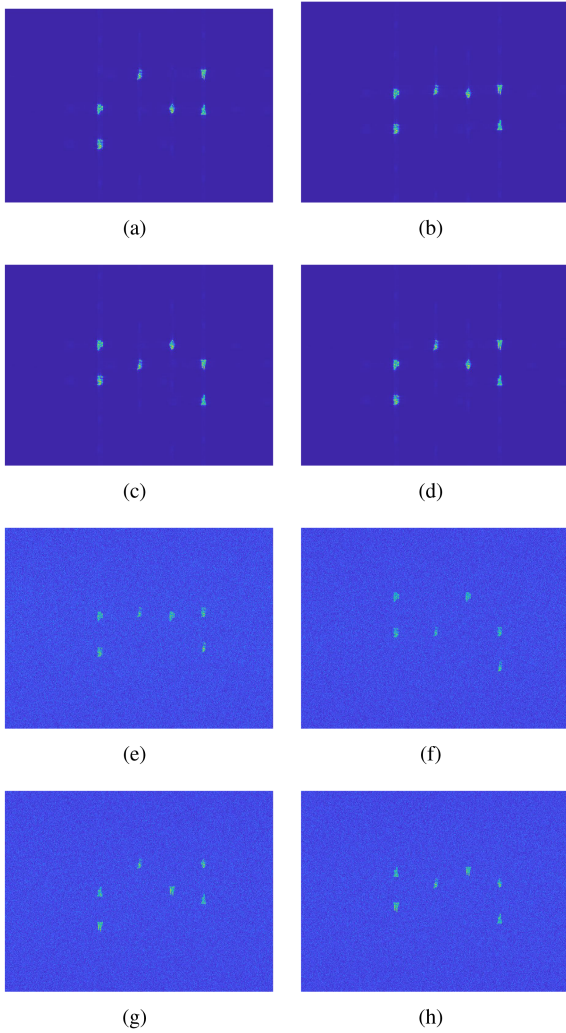
(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig. 9. Simulation results of ground moving targets. (a)–(d) Imaging results of moving targets. (e)–(h) Imaging results of moving targets under the noise background.
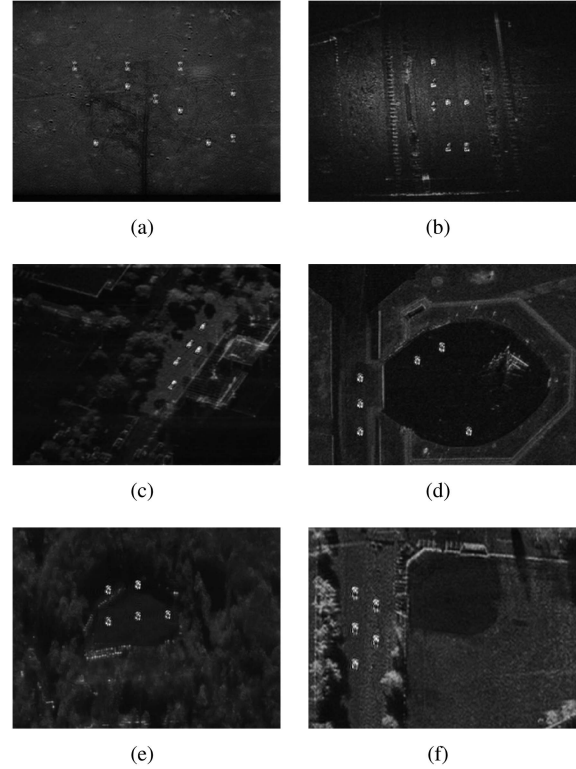

(a) (b)

(c) (d)

(e) (f)

Fig. 10. Simulation sample sets of moving targets under different backgrounds. (a) Scene 1 (desert). (b) Scene 2 (urban road). (c) Scene 3 (urban building). (d) Scene 4 (apron). (e) Scene 5 (forest). (f) Scene 6 (baseball diamond).

to the echo simulation of moving targets, and the signal-to-noise ratio is 15 dB, as shown in Fig. 9(e)–(h). The obtained moving target sample sets under the noise background are used for network training.

The azimuth position coordinate of moving target is traversed within a reasonable range (step size: 1). Accordingly, 350 SAR image sample sets under the noise background are obtained, and then, the quantity is expanded to 4000 by image enhancement (image rotation, cropping, translation, etc.). The obtained moving target image sample sets under the noise background are fed into neural network for training. Furthermore, the echo signal of the moving target is separately fused with that of each SAR scene shown in Fig. 3 to obtain the test sample sets. Some of the imaging results are shown in Fig. 10.

In particular, the variation of observation angle of some SAR systems (such as circular spotlight SAR) will lead to the rotation of the imaging scene. Therefore, the simulation of moving target at a certain rotation angle (30°) is added to construct the test sample set, as shown in Fig. 9(c). This usually requires image

registration for accurate detection, but it will also generate additional calculation load and work complexity. Comparatively, the ASLE-DM network proposed in this article can still realize the robust detection of moving targets in the rotated scene due to its strong adaptive modeling capability of the input targets with different shapes. So far, the construction of diverse simulation sample sets has been completed.

B. Results of Moving Target Detection Using ASLE-DM

The simulation sample set of moving targets under the noise background constructed in Section V-A is input into the ASLE-DM network for training, and the trained model is used to detect the SAR images under different backgrounds in the test set. The partial detection results of moving targets under different backgrounds obtained by using the ASLE-DM network are shown in Fig. 11(s)–(x).

Except that some sandstones show strong scattering properties, the overall characteristics of scene 1 (desert) are similar to the noise background, and most of the moving targets in this scene can be correctly detected. However, some moving targets are submerged in background clutter, which will produce a certain number of missing alarms, as shown in Fig. 11(s). For the moving targets in other scenes, although the imaging results of the vegetation, artificial buildings, and some strong scattering points in the scene are similar to the shapes of targets, which will bring some missing and false alarms to the detection results, the
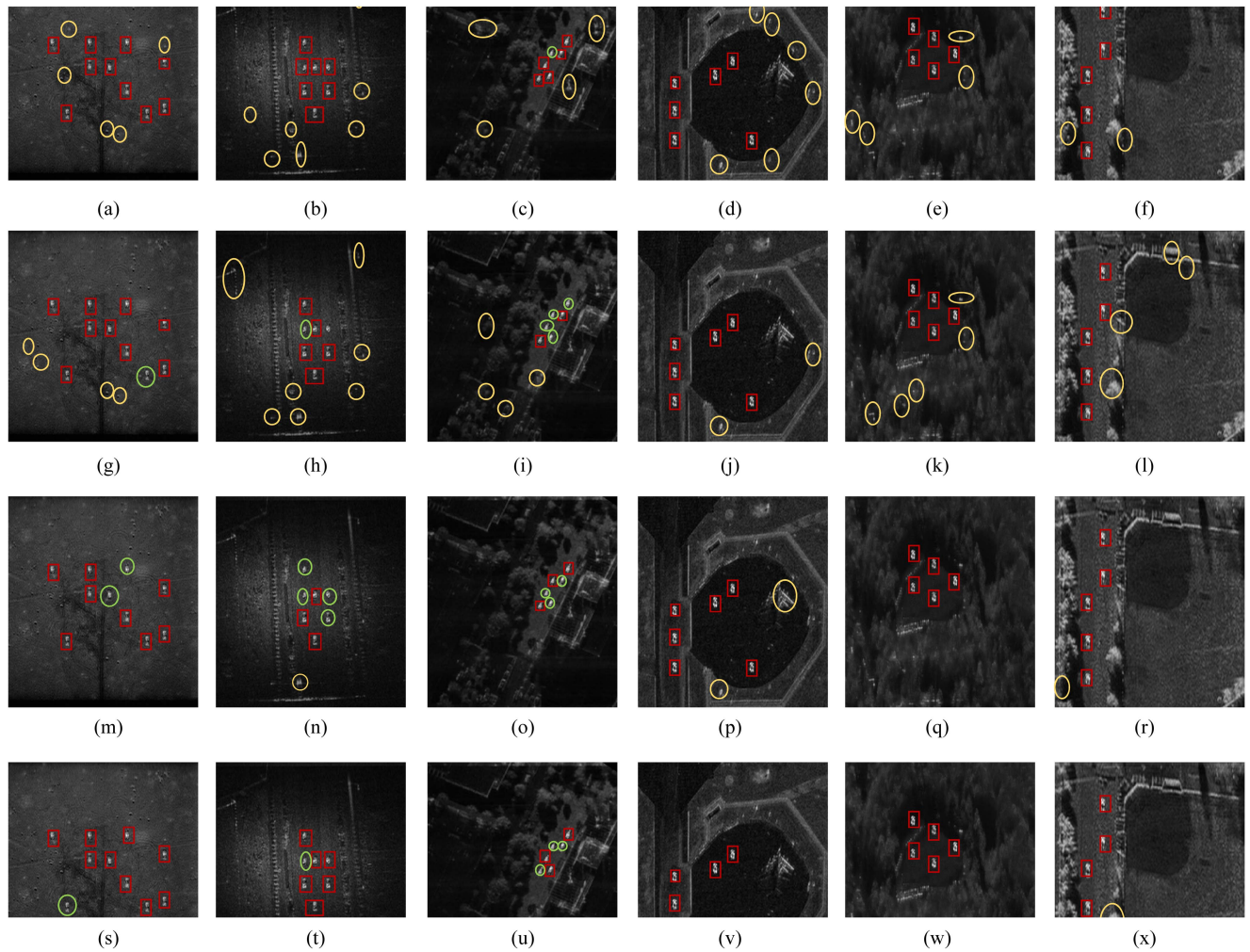
Fig. 11. (a)–(x) Detection results of moving targets under multiple backgrounds based on different methods. (Red rectangle represents correctly detected moving target, green ellipse represents missing alarm, and yellow ellipse represents false alarm.) Each row represents different detection methods (rows 1–4 are Faster-RCNN, Detr, Yolov8, and ASLE-DM, respectively), and each column represents different scenes (columns 1–6 are scenes 1–6, respectively).

ASLE-DM network can still robustly detect most of the targets under different backgrounds, as shown in Fig. 11(s)–(x). Especially, as to the moving targets in scene 3, the rotation of SAR imaging result occurs due to the change of observation angle, while most of moving targets can still be accurately detected by the ASLE-DM network due to its strong adaptive modeling capability of input targets with different shapes. Generally, the ASLE-DM network can realize robust detection of the moving targets under different backgrounds with a certain number of missing and false alarms.

C. Comparative Experiment of Different Methods

In order to further verify the effectiveness of the proposed method, the experimental results of Faster-RCNN [34], Detr [35], Yolov8 [36], and ASLE-DM are quantitatively compared in Table II. Note that the frames per second (FPS) is used to measure the processing speed of model prediction. With the results shown in Table II, it can be seen that the mean average precision (mAP) and F1-score are superior to those of other

TABLE II
COMPARISON OF DIFFERENT NEURAL NETWORKS

Methods	mAP@0.5	mAP@[0.5:0.95]	F1-score	FPS
Faster-RCNN	0.893	0.614	0.532	17
Detr	0.759	0.582	0.625	12
Yolov8	0.925	0.813	0.815	11
ASLE-DM	0.925	0.824	0.835	37

methods and the speed of model prediction is also optimized . By comparison, the ASLE-DM network comprehensively shows better performance to other methods.

In addition, although Table II reflects the better performance of the ASLE-DM model, further validation is needed to determine whether the proposed method can better realize robust detection of moving targets under different backgrounds. Therefore, the moving target simulation dataset under the noise background is fed into Faster-RCNN, Detr, and Yolov8 networks for training, respectively, and the trained model is used to detect

TABLE III
DETECTION RESULTS OBTAINED BY DIFFERENT NEURAL NETWORKS

Methods	Correctly detected targets	False alarms	Missing alarms	Recall rate	Precision
CA-CFAR	4550	1636	975	0.824	0.736
Faster-RCNN	5487	6996	38	0.993	0.440
Detr	4765	5118	760	0.862	0.482
Yolov8	4583	560	942	0.830	0.891
ASLE-DM	4807	446	718	0.870	0.915

TABLE IV
DETECTION RESULTS OF DIFFERENT NEURAL NETWORKS AFTER CLUTTER SUPPRESSION

Methods	Correctly detected targets	False alarms	Missing alarms	Recall rate	Precision
CA-CFAR	4683	1021	842	0.847	0.821
Faster-RCNN	5525	2194	0	1.000	0.716
Detr	4910	2419	615	0.889	0.670
Yolov8	4621	522	904	0.836	0.898
ASLE-DM	4938	352	587	0.894	0.933

TABLE V
ABLATION STUDY OF THE SPATIAL DEFORMABLE MODULE

Methods	Backbone	map@0.5	mAP@[0.5:0.95]	FPS	Training time
Baseline (Yolov5s)	CSPDarkNet53	0.910	0.814	29	5.072
+Deformable module (ASLE-DM)		0.925	0.824	37	4.311
Baseline (Faster-RCNN)	ResNet-50	0.893	0.614	17	7.483
+Deformable module		0.904	0.725	21	6.732

TABLE VI
DETECTION RESULTS OF DIFFERENT NETWORK ARCHITECTURES ON THE SIMULATION DATA

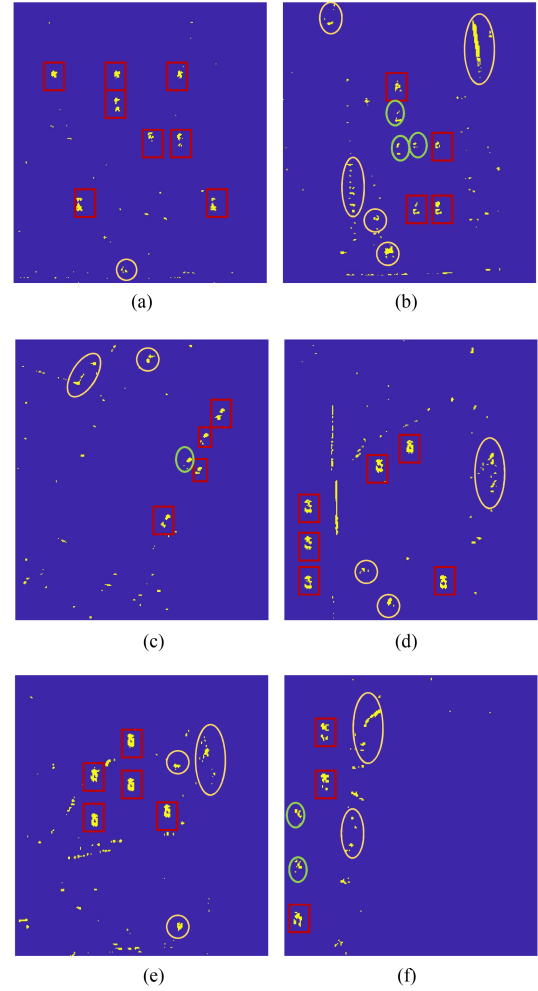| Methods	Correctly detected targets	False alarms	Missing alarms	Recall rate	Precision
Baseline (Yolov5s)	4502	441	1023	0.815	0.911
+Deformable module (ASLE-DM)	4938	352	587	0.894	0.933
Baseline (Faster-RCNN)	5513	2194	12	0.998	0.715
+Deformable module	5519	1368	6	0.999	0.801

Fig. 12. Detection results by CA-CFAR. (Red rectangle represents correctly detected moving target, green ellipse represents missing alarm, and yellow ellipse represents false alarm.) (a)–(f) Scenes 1–6.

the moving targets in the test sample set. The quantitative comparison of detection results obtained by different methods is shown in Table III. Apart from the intelligent target detection algorithms, the echo-based CA-CFAR algorithm [37], [38] is also implemented to make a comparison. The 2-D cell-averaging CFAR (CA-CFAR) method adopts the sliding window mechanism, which consists of cell under detection (CUT), guard band, and training band. When the power of CUT is greater than the detection threshold T, it is determined that there exists a target in the current unit; otherwise, there exists no target. Similarly, the above operation is conducted for the next CUT until all the cells are traversed. The detection results of CA-CFAR are also recorded in Table III to make a comparison.

The simulation test set contains SAR image samples of moving targets under six different backgrounds, and the total number of moving targets to be detected is 5525. The detection threshold (confidence) of each neural network model is set as the same to make the results comparable.

In particular, recall rate and precision indicators are further calculated to display the comparison results more intuitively.

The definition of recall rate and precision can be expressed by

$$\text{Recallrate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{18}$$

where TP, FN, and FP separately stand for true positive, false negative, and false positive, respectively. Actually, TP, FN, and FP correspond to the number of correctly detected targets, missing alarms, and false alarms, respectively.

Faster-RCNN is a typical two-stage object detection algorithm, and the detection accuracy of the algorithm is improved by introducing region proposal network structure. It can correctly detect most of moving targets in the test set with the highest recall rate, but the detection precision is extremely unacceptable, which will not operate well in practical applications. Detr is an end-to-end target detection model based on the transformer architecture, and it leverages CNN, encoding, and decoding structure to implement feature extraction and inference. However, for the relatively complex training process and insufficient
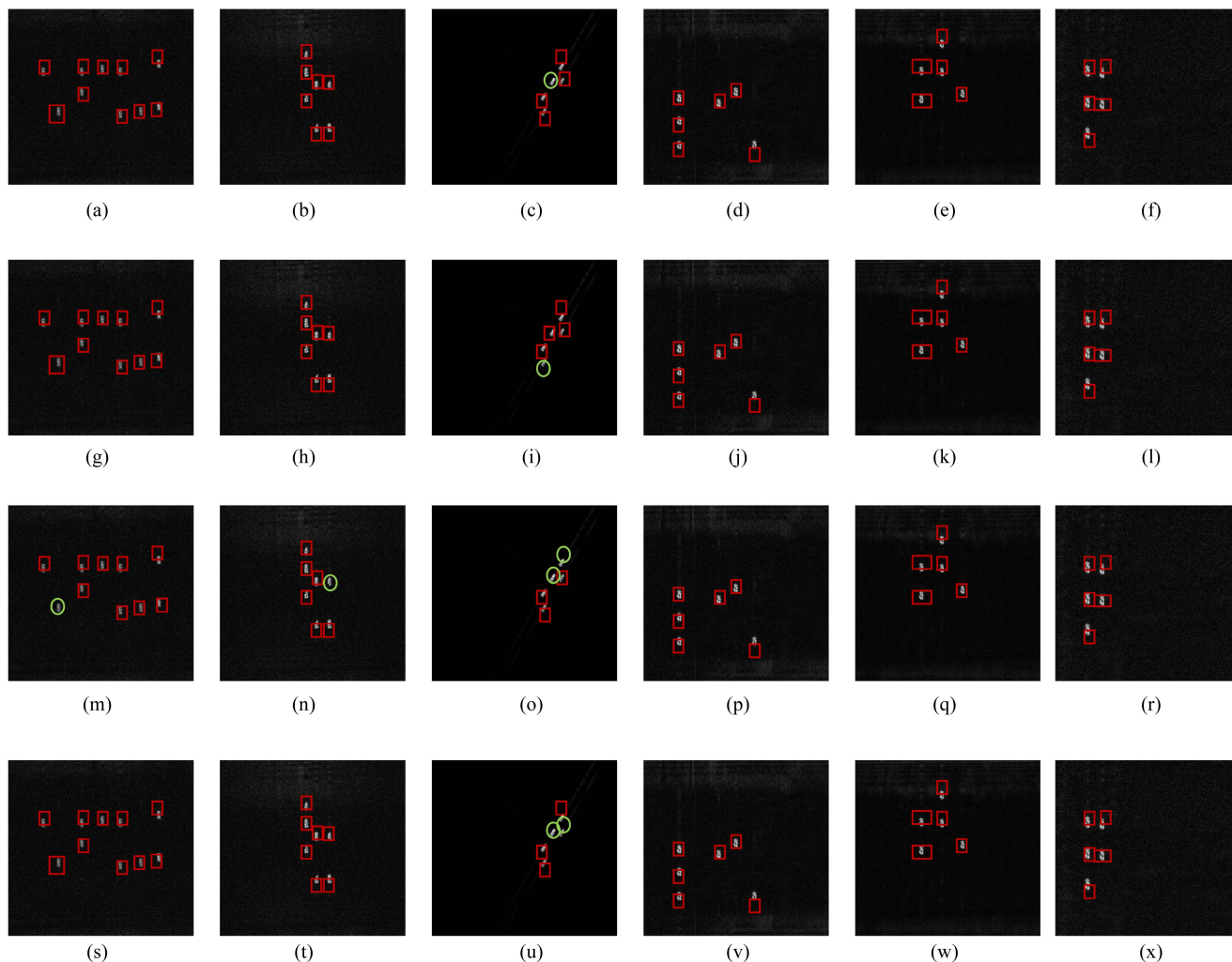
Fig. 13.   (a)–(x) Detection results of the targets after clutter suppression by different methods. (in ideal conditions). Each row represents different detection methods (row 1-4 are Faster-RCNN, Detr, Yolov8 and ASLE-DM, respectively), and each column represents different scenes (column 1-6 are scene 1-6, respectively).

feature extraction of small targets, the performance of Detr is relatively poor in detection precision. Yolov8 is a typical one-stage object detection algorithm, which regards target detection as a regression problem and inputs the whole image into the network directly for inference. According to Table III, although the detection performance of Yolov8 is relatively satisfactory with decreasing false alarms, it is still inferior to that of the ASLE-DM network. Compared to the ASLE-DM network, the detection performance of CA-CFAR is easy to be interfered by complex background clutter, which brings about several missing and false alarms consequently and make it perform badly in recall rate and precision.

Moreover, some detection results of moving target under different backgrounds based on the neural network models and CA-CFAR are shown in Figs. 11 and 12, respectively. For the detection of moving targets in desert, the echo signal of some targets is merged in background clutter, which brings several missing alarms to the detection results, as shown in Fig. 11(g), (m), and (s). And there still exist a few false alarms in the detection results of Faster RCNN and Detr, as shown in Fig. 11(a)

and (g). For the rotated imaging scene due to the change of observation angle, some targets can still be correctly detected by virtue of its adaptive modeling capability, as shown in Fig. 11(u). In relatively complex backgrounds, the number of missing and false alarms will increase in the detection results of the other three neural networks, while better detection results can be obtained by the ASLE-DM network, as shown Fig. 11(b), (h), (n), and (t).

It can be seen from Fig. 12 that there are many scattering points in the detection results of CA-CFAR, which can be ignored since the moving target to be monitored occupies a certain area instead of isolated points, and only the region composed of several pixel points is judged as a target. Combined with Table III, CA-CFAR can correctly detect partial moving targets from the noise interference. However, the detection performance of the CA-CFAR algorithm will deteriorate sharply due to the disturbance of strong background clutter, and a large number of false alarms and missing alarms will be generated especially under a relatively complex backgrounds, as shown in Fig. 12(b)–(f). By comparison, the ASLE-DM network can overall realize robust
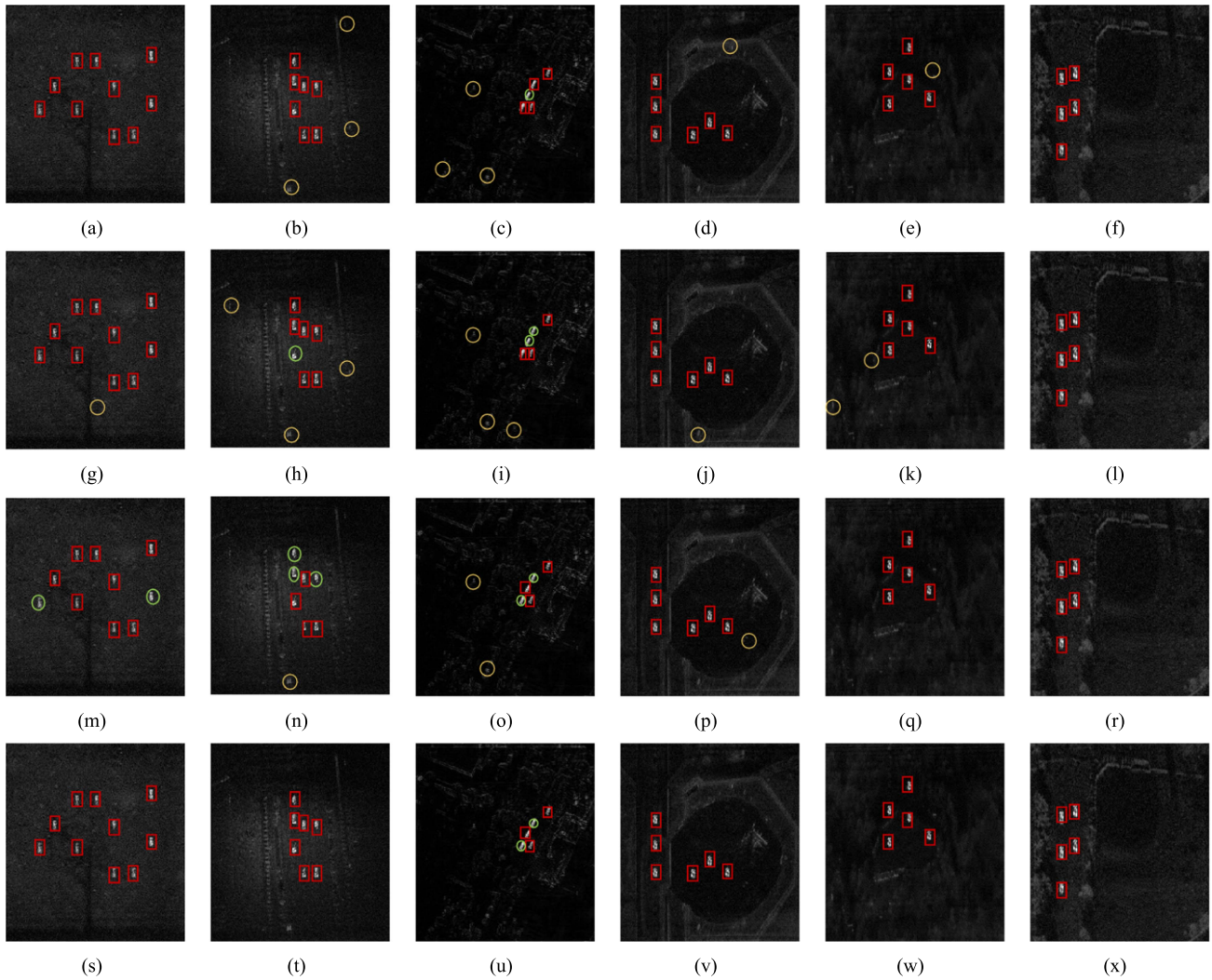
Fig. 14. (a)–(x) Detection results of the targets after clutter suppression by different methods (with channel error). Each row represents different detection methods (rows 1–4 are Faster-RCNN, Detr, Yolov8, and ASLE-DM, respectively), and each column represents different scenes (columns 1–6 are scenes 1–6, respectively).

detection of the moving targets under different backgrounds and shows better detection performance than the other methods with fewer false and missing alarms.

D. Robust Moving Target Detection Based on Clutter Suppression

Although the ASLE-DM network can correctly detect most of the moving targets under different backgrounds according to the comparative experimental results in Section V-C, there still exist a certain number of false alarms and missing alarms in the detection results due to the interference of complex background clutter. In order to further improve the detection performance of the ASLE-DM network, multichannel clutter suppression technology is combined to suppress background clutter significantly. Taking dual-channel DPCA technology ($d = 1$ m) as an example, the network is used to detect the SAR images under different backgrounds after clutter suppression. From the simulation results of clutter suppression, in the ideal simulation

conditions (that is, there exists no channel error), the background clutter can be significantly suppressed, and the moving target component is fully reserved in the final imaging results. On this basis, almost all of the moving targets in the test sample set can be correctly detected, as shown in Fig. 13. Consequently, the proposed method can effectively realize the robust detection of moving targets under different backgrounds.

However, in the actual SAR system, the channel error (such as antenna phase center error, inconsistent response of each channel pattern, and receiver thermal noise) exists inevitably, which will restrict the performance of clutter suppression technology and inhibit the detection accuracy of ground moving targets. In the case of amplitude and phase error between channels ($A_e = 1.08$, $P_h = -\pi/180$), the experimental results of CA-CFAR, Faster-RCNN, Detr, Yolov8, and ASLE-DM are quantitatively compared in Table IV. From Table IV, except that the detection performance of the ASLE-DM network is comprehensively superior to the other methods, clutter suppression technology can also effectively improve the detection results of different
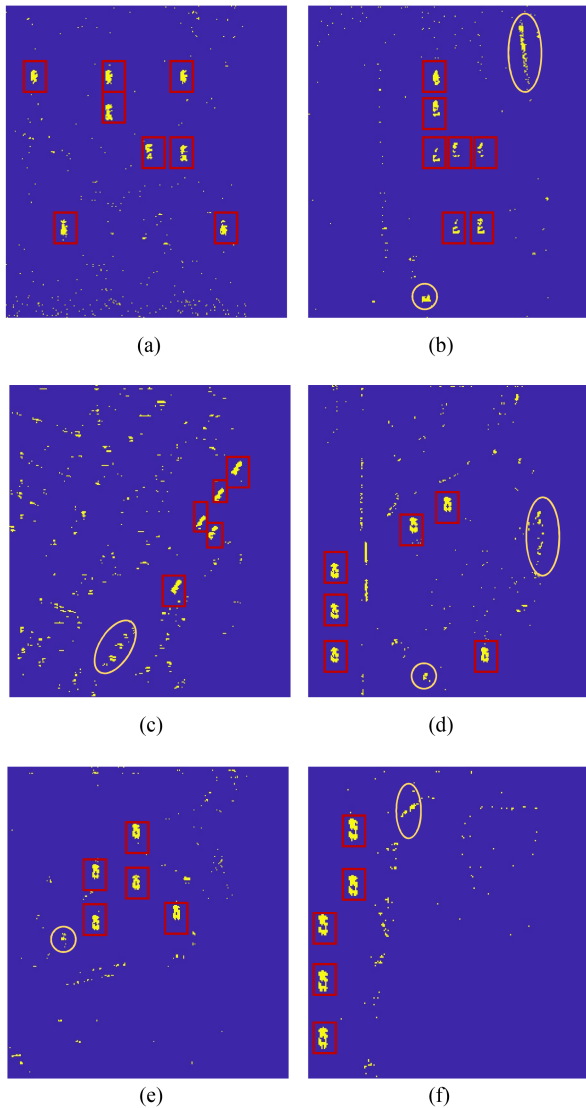
Fig. 15. Detection results of CA-CFAR based on clutter suppression. (Red rectangle represents correctly detected moving target, green ellipse represents missing alarm, and yellow ellipse represents false alarm.) (a)–(f) scenes 1–6.

methods as a whole. In particular, combined with Table III, the recall rate of the detection results by using the ASLE-DM network increases (from 0.87 to 0.894 and the precision increases from 0.915 to 0.933), and so are the other methods.

Moreover, the corresponding detection results after clutter suppression by different methods are partially shown in Figs. 14 and 15. Due to the existence of channel error, the background clutter cannot be completely suppressed, and there is still some residual clutter in the SAR imaging results. Compared with the detection results in Fig. 11, the false alarms and missing alarms have been significantly mitigated, such as Fig. 14(d)–(v). From Fig. 15, it can also be seen that the detection performance of CA-CFAR is also improved when compared to Fig. 12. Meanwhile, even if there is residual clutter in the background, the ASLE-DM network shows better detection performance with fewer false alarms and missing alarms generated in the detection results,

which can effectively realize the robust detection of moving targets under different backgrounds.

As to the intelligent target detection algorithms, further study of the detection results at different confidence levels is conducted for supplement to demonstrate the impact of detection threshold on the number of missing and false alarms. The detection threshold (confidence) is set as 0.65, 0.7, 0.75, 0.8, and 0.85, respectively, and the relevant experimental results obtained by different detection methods are calculated, as shown in Fig. 16.

The total number of missing and false alarms in detection results obtained by different methods is compared in Fig. 16(a) and (b), respectively. As the detection threshold increases, the number of missing alarms embodies an upward trend, while the number of false alarms declines as a whole. From Fig. 16(a), it is not difficult to see that the number of missing alarms obtained by ASLE-DM is relatively less than those by other methods. In particular, although Faster-RCNN shows slightly better performance to the proposed method in missing alarms, much more false alarms are produced by Faster-RCNN combined with Fig. 16(b), which is unacceptable in practical applications. Moreover, from Fig. 16(b), the detection results of false alarms by ASLE-DM are mostly better than the other methods. Although Yolov8 shows similar performance to the proposed method in false alarms, the missing alarms extremely increases in Fig. 16(a). Therefore, it is well verified that the proposed method shows superior detection performance to the other methods mentioned in this article.

E. Ablation Study of the Spatial Deformable Module

In order to further verify the effectiveness of spatial deformable module, comparative experiments with different network architectures are conducted, and relevant results are shown in Table V. All of the experiments are conducted on the platform NVIDIA GeForce RTX 3090 GPU, the deep learning framework is PyTorch (version 1.10.0), and the training process consists of 200 epochs.

The ablation study of the deformable module is conducted by using Yolov5s and Faster-RCNN as baseline networks, in which CSPDarkNet53 and ResNet-50 are used as the backbone architectures, respectively. On this basis, further experiment of applying baseline architecture with three deformable layers on the simulation datasets is conducted to make a comparison.

From Table V, it is verified that the proposed architecture is superior to other methods and thus will show better detection performance of the targets in the test sample sets. The deformable module can effectively increase the mAP value of the model, and the prediction speed and the training time of the model are also optimized. As described in Section III, the spatial deformable module can further enhance the sampling positions of the feature maps and fully preserve high-level semantic information of ROI in SAR images during the process of feature extraction, which will significantly promote the final detection results of the moving targets.

In addition, further comparative experiments of the network architectures are performed on the detection results of the
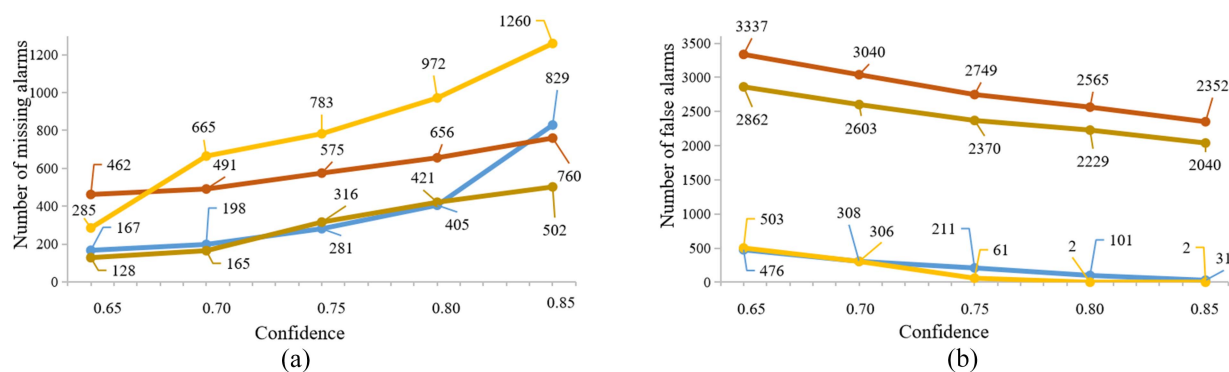
Fig. 16. Calculation of detection results at different confidence levels. (a) Number of missing alarms. (b) Number of false alarms. (The blue, brown, red, and yellow lines correspond to ASLE-DM, Faster-RCNN, Detr, and Yolov8, respectively.)

simulation test sample sets, and the relevant results are recorded in Table VI.

As given in Table VI, when using the trained network model to detect the SAR imaging results after clutter suppression in test sample sets, the recall rate and precision of the detection results are effectively increased due to the application of the spatial deformable module. By comparison, it is verified that the proposed method can effectively improve the detection performance of the model by virtue of its more powerful ability of feature representation and can further realize robust detection of the moving targets under different backgrounds.

Moreover, it is noted that although the introduction of the deformable module can effectively optimize the detection performance of the model with more correctly detected targets and fewer false alarms generated, it also brings more parameters and raises the calculation complexity of the model, which will significantly increase the training time. In particular, during the experiment process, it is found that the training results and detection performance of the model are no longer promoted remarkably with three more network layers substituted by the deformable module, while the training time significantly increases instead. Therefore, the number of the network layers substituted by the deformable module depends comprehensively on the calculation complexity and detection performance of the model.

VI. CONCLUSION

In this article, a new method of SAR ground moving target detection is proposed on the basis of associating the ASLE-DM network with traditional clutter suppression technology, which focuses on optimizing the detection performance of moving targets in varying SAR scenes for further practical applications. Combined with the SAR imaging algorithm, the simulation sample set is constructed by traversing the position parameters of moving targets, which can tackle the problem of insufficient measured data in the SAR-GMTI algorithm based on deep learning. In order to enhance the detection performance of polymorphous input targets, the spatial deformable module is applied to make it capable to model various geometry transformations. Moreover, the multichannel clutter suppression technology is used to suppress the background clutter and then realize robust

detection of the moving targets under different backgrounds. The effectiveness and robustness of the proposed method are further verified compared with other detection methods. Eventually, it is acknowledged that although we pursue ensuring the diversity and representativeness of the simulation data, the experimental sample data are impossible to cover all the SAR scenes, which means that the proposed method still needs to be replenished and improved continuously with the accumulation of diverse SAR sample data.

REFERENCES

[1] H. Xu, Z. Yang, M. Tian, Y. Sun, and G. Liao, "An extended moving target detection approach for high-resolution multichannel SAR-GMTI systems based on enhanced shadow-aided decision," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 715–729, Feb. 2018.

[2] J.-C. Ni, Y. Luo, D. Wang, J. Liang, and Q. Zhang, "Saliency-based SAR target detection via convolutional sparse feature enhancement and Bayesian inference," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5202015.

[3] Z. Wang et al., "Visual attention-based target detection and discrimination for high-resolution SAR images in complex scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1855–1872, Apr. 2018.

[4] M. Newey, G. R. Benitz, D. J. Barrett, and S. Mishra, "Detection and imaging of moving targets with LiMIT SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3499–3510, Jun. 2018.

[5] C. Han, Z. Yang, G. Liao, Q. Zhang, S. He, and H. Xu, "A SAR-GMTI approach aided by online knowledge with an airborne multichannel quad-pol radar system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8668–8681, 2022.

[6] B. Liu, K. Yin, Y. Li, F. Shen, and Z. Bao, "An improvement in multichannel SAR-GMTI detection in heterogeneous environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 810–827, Feb. 2015.

[7] B. Ge, D. An, J. Liu, D. Feng, L. Chen, and Z. Zhou, "Modified adaptive 2-D calibration algorithm for airborne multichannel SAR-GMTI," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4004805.

[8] G. Gao, X. Wang, and T. Lai, "Detection of moving ships based on a combination of magnitude and phase in along-track interferometric SAR—Part II: Statistical modeling and CFAR detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3582–3599, Jul. 2015.

[9] Y. Guo, G. Liao, J. Li, and X. Chen, "A novel moving target detection method based on RPCA for SAR systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6677–6690, Sep. 2020.

[10] B. Ge, D. An, L. Chen, W. Wang, D. Feng, and Z. Zhou, "Ground moving target detection and trajectory reconstruction methods for multichannel airborne circular SAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2900–2915, Aug. 2022.

[11] J. Yang, C. Liu, and Y. Wang, "Detection and imaging of ground moving targets with real SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 920–932, Feb. 2015.

[12] P. Huang et al., "Imaging and relocation for extended ground moving targets in multichannel SAR-GMTI systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5214024.

[13] H. Wan et al., "AFSar: An anchor-free SAR target detection algorithm based on multiscale enhancement representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5219514.

[14] H. Mu, Y. Zhang, Y. Jiang, and C. Ding, "CV-GMTINet: GMTI using a deep complex-valued convolutional neural network for multichannel SAR-GMTI system," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5201115.

[15] Z. Wang et al., "Land-sea target detection and recognition in SAR image based on non-local channel attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5237316, doi: 10.1109/TGRS.2022.3217035.

[16] D. Meng et al., "On the processing of very high resolution spaceborne SAR data: A chirp-modulated back projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 191–201, Jan. 2018.

[17] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.

[18] G. Xu, Y. Chen, A. Ji, B. Zhang, C. Yu, and W. Hong, "3-D high-resolution imaging and array calibration of ground-based millimeter-wave MIMO radar," *IEEE Trans. Microw. Theory Techn.*, pp. 1–13, 2024.

[19] G. Xu, B. Zhang, H. Yu, J. Chen, M. Xing, and W. Hong, "Sparse synthetic aperture radar imaging from compressed sensing and machine learning: Theories, applications, and trends," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 32–69, Dec. 2022.

[20] Y. Cheng, J. Ding, Z. Sun, and C. Zhong, "Processing of airborne video SAR data using the modified back projection algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5238013.

[21] L. M. Ulander, H. Hellsten, and G. Stenstrom, "Synthetic-aperture radar processing using fast factorized back-projection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 760–776, Jul. 2003.

[22] G. Cheng et al., "Prototype-CNN for few-shot object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5604610.

[23] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10070–10087, Dec. 2019.

[24] J. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 764–773.

[25] Z. Xia, X. Pan, S. Song, L. E. Li, and G. Huang, "Vision transformer with deformable attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4784–4793.

[26] Z. Li, E. Li, T. Xu, A. Samat, and W. Liu, "Feature alignment FPN for oriented object detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6001705.

[27] E. Makhoul, S. V. Baumgartner, M. Jäger, and A. Broquetas, "Multichannel SAR-GMTI in maritime scenarios with F-SAR and TerraSAR-X sensors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5052–5067, Nov. 2015.

[28] K. Suwa, K. Yamamoto, M. Tsuchida, S. Nakamura, T. Wakayama, and T. Hara, "Image-based target detection and radial velocity estimation methods for multichannel SAR-GMTI," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1325–1338, Mar. 2017.

[29] W.-Q. Wang and H. Shao, "Two-antenna SAR with waveform diversity for ground moving target indication," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2154–2158, Dec. 2014.

[30] D. Cerutti-Maori and I. Sikaneta, "A generalization of DPCA processing for multichannel SAR/GMTI radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 560–572, Jan. 2013.

[31] T. K. Sjogren et al., "Suppression of clutter in multichannel SAR GMTI," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4005–4013, Jul. 2014.

[32] S. Hu, J. Yi, X. Wan, F. Cheng, and Y. Tong, "Doppler separation-based clutter suppression method for passive radar on moving platforms," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4701714.

[33] F. Xiao, Z. Ding, Z. Li, and T. Long, "Channel error effect analysis for reconstruction algorithm in dual-channel SAR imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1563–1567, Sep. 2020.

[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[35] M. L. Mekhalfi, C. Nicolo, Y. Bazi, M. M. A. Rahhal, N. A. Alsharif, and E. A. Maghayreh, "Contrasting YOLOv5, transformer, and EfficientDet detectors for crop circle detection in desert," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 3003205.

[36] Q. Guo, J. Liu, and M. Kaliuzhnyi, "YOLOX-SAR: High-precision object detection system based on visible and infrared sensors for SAR remote sensing," *IEEE Sens. J.*, vol. 22, no. 17, pp. 17243–17253, Sep. 2022.

[37] C. Kuang, C. Wang, B. Wen, Y. Hou, and Y. Lai, "An improved CA-CFAR method for ship target detection in strong clutter using UHF radar," *IEEE Signal Process. Lett.*, vol. 27, pp. 1445–1449, 2020.

[38] A. Maali, A. Mesloub, M. Djeddou, H. Mimoun, G. Baudoin, and A. Ouldali, "Adaptive CA-CFAR threshold for non-coherent IR-UWB energy detector receivers," *IEEE Commun. Lett.*, vol. 13, no. 12, pp. 959–961, Dec. 2009.
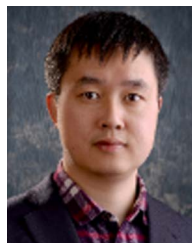
He Yan received the B.S. degree in electronic information science and technology from Jilin University, Changchun, China, in 2008, and the Ph.D. degree in communication and information system from the University of Chinese Academy of Sciences, Beijing, China, in 2013 .

From 2013 to 2014, he was an Engineer with the 14th Research Institute of China Electronics Technology Group Corporation, Nanjing, China. Since 2015, he has been an Associate Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing. His main research interests include synthetic aperture radar ground moving target indication, wide area surveillance ground moving target indication, video synthetic aperture radar ground moving target indication, and ocean current measurements based on spaceborne synthetic aperture radar.

Hui Liu was born in Huaian, China, in 2001. She received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2018. She is currently working toward the M.S. degree in communication and information systems with the Nanjing University of Aeronautics and Astronautics, Nanjing.

Her research interests include signal and information processing.

Jialin Hao was born in Changchun, China, in 2000. She received the B.S. degree in communication engineering from Jilin University, Changchun, in 2018. She is currently working toward the M.S. degree in communication and information systems with the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Her research interests include signal and information processing.

Wenshuo Xu was born in Nantong, China, in 2000. She received the B.S. degree in electronic information engineering from the Nanjing University of Information Science and Technology, Nanjing, China, in 2018. She is currently working toward the M.S. degree in communication and information systems with the Nanjing University of Aeronautics and Astronautics, Nanjing.

Her research interests include signal and information processing.

Jingdong Zhang (Member, IEEE) was born in Nantong, China, in 1981. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2004, 2006, and 2010, respectively.

In 2010, he joined the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, where he is currently an Assistant Professor. He developed a graduate-level course on radar signal analysis and processing. He is also a Member of the Key Laboratory of Radar Imaging (Nanjing University of Aeronautics and Astronautics), Ministry of Education. His current research interests include cognitive radar and adaptive radar waveform design.

Dr. Zhang received the Second Class Prize of the State Scientific and Technological Progress Award in 2019 for radar systems development as a Member of the Technical Team.

Gong Zhang (Member, IEEE) received the Ph.D. degree in electronic engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2002.

From 1990 to 1998, he was a Member of the Technical Staff of the No. 724 Institute China Shipbuilding Industry Corporation, Nanjing. Since 1998, he has been with the College of Electronic and Information Engineering, NUAA, where he is currently a Professor. His research interests include signal processing for communications, radar, and sensor processing; statistical and array signal processing; applications of linear algebra and optimization methods in signal processing and communications; estimation and detection theory; sampling theory; and cooperative and cognitive systems.

Dr. Zhang is a Member of the Committee of Electromagnetic Information of the Chinese Society of Astronautics. He is on the Editorial Board of *Journal of Radars*. He is a Senior Member of the Chinese Institute of Electronics.

Di Wu was born in Anyang, China, in 1982. He received the B.S. and Ph.D. degrees in information engineering and communication and information systems from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2005 and 2011, respectively.

In 2011, he joined the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, where he worked with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, and is currently a Professor. From 2018 to 2019, he was a Visiting Scholar with the Department of Informatics, University of Leicester, Leicester, U.K. His current research interests include radar forward-looking imaging algorithms, space-time adaptive processing, and synthetic aperture radar ground moving target indication.
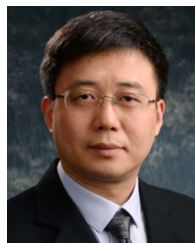
Daiyin Zhu (Member, IEEE) was born in Wuxi, China, in 1974. He received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, in 1996, and the M.S. and Ph.D. degrees in electronics from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, in 1998 and 2002, respectively.

From 1998 to 1999, he was a Guest Scientist with the Institute of Radio Frequency Technology, German Aerospace Center, Cologne, Germany, where he worked in the field of synthetic aperture radar (SAR) interferometry. In 1998, he joined the Department of Electronic Engineering, NUAA, where he is currently a Professor. He has developed algorithms for several operational airborne SAR systems. His current research interests include radar imaging algorithms, SAR ground moving target indication, SAR/inverse SAR autofocus techniques, and SAR interferometry.

Xudong Wang received the B.S., M.S., and Ph.D. degrees in communication engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2001, 2004, and 2008, respectively.

In 2004, he joined the College of Electronic and Information Engineering and the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, NUAA, where he is currently an Associate Professor. His research interests include radar signal processing and field-programmable gate array hardware implementation.