Dynamic Estimation of Spin Satellites Based on the Feature Translation Between Optical and Radar Images

Yejian Zhou, Weifeng Li, Man Zhang, Lei Zhang, Wenan Zhang

Abstract—Nowadays, dynamic estimation of on-orbit satellites is significantly important in space situation awareness applications. Most existing methods are based on high-resolution remote sensing technologies from the ground, which can hardly meet the monitoring demand for mega-constellations due to the limitation of observation scope. In order to achieve the automatic dynamic estimation of satellites in a certain region, a novel approach is proposed based on the synchronous imaging cooperation between spaceborne optical and ground radar devices. Unlike optical imagery, part of the target components is incomplete in inverse synthetic aperture radar (ISAR) images due to its imaging mechanism. Therefore, the radar-to-optical feature translation network is also proposed to support feature extraction among ISAR images. In this way, the extracted shape feature is used to estimate the target instantaneous attitude and spin parameters in steps. Simulation experiments of two typical satellites confirm the robustness and advantages of the proposed method.

Index Terms—Dynamic state estimation, cross-domain feature translation, spin satellites, spaceborne observation.

I. INTRODUCTION

AS the frequent launch of mega-constellations have increased the number of satellites in space orbit, dynamic estimation of these targets plays an important role in various space situation awareness (SSA) applications [1], [2]. The major mega-constellations populated areas are the Low Earth Orbit (LEO) region. It becomes a significant challenge to achieve the efficient monitoring of massive satellites' status with the limited observation period [3], [4]. Especially some of them lose contact with the ground, and their gravity-gradient torque changes when their batter exhaust and attitude control systems (ACS) fail. As a result, this kind of satellite tumbles and spins for a long time, threatening nearby active satellites [5], [6]. Therefore, it becomes a hot topic to analyze target dynamic parameters from the exterior measurement.

Most of the relevant works involve ground-based sensors for the dynamic estimation of on-orbit satellites [7]–[22]. For example, laser ranging (LR) technique is adopted to estimate

Yejian Zhou, WeiFeng Li, and Wenan Zhang are with College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310023, P.R.China.

Man Zhang is with School of Electronic and Communication Engineering, Guangzhou University, Guangzhou, 510006, P.R.China;

Lei Zhang is with School of Electronics and Communication Engineering, Sun Yat-Sen University (Shenzhen Campus), Shenzhen, 518100, P.R.China.

the target spin parameters by matching one-dimensional range residual features in the existing observation dataset [7]–[9]. Organized by International Laser Ranging Service (ILRS), a global monitoring activity was implemented to track the Envisat from 2013 to 2015 [7], [8]. Similarly, some researchers proposed to analyze the target RCS sequence measured by the ground-based radar for the target attitude estimation [10], [11]. This sort of data-driven strategy works rely on long-term historical data, and is limited by observation conditions. Another typical method is based on interpreting the high-resolution two-dimensional observation techniques, such as the ground inverse synthetic aperture radar (ISAR) imaging [12]–[18]. For example, the researchers from European Space Agency (ESA) have used the Tracking and Imaging Radar (TIRA) system to track the Envisat for many years. According to the obtained ISAR image, the motion parameters including its attitude and spin period are determined by feature matching [12]–[14]. Besides, the target shape feature is extracted from several ISAR images, and then used to calculate its dynamic parameters in some optimizations [15]–[17]. There are also some explosive methods proposed on high-resolution images [19]–[22]. With the deep network trained by the simulation dataset, the implicit mapping between the target state and optical image is learned to determine the target attitude in a given image [19]. However, the applicability of single-sensor images to the dynamic estimation of spin satellites is not trivial. It needs long-term spaceborne observation or the collaborative imaging of several ground ISAR devices, which hardly can meet the efficiency demand of mega-constellation observation.

On the other hand, the joint optical-and-radar observation has shown significant improvements in the image analysis field [23]–[26]. Due to its special imaging mechanism, the target's relative motion can be directly related to its Doppler feature in the ISAR image. If the additional shape feature is extracted from the synchronous spaceborne optical image, the target three-dimensional attitude and spin parameters can be calculated in two steps. In order to achieve the dynamic estimation of in-orbit satellites in an automated way, the deep learning technique is introduced. A feature translation network is proposed based on a cycle adversarial strategy to support the precious extraction of target shape features from ISAR images with optical images. After that, the same UNet-based feature extractor is adopted to obtain target projection lengths from synchronous optical and ISAR images. Then, these features are substituted into two optimizations in steps. Solved by the

particle swarm optimization (PSO) algorithm, target attitude and spin parameters are estimated to reflect its operation state in the orbit.

Compared with the existing methods, the proposed algorithm has some innovations.

- We present, to our knowledge, the first application of the imaging cooperation between the spaceborne camera and ground radar for the dynamic estimation of satellites. It makes up for the angular limitation of the ground-based observation and also makes full use of the advantages of these two imaging devices. Without the limitation of lighting and atmosphere conditions, high-resolution images can be generated by the spaceborne optical camera. For fixed region monitoring, ground-based ISAR is able to ensure robust sequential imaging for pass-by satellites. Therefore, we believe it is suitable to meet the increasing demands of the security maintenance of meta-constellations.

- Another major contribution is designing the feature translation network to integrate the strengths of deep networks in optical imaging processing. By learning the target optical feature in an adversarial way, the generated images retain the shape feature of ISAR samples and also is filled with pixel-level optical features. In this way, it is easier to get fine feature extraction even part of the target components is incomplete in the ISAR image due to its imaging mechanism. It is beneficial to achieve the automatic dynamic estimation of on-orbit satellites.

The remainder of this article is organized as follows. Section II introduces the fundamental of the target dynamic estimation. Combined with the dynamic model of the on-orbit satellite, the explicit connection between target dynamic parameters and projection features in the images is derived. Section III describes the proposed feature translation network, which is based on the cycle adversarial learning framework. Based on it, Section IV describes the proposed dynamic estimation method. Section V demonstrates the feasibility of the proposed algorithm in three experiments. Some crucial conclusions of this work are drawn in Section VI.

II. DYNAMIC ESTIMATION OF SATELLITES FROM SYNCHRONIZED IMAGES

A. Synchronized Imaging Model

As shown in Fig. 1, two ground-based ISAR radars and one spaceborne optical satellite are adopted to make up the synchronized imaging observation for dynamic estimation of satellites within their visible scope. Unlike optical imaging, ISAR imagery is obtained by analyzing the target range and Doppler characteristic of in the echo during the coherent processing interval (CPI). Therefore, the imaging moment of the optical sensor and the central moment of the radar CPI are required to be synchronized. Besides, in order to describe target spin motion in a common coordinate system, the Cartesian coordinate system is established on the observed target. As depicted in Fig. 1, the W axis points to the earth's core, the V axis is consistent with the target trajectory motion, and the N axis is determined by the right-hand rule.
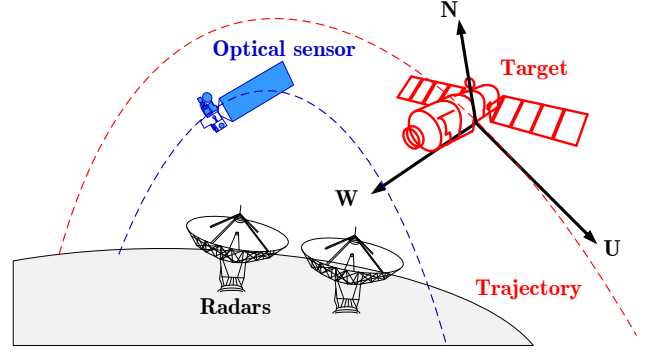

Fig. 1: The imaging geometry of the radar-and-optical joint observation.

Under this coordinate system, imaging geometry vectors are defined to describe the optical imaging plane and ISAR imaging plane, shown in Fig.2. Each vector is determined by elevation angle q and azimuth angle f, as in the following equation.

$$\vec{\mathbf{k}} = (\cos\theta\sin\phi, \cos\theta\cos\phi, \sin\theta)^{\mathrm{T}} \qquad (1)$$

where elevation angle θ is the intersection angle between the vector and the UON plane, and azimuth angle ϕ is the intersection angle between N axis and the projection of the vector in the UON plane.
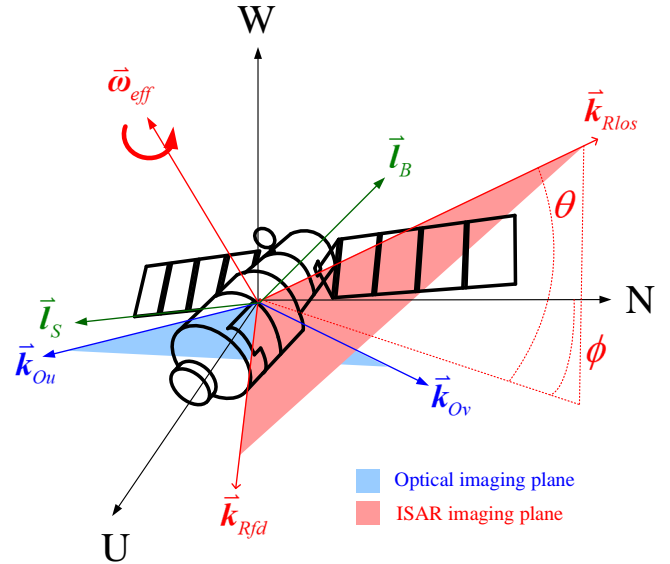

Fig. 2: The spin motion model of spin spacecraft.

Among these vectors, $\vec{\mathbf{k}}_{Ou}$ and $\vec{\mathbf{k}}_{Ov}$ denotes the vertical and horizontal dimensions of the optical imaging plane, respectively. $\vec{\mathbf{k}}_{Rlos}$ refers to the radar line of sight (LOS) at the imaging central moment, which is consistent with the range axis of the ISAR image. These vectors can be directly obtained according to the target observation trajectory data in the tracking systems of space-borne cameras and ground-based ISAR devices.

However, the other dimension of ISAR image, the Doppler axis $\vec{\mathbf{k}}_{Rfd}$, is hard to be calculate when target spins. As we know, target image information in Doppler dimension reflects its effective rotation between the target and the radar. It contains the radar LOS change and its spin motion as follows.

$$\vec{\mathbf{w}}_{eff} = \vec{\mathbf{w}}_{los} + \vec{\mathbf{w}}_{rot} \tag{2}$$

where the radar LOS changing vector $\vec{\mathbf{w}}_{los}$ can be calculated with target LOS sequences, but the target spin vector $\vec{\mathbf{w}}_{rot}$ is unknown and needs to be solved. With $\vec{\mathbf{w}}_{los}$ and $\vec{\mathbf{w}}_{rot}$, $\vec{\mathbf{k}}_{Rfd}$ can be determined according to right hand principle.

From relevant research [5], [6], [17], it is assumed that the target rotates around a certain rotation shaft with a constant speed during the ISAR coherence accumulation interval (CPI). In this work, it is represented as the following equation.

$$\vec{\mathbf{w}}_{rot} = (\cos\theta_{rot}\sin\phi_{rot}, \cos\theta_{rot}\cos\phi_{rot}, \sin\theta_{rot})^{\mathrm{T}}\omega_{rot} \tag{3}$$

where θ_{rot} and ϕ_{rot} are used to describe the direction of the rotation shaft, and ω_{rot} is the spin speed. Therefore, the Doppler axis $\vec{\mathbf{k}}_{Rfd}$ is related to these spin parameters in each ISAR image.

For most satellites, there are more than two typical structural features that can be observed in the 2D image, like the spacecraft body and the solar wing. For each structure, two attitude parameters are adopted to describe its instantaneous attitude vector as follows.

$$\vec{\mathbf{l}} = (\cos\alpha\sin\beta, \cos\alpha\cos\beta, \sin\alpha)^{\mathrm{T}} \tag{4}$$

where α is the intersection angle between the instantaneous attitude vector and the UON plane, and β is the intersection angle between the N axis and the projection of the instantaneous attitude vector in the UON plane.

Then, the target projection feature in the known axes of these synchronous images can be expressed as

$$u = L\vec{\mathbf{k}}_{Ou} \bullet \vec{\mathbf{l}} \tag{5}$$

$$v = L\vec{\mathbf{k}}_{Ov} \bullet \vec{\mathbf{l}} \tag{6}$$

$$r_1 = L\vec{\mathbf{k}}_{Rlos1} \bullet \vec{\mathbf{l}} \tag{7}$$

$$r_2 = L\vec{\mathbf{k}}_{Rlos2} \bullet \vec{\mathbf{l}} \tag{8}$$

where L refers to the 3D length of the typical structure, and $\bullet$ denotes the inner product.

B. Dynamic Parameters estimation

Once these projection lengths are extracted from the optical and ISAR images, both the instantaneous attitude vector and length of each typical structure can be estimated in the following optimization independently.

$$\min_{\alpha,\beta,L} (r_1 - \hat{r}_1)^2 + (r_2 - \hat{r}_2)^2 + (u - \hat{u})^2 + (v - \hat{v})^2 \tag{9}$$

where $\hat{\ast}$ denotes the measured value from the images.

The optimization (9) can be solved by various stochastic intelligent methods, like the PSO algorithm. After solving for the optimal $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{L})^T$, the attitude vector of the structure can be calculated according to Eq.(10).

$$\vec{\widetilde{\mathbf{l}}} = \widetilde{L}(\cos\widetilde{\alpha}\sin\widetilde{\beta}, \cos\widetilde{\alpha}\sin\widetilde{\beta}, \sin\widetilde{\alpha})^T \tag{10}$$

When there are more than two attitude vectors obtained in the former stage, the target spin parameters can be estimated according to the target projection lengths in the Doppler dimension of ISAR imagery.

Similar to imaging projection in the range axis, the target Doppler feature is described by the inner product between the typical structure's attitude vector and the equivalent rotation of the target, as follows.

$$d_{j,1} = \omega_{eff1}\left(\vec{\mathbf{w}}_{los1}, \vec{\mathbf{w}}_{rot}\right)\vec{\mathbf{k}}_{Rfd1}\left(\vec{\mathbf{w}}_{los1}, \vec{\mathbf{w}}_{rot}\right) \bullet \vec{\widetilde{\mathbf{l}}}_j \tag{11}$$

$$d_{j,2} = \omega_{eff1}\left(\vec{\mathbf{w}}_{los2}, \vec{\mathbf{w}}_{rot}\right)\vec{\mathbf{k}}_{Rfd2}\left(\vec{\mathbf{w}}_{los2}, \vec{\mathbf{w}}_{rot}\right) \bullet \vec{\widetilde{\mathbf{l}}}_j \tag{12}$$

where $\vec{\mathbf{w}}_{eff}$ is the equivalent rotation speed between the target and radar during the current CPI, which can be represented by the spin according to Eq.(2). Besides, the Doppler axis is always perpendicular to the range axis in each ISAR image.

$$\vec{\mathbf{k}}_{Rfd1}\left(\vec{\mathbf{w}}_{los1}, \vec{\mathbf{w}}_{rot}\right) \bullet \vec{\mathbf{k}}_{Rlos1} = 0 \tag{13}$$

$$\vec{\mathbf{k}}_{Rfd2}\left(\vec{\mathbf{w}}_{los2}, \vec{\mathbf{w}}_{rot}\right) \bullet \vec{\mathbf{k}}_{Rlos2} = 0 \tag{14}$$

Therefore, the spin parameters, ω_{eff}, θ_{rot}, ϕ_{rot} are solved in the below optimization.

$$\begin{aligned} \min_{\omega_{eff},\theta_{rot},\phi_{rot}} \sum_j &\left(d_{j1} - \hat{d}_{j1}\right)^2 + \left(d_{j2} - \hat{d}_{j2}\right) \\ &+ \lambda\vec{\mathbf{k}}_{Rfd1}\left(\vec{\mathbf{w}}_{los1}, \vec{\mathbf{w}}_{rot}\right) \bullet \vec{\mathbf{k}}_{Rlos1} \\ &+ \lambda\vec{\mathbf{k}}_{Rfd2}\left(\vec{\mathbf{w}}_{los2}, \vec{\mathbf{w}}_{rot}\right) \bullet \vec{\mathbf{k}}_{Rlos2} \end{aligned} \tag{15}$$

where λ is the confidence factor.

C. Feature extraction from optical and ISAR images

According to the previous derivation, we can find that the precious extraction of these projection lengths from the optical and ISAR images becomes the core of estimating target dynamic parameters in this work. However, from the public ground-based ISAR images of satellites [27], the shape of the target shape feature is often incomplete due to the scattering sensitivity to the radar LOS. It is difficult to achieve robust feature extraction further decreasing the accuracy of target state estimation. On the contrary, these shape features can be easily observed from the optical image. When the target shows the same shape characteristics in the ISAR and optical images, these images can be paired to assist the feature extraction in the ISAR image.

In order to achieve this task, the optical features can be directly used as supervision during the annotation of the ISAR image. However, it is hard to be understood why those missing

components are still marked in the ISAR image without pixel-level supplementary information. Therefore, the target pixel-level features between the optical domain and ISAR domain should be related first. Then, the annotation label of the optical image can be shared to improve the feature extraction of the target component in the ISAR image. That is to say, our goal is training a feature-fusion network to map target shape features from the original optical-domain samples $X_s = X_i|_{i=1}^N$ and ISAR-domain samples $Y_S = y_i|_i^M = 1$.

III. FEATURE CROSS-DOMAIN TRANSLATION BETWEEN OPTICAL IMAGES TO ISAR IMAGES

A. Related work

Reviewing the related works on image-to-image translation, Generative Adversarial Network (GAN) is considered to be one of the most effective methods in this field [28]–[32]. Through the antagonistic training of generating network (Generator) and discriminating network (Discriminator), the generated imagery becomes close to the image in the target domain. In this way, the style translation is realized among different domains. In order to improve the training efficiency, Deep Convolution Generative Adversarial Network (DCGAN) is further proposed by optimizing the structure of the original GAN model and the activation functions of generator and discriminator [28]. More recent approaches adopt additional label information as part of the input to train the parameters of the generator and discriminator, such as Information Maximizing GAN (InfoGAN) and auxiliary classifier GAN (ACGAN) [29]–[32]. Unlike the random generation strategy, the output of the generator is continuously close to the real sample with the difference decreasing or auxiliary classification principle. Another approach is introducing the cycle loss to ensure that the changing of generated sample keeps consistent during the cycle training. As a domain adaptation model, CycleGAN can be trained without pair-matched images.

On the other hand, deep feature fusion is the core of the image cross-domain translation task. A sort of classical method is based on the fine-tuning of the original domain model [33], [34]. Some few-shot learning strategy is used to make the re-trained network simultaneously adaptive to both the original and new tasks. For example, part of the original samples is encoded and superimposed with noise in the framework of Data Augmentation GAN (DAGAN) [35]. After stitching with the real ones, these new samples help the decoder to generate cross-domain images. Similarly, other works achieve sample augmentation by interpolating existing samples with certain weights [36], [37]. Furthermore, an additional fusion step is introduced in the fusing-and-filling GAN (F2GAN) [38]. Different from the existing linear interpolation, image features are extracted and fused in the deep perceptual domain after convolution processing. That is why the generated image preserves the semantic information of conditional images and the detailed texture features of the original image. These works inspire us to think about the feature translation between the optical image and ISAR image of satellites.

B. Radar-to-Optical Feature Translation Network

The proposed radar-to-optical feature translation (R2OFT) network is depicted in Fig.3, which is composed of two branches (i.e., the Optical-ISAR and ISAR-Optical translation branches). Each branch consists of a feature generator and discriminator, which aim to optimize the feature translation process via adversarial gaming. Specifically, the cycle adversarial training strategy is employed to ensure consistency when the input is translated to the target domain and subsequently returned to the original domain. This strategy can effectively preserve key features throughout the unpaired Optical-ISAR translation process. Besides, a Non-local Attentional Fusion (NAF) module is introduced to utilize the local detail and global context from multi-layer features to improve the quality of the synthesized images. After the optimization, the optical generator can be used to assist in the extraction of targeted key points from ISAR images by synthesizing corresponding optical images.

1) Feature Generator: The feature generator aims to preserve the domain-specific feature while performing the translation procedure from the optical domain to the ISAR domain, and inversely. Specifically, two autoencoder-based networks initially take the optical and ISAR images and reconstruct them under the L1-norm constraint to prevent domain feature distortion. Following Ref. [38], we introduce a non-local attention fusion block at the end of each decoder block to prevent information loss during the up-sampling stage and encourage multi-level feature integration. The non-local attention takes a few conditional optical images $\mathcal{X}^O = \{x_k^O|_{k=1}^K\}$ and ISAR images $\mathcal{X}^I = \{x_k^I|_{k=1}^K\}$ with corresponding random interpolation coefficients $\mathbf{a}^O = [a_1^O, \cdots, a_k^O]$ and $\mathbf{a}^I = [a_1^I, \cdots, a_k^I]$ to generate multi-level feature concentrated images for the subsequent translation procedure. The operation of the non-local attention fusion mechanism is demonstrated as follows:

$$\mathcal{L}_{NAF} = \sum_{k=1}^K a_i^O||x_i^O - \hat{x}^O||_1 + a_i^I||x_i^I - \hat{x}^I||_1 \quad (16)$$

where $\hat{x}^I$ and $\hat{x}^O$ represent the generated multi-level fused ISAR and optical images respectively. The K represents the number of conditional images and k denotes the k-th images. Then, to keep the information consistency of the translated images, the l1-norm constraint optimization function of the feature generator can be written as:

$$\mathcal{L}_{Norm} = ||\hat{x}^O - G_{Op}(\hat{x}^I)||_1 \\ + ||\hat{x}^I - G_{ISAR}(\hat{x}^O)||_1 \quad (17)$$

where G_{ISAR} and G_{Op} represents the ISAR generator and Optical generator respectively.

Furthermore, to ensure that the translated images embed the style of the target domain, the generator is designed to produce images such that the domain discriminator cannot distinguish them. This is achieved through the following adversarial optimization:

$$\mathcal{L}_{Adv}^G = - \mathbb{E}_{G_{ISAR}(\hat{x}^O)}[D_{ISAR}(G_{ISAR}(\hat{x}^O))] \\ - \mathbb{E}_{G_{Op}(\hat{x}^I)}[D_{Op}(G_{Op}(\hat{x}^I))] \quad (18)$$
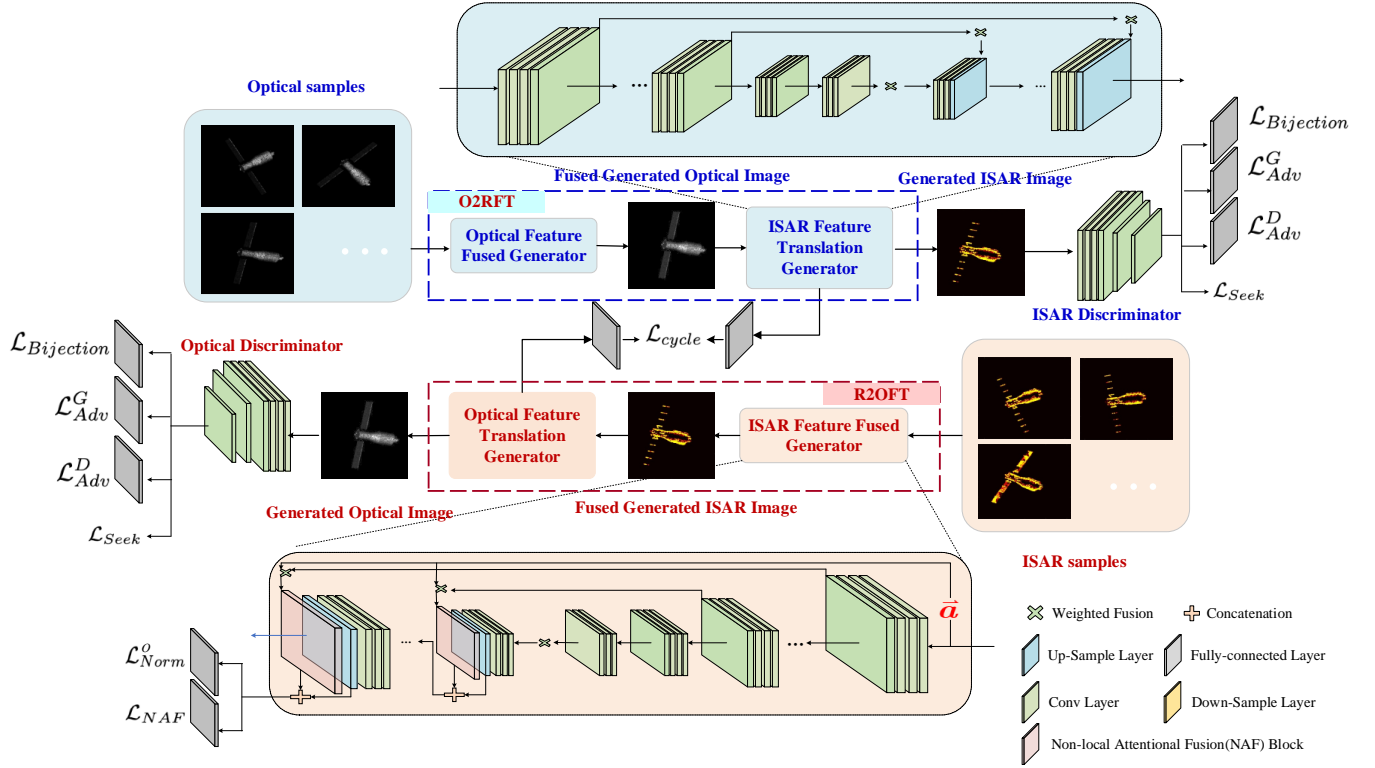
Fig. 3: The architecture of the proposed feature translation framework.

where D_{ISAR} and D_{Op} represents the ISAR discriminator and Optical discriminator respectively.

Finally, the total optimization function of the generator could be written as:

$$\mathcal{L}_G = \mathcal{L}_{NAF} + \mathcal{L}_{Norm} + \mathcal{L}_{Adv}^G \tag{19}$$

2) Feature Discriminator: The feature discriminator aims to distinguish the fake generated images so as to encourage the generator to produce more realistic images. The adversarial optimization function of the discriminator could be written as:

$$\begin{aligned}
\mathcal{L}_{Adv}^D =& \mathbb{E}_{G_{ISAR}(\hat{x}^O)}[max(0, 1 + D_{ISAR}(G_{ISAR}(\hat{x}^O)))] \\
& + \mathbb{E}_{G_{Op}(\hat{x}^I)}[max(0, 1 + D_{Op}(G_{Op}(\hat{x}^I)))] \\
& + \mathbb{E}_{\hat{x}^I}[max(0, 1 - D_{ISAR}(\hat{x}^I))] \\
& + \mathbb{E}_{\hat{x}^O}[max(0, 1 - D_{Op}(\hat{x}^O))]
\end{aligned} \tag{20}$$

Furthermore, to simulate the generator's capacity to synthesize more diverse images (i.e., diverse interpolation coefficients $\mathbf{a}$), we introduce two strategies to prevent the mode collapse of the generator (i.e., generate identical or similar images). Specifically, we introduce the mode seeking loss, as described in [51], to maintain the positive proportional relationship between the differences in latent representation and the synthesized images. This strategy enhances the discrepancies of generated images with only minor discrepancies of its corresponding interpolation coefficients in the latent space. The optimization function could be written as:

$$\mathcal{L}_{Seek} = \frac{\left\| \hat{D}\left(x_{\mathbf{a}1}\right) - \hat{D}\left(x_{\mathbf{a}_2}\right) \right\|_1}{\left\| \mathbf{a}_1 - \mathbf{a}_2 \right\|_1} \tag{21}$$

where $\hat{D}$ denotes the discriminator D that removed the classification layer for feature extraction usage. The $\hat{x}_{\mathbf{a}_1}$ and $\hat{x}_{\mathbf{a}_2}$ represent the images generated with the corresponding interpolation coefficients $\mathbf{a}_1$ and $\mathbf{a}_2$ respectively.

The second strategy is to leverage the interpolation regression loss [38] for bijective relationship maintenance. Specifically, to maintain the one-to-one mapping between the interpolation coefficients and generated images, we utilize a fully connected layer $E()$ to predict the similarity score s_k between the features of conditional images $\hat{D}(x_k)$ and generated images $\hat{D}(\hat{x})$.

$$\begin{aligned}
s &= [s_1, \ldots, s_K] \\
&= E([D(x_k), D(\hat{x})])_{k=1}^K
\end{aligned} \tag{22}$$

Then, we apply the softmax function to the similarity score and obtain the predicted interpolation coefficient, which is enforced to match the ground truth $\mathbf{a}$.

$$\mathcal{L}_{Bijection} = ||softmax(s) - \mathbf{a}||_2 \tag{23}$$

Finally, the total optimization function of the discriminator could be written as:

$$\mathcal{L}_D = \mathcal{L}_{Adv}^D - \lambda_m \mathcal{L}_{Seek} + \lambda_a \mathcal{L}_{Bijection} \tag{24}$$

where λ_m and $\lambda_{\mathbf{a}}$ are trade-off parameters.

3) Cycle adversarial strategy: Unlike traditional adversarial training methods, our approach incorporates a cycle strategy to learn the mapping function between paired optical and ISAR samples [33]. Specifically, each optical image is converted into an ISAR sample using our specially designed ISAR Generator. This ISAR sample can then be reverted back

into the original optical image using our Optical Generator. Similarly, an ISAR image can also be reverted using the Optical Generator and ISAR Generator in sequence. This cycle consistency is enforced by minimizing the following function:

$$
\begin{aligned}
\mathcal{L}_{cycle} = \; & \mathbb{E}_{\hat{x}^O}[||G_{Op}(G_{ISAR}(\hat{x}^O)) - \hat{x}^O|| \\
& + \mathbb{E}_{\hat{x}^I}[||G_{ISAR}(G_{Op}(\hat{x}^I)) - \hat{x}^I||
\end{aligned}
\tag{25}
$$

In all, the overall optimization function for both the generator and discriminator could be written as:

$$
\mathcal{L}_{total} = \mathcal{L}_G + \mathcal{L}_D + \mathcal{L}_{cycle}
\tag{26}
$$

IV. THE PROPOSED AUTOMAGICAL DYNAMIC ESTIMATION METHOD
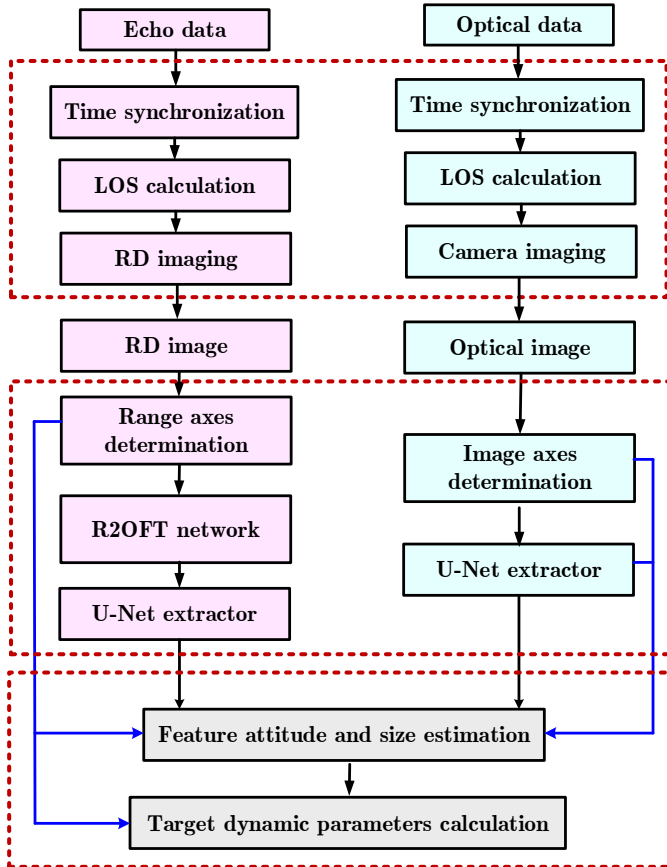

Fig. 4: The flowchart of target dynamic estimation.

The dynamic estimation of spin satellites is comprised of three components: image preprocessing, feature extraction, target attitude estimation, and spin parameter estimation. The flow of the proposed method is depicted in Fig.4, and the detailed steps are below.

- Step 1: The time synchronization between optical and radar sensors is performed according to the tracking data from their observation systems.
- Step 2: The synchronized ISAR and optical images are generated with the high-resolution imaging technique, respectively.
- Step 3: The range axis of the ISAR image is calculated with the radar LOS parameters, while the vertical and

TABLE I: The main parameters of the ISAR image dataset

Size of a single image	256×256
Radar image range resolution	0.05 m
Signal Center frequency	10 GHz
Pulse repetition frequency	100 Hz
Optical image resolutions	0.05 m $\times 0.05 m$

horizontal axes of the optical image are determined by external parameter calibration.
- Step 4A: A five-layer U-Net is used to extract projection lengths of typical structural features in the optical image [39]. This feature extractor is trained by the optical samples of the target in advance.
- Step 4B: By contrast, the shape feature of the ISAR image is strengthened by the proposed R2OFT network to map the target shape feature in the optical domain. Then, the same U-Net extractor is used to extract target projection lengths from the generated image.
- Step 5: With the observation vectors of three imaging axes, the obtained range, vertical and horizontal lengths are substituted into the optimization (9). Using the PSO algorithm, the instantaneous attitude vectors and 3D lengths of typical structural features can be solved.
- Step 6:With estimated target instantaneous attitude parameters, the obtained Doppler, vertical and horizontal lengths are substituted into optimization (15). Also using the PSO algorithm, the target spin parameters can be solved.

V. EXPERIMENTS
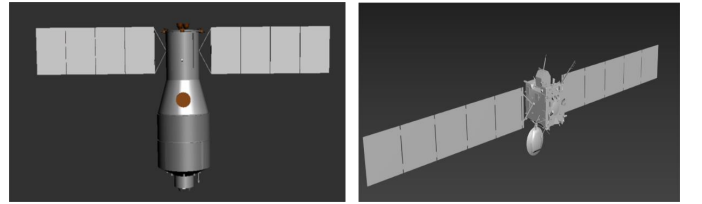
A. Datasets and Implementation Details

Fig. 5: The 3D models of satellites. (left: TG-I; right: Rosetta.)

As shown in Fig.5, two typical satellites, TianGong-I(TG-I) and Rosetta, are adopted as the target satellite to build two simulation datasets in this section, respectively. The optical images are generated by the ray-tracing algorithm [40], [41], while the ISAR images are generated by the fast physical optics (FPO) algorithm [42], [43]. The signal-to-noise ratio (SNR) of each image is set to 10 dB, and other simulation parameters are given in Table 1. Each dataset consists of 280 optical images and 280 ISAR images, which almost covers the all-aspect observation scope. Among them, three-quarters of the samples are used for training, and the rest become the test set.

In the feature translation experiment, the detailed structure of the proposed R2OFT network is listed in Table.II, and the Adam optimizer is used during the training. The trade-off

TABLE II: The structure parameters of the R2OFT network

Stages	Detailed Structure
The generator of R2OFT network	
Stage 1	[Conv.64@3*3, ReLU]*3 Maxpool 3*3
Stage 2	[Conv.128@3*3, ReLU]*3 Maxpool 3*3
Stage 3	[Conv.256@3*3, ReLU]*3 Maxpool 3*3
Stage 4	[Conv.512@3*3, ReLU]*3 Maxpool 3*3
Stage 5	[Conv.1024@3*3, ReLU]*3 Maxpool 3*3
Stage 6	[Conv.1024@3*3, ReLU]*3 Maxpool 3*3
Stage 7	[Conv.512@3*3, ReLU]*3 Maxpool 3*3
Stage 8	[Conv.256@3*3, ReLU]*3 Maxpool 3*3
Stage 9	[Conv.128@3*3, ReLU]*3 Maxpool 3*3
Stage 10	[Conv.1@3*3, ReLU]*3 Maxpool 3*3
The discriminator of R2OFT network	
Stage 1	[Conv.1@7*7, ReLU]
Stage 2	[Conv.64@1*1, ReLU, Conv.64@3*3, ReLU, Conv.256@1*1, ReLU]*3 Maxpool 3*3
Stage 3	[Conv.128@1*1, ReLU, Conv.128@3*3, ReLU, Conv.512@1*1, ReLU]*4
Stage 4	[Conv.256@1*1, ReLU, Conv.256@3*3, ReLU, Conv.1024@1*1, ReLU]*6
Stage 5	[Conv.512@1*1, ReLU, Conv.512@3*3, ReLU, Conv.1024@1*1, ReLU]*3

TABLE III: The structure parameters of the Unet-based feature extractor

Stages	Detailed Structure
Stage 1	[Conv.570@3*3, ReLU, Conv.568@3*3, ReLU] MaxPool 2*2
Stage 2	[Conv.282@3*3, ReLU, Conv.280@3*3, ReLU] MaxPool 2*2
Stage 3	[Conv.138@3*3, ReLU, Conv.136@3*3, ReLU] MaxPool 2*2
Stage 4	[Conv.66@3*3, ReLU, Conv.64@3*3, ReLU] MaxPool 2*2
Stage 5A	[Conv.32@3*3, ReLU]
Stage 5B	[Conv.28@3*3, ReLU] Up-Conv 2*2
Stage 6	[Conv.54@3*3, ReLU, Conv.52@3*3, ReLU] Up-Conv 2*2
Stage 7	[[Conv.102@3*3, ReLU, Conv.100@3*3, ReLU] Up-Conv 2*2
Stage 8	[Conv.198@3*3, ReLU,Conv.196@3*3, ReLU] Up-Conv 2*2
Stage 9	[Conv.390@3*3, ReLU, Conv.390@3*3, ReLU] Conv.388@1*1

TABLE IV: The similar evaluation results of image cross-domain generation by different methods

Methods	TG1			Rosetta		
	FID	IS	LPIPS	FID	IS	LPIPS
Pix2Pix	150.76	5.14	0.09	220.78	2.49	0.08
CycleGAN	210.11	3,97	0.06	191.09	4.57	0.05
BiCycleGAN	139.41	6.87	0.15	178.34	4.43	0.10
NiceGAN	190.12	3.93	0.08	146.98	5.00	0.10
Ours	**106.55**	**8.35**	**0.20**	**110.96**	**7.51**	**0.19**

parameters λ_m and λ_a are set to be 3, the learning rate is set to be 0.003, and the number of training epochs is 200.

In the feature extraction experiment, the detailed structure of the feature extraction is listed in Table.III. The label of the image feature is manually annotated before network training. During the training, the Adam optimizer is used, and the mean square error (MSE) is adopted to calculate the loss function. The learning rate was set as 0.003, and the number of training epochs is 50.

B. Feature translation experiment

To quantitatively evaluate the cross-domain feature quality of the proposed R2OFT network's generator, we use three evaluation metrics: Inception Scores (IS), Fréchet Inception Distance (FID), and Learned Perceptual Image Patch Similarity (LPIPS). The number of IS positively correlates with the visual quality of the generated image [44]. The FID measures the feature similarity of the two images [45]. The pre-trained Inception-V3 model is adopted as the feature extractor, and its last average pooling layer is removed. LPIPS can reflect the average distance of the shape feature among the generated samples [46]. And the performance of the R2OFT network is also compared with four existing methods, Pix2Pix, CycleGAN, BiCycleGAN and NiceGAN [47]–[50].

The visual comparison of image cross-domain generation is depicted in Fig.6 and Fig.7. Compared with the ground truth, the cross-domain generation of NiceGAN and CycleGan perform well in the case of Rosetta. But the tail of the TG-I's body is generated with the wrong shape by these two

methods. The major shape features of Rosetta and TG-I can be found in the cross-domain images generated by Pix2Pix and BiCycleGAN, but some detailed features are not accurate, such as the contour profile of the target body. By contrast, the image generated by our method is more realistic. Because the shape feature of several images in different observation LOS conditions is coherently integrated during the training of each epoch. The numerical comparison of the evaluation metrics is listed in Table. IV. The image generated by the proposed R2OFT network gets the highest score of IS and LPIPS, and the lowest of FID, which demonstrates its advantages in the cross-domain generation of satellite images.

C. Feature extraction experiment

In order to illustrate the advantage of the proposed R2OFT network on ISAR image interpretation, the target shape feature is extracted by two approaches. In the first way, the key points are extracted from the ISAR image by the U-net extractor directly after training. In the other way, these ISAR samples are used to generate cross-domain images by the R2OFT

Fig. 6: Visual comparison of Image generation on TG-I samples. (The first row: Inputs; The second row: the ground truth; The third row: cross-domain images generated by Pix2Pix; The forth row: cross-domain images generated by CycleGAN; The fifth row: cross-domain images generated by BiCycleGAN; The sixth row: cross-domain images generated by NiceGAN; The last row: cross-domain images generated by the proposed R2OFT network;)

network, and then the key points are extracted from these images by the U-net extractor, which is trained according to the optical samples. For the TG-I image, 10 key points are annotated in the image, including eight corner points of two solar wings, and the head and tail of the target body. For the Rosetta image, four corner points of the long solar wing and four corner points of the passenger cabin are annotated as labels.

Compared with the ground truth, the shift of extracted points' position is calculated. The average bias of the extraction of the first method is 12.6 and 10.8 pixels in the TG-I dataset and Rosetta dataset, respectively. By contrast, that of the second method reduces to 2.3 and 2.1 pixels in these two datasets, respectively. Due to the scattering mechanism of ISAR imaging, part of the target feature might be weakened or lost, and can not be observed in some images (depicted with white circles in the sub-figures of Fig. 8). As the target shape feature of the optical sample is more complete in the same observation view, the ISAR image is used to generate

the cross-domain image. Therefore, the feature extraction of the proposed method has a better performance in both of these two datasets.

Besides, to investigate the robustness of the proposed method, the extraction experiment of key points is repeated with the same samples under different SNR conditions, from -10 dB to 10 dB. As shown in Figs.9 and 10, the extracted key points (marked in red) are still close to the ground truth (marked in green) in most images even though their SNR has decreased to -5 dB. It confirms the robustness of the proposed method.

D. Target dynamic estimation experiment

To exemplify the proposed target dynamic estimation method, a continuous observation scene is built to obtain the dynamic parameters of a Starlink satellite (53918U). Two ground-based ISAR devices and an adjacent Starlink satellite (53914U)in the same orbit are used, and the geometrical parameters are listed in Table.V. The target satellite spins
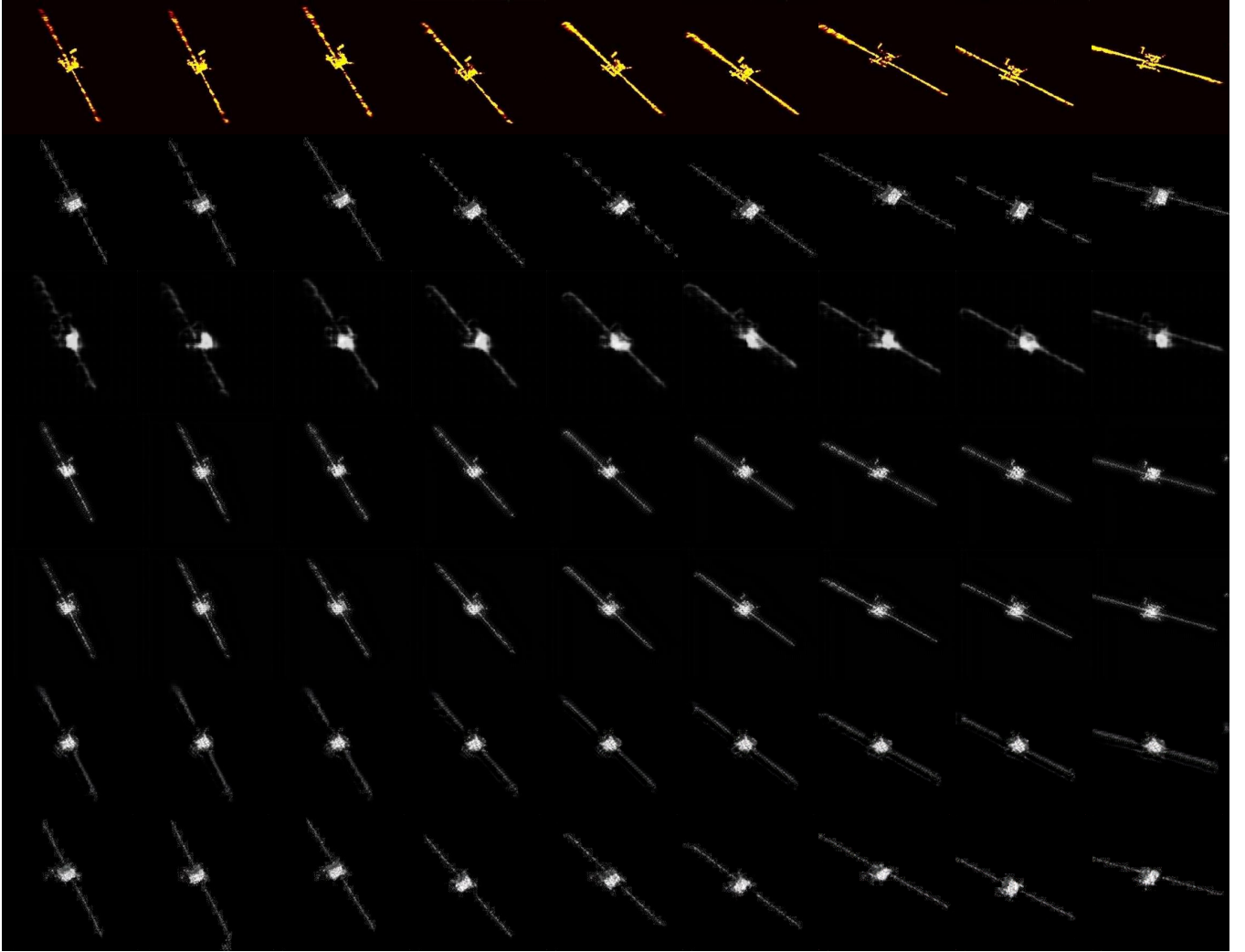
Fig. 7: Visual comparison of Image generation on Rosetta samples. (The first row: Inputs; The second row: the ground truth; The third row: cross-domain images generated by Pix2Pix; The fourth row:cross-domain images generated by CycleGAN; The fifth row: cross-domain images generated by BiCycleGAN; The sixth row: cross-domain images generated by NiceGAN; The last row: images generated by the proposed R2OFT network;)

around a certain shaft at a constant speed, and its instantaneous attitude and spin speed will be measured by the proposed method at ten moments in the visual scope of this observation system. With the target shape feature extracted by the proposed network, target instantaneous attitude parameters are obtained by solving the optimization Eq.(9). Then, these estimated results are substituted into Eq.(15) to determine target spin parameters, including spin speed and shaft. The numerical results are compared with the true values in Table. VI. It confirms the feasibility of the proposed method, and its precision is able to achieve the state-of-art result of existing methods which rely on the accurate extraction of target shape features in ISAR images [17], [23].

Overall, we believe the proposed method is more suitable for the state monitoring of pass-over satellites in a certain LEO region. At the current stage, different space optical satellites are chosen from the mega-constellation to build the observation system with fixed ground-based radars. However, it can be extended to use geostationary satellites for a fixed

TABLE V: The geometrical parameters of the observers and target satellite in the observation scene

ECEF coordinates of ground ISAR stations	
Station 1	(34.15 N, 108.56 E, 88 m)
Station 2	(34.37 N, 112.26 E, 0 m)
TLE parameters of satellites	
Observer	1 53918U 22119AJ 22316.91667824 -.00091872 00000-0 -65537-3 0 9996
	2 53918 53.2169 10.7314 0003590 5.9101 36.2310 15.73258859 8144
Target	1 53914U 22119AE 22316.91667824 -.00078552 00000-0 -55904-3 0 9999
	2 53914 53.2169 10.7338 0003878 1.8002 39.1983 15.73247937 8143

system if high-resolution optical imaging can be ensured in this condition.

CONCLUSIONS

In this article, an approach is proposed based on spaceborne optical and ground ISAR images to estimate the dynamic
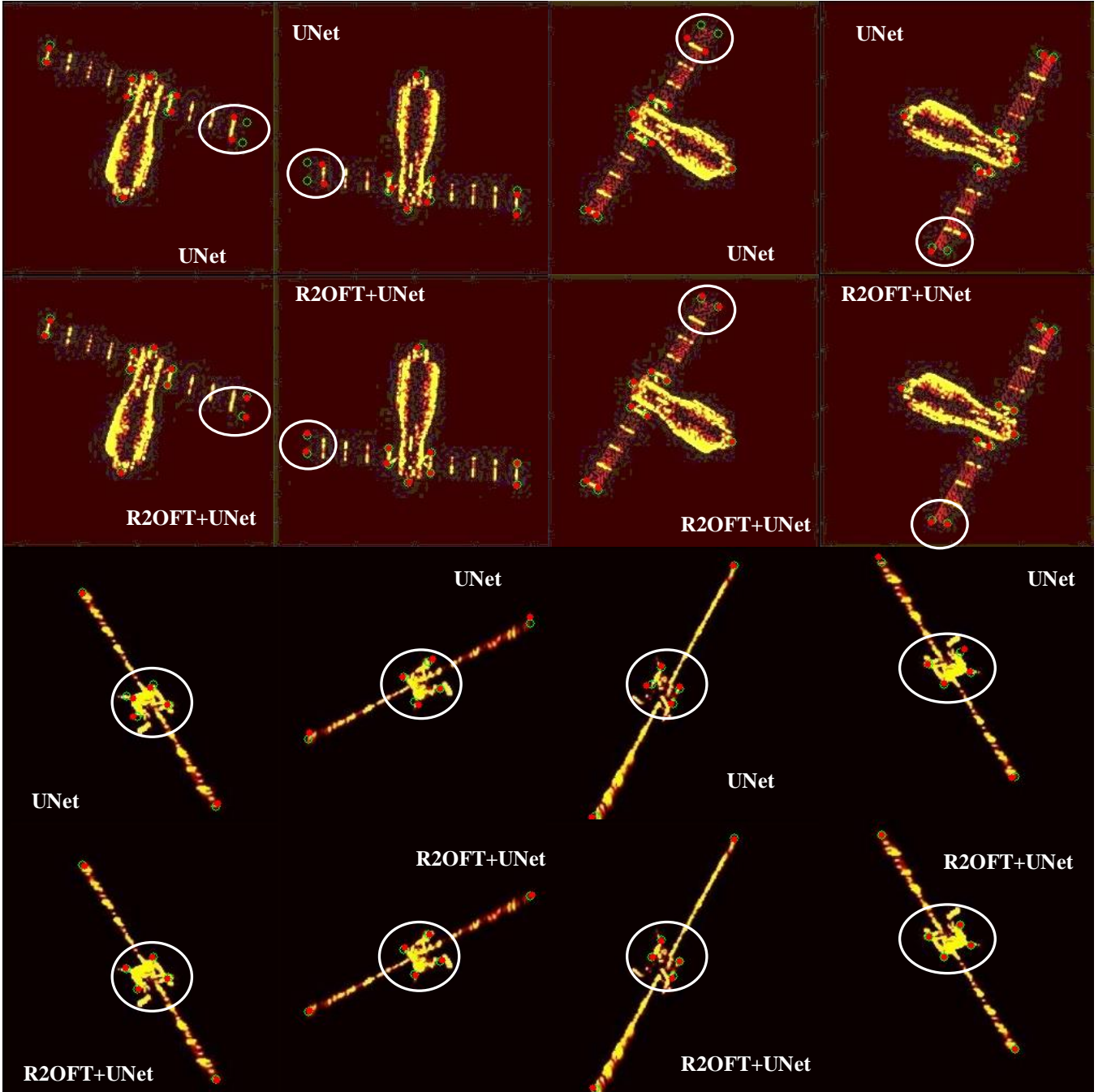
Fig. 8: The ablation results of R2OFT network.

parameters of spin spacecraft. The explicit expression is derived to bridge the synchronous imagery and target dynamic parameters. In order to achieve precious feature extraction, the R2OFT network is also designed in a cycle adversarial learning way. Then, the target projection feature extracted by the U-net is used to solve the target instantaneous attitude and spin parameters with the PSO algorithm. The experimental results have confirmed the effectiveness of the proposed algorithm. Compared with the existing methods, the proposed method breaks up the limitation of the observation view. We believe this article provides an effective approach to accomplishing target dynamic estimation in mega-constellations applications.

In the future, we plan to investigate a novel processing framework for the dynamic estimation of multiple targets in a certain space region.

REFERENCES

[1] Boley A., and Byers M. "Satellite mega-constellations create risks in Low Earth Orbit.", the atmosphere and on Earth. Sci Rep 11, 10642, 2021.
[2] McDowell J., "The low earth orbit satellite population and impacts of the SpaceX Starlink constellation.", The Astrophysical Journal Letters 892.2, 2020.
[3] Le S., Gehly S., Carter B., and et al., "Space debris collision probability analysis for proposed global broadband constellations.", Acta Astronautica 151 : 445-455, 2018.
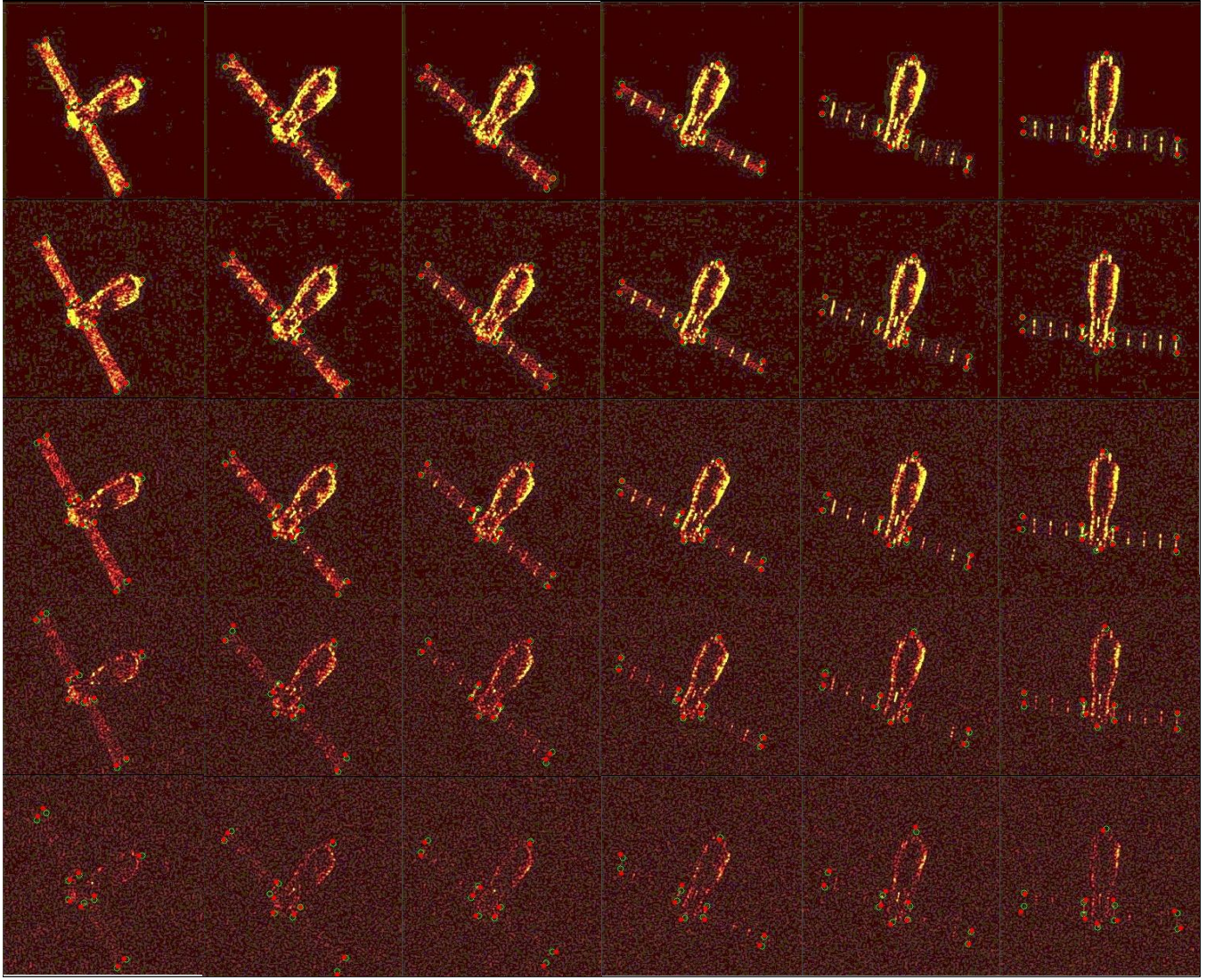
Fig. 9: Extraction results of TG-I in different noise conditions.(The first row: SNR = 10 dB; The second row: SNR = 5 dB; The third row: SNR = 0 dB; The fourth row: SNR = -5 dB; The fifth row: SNR = -10 dB.)

TABLE VI: The dynamic estimation results

Observation time	Estimation error of target body (degrees)	Estimation error of solar wings (degrees)	Estimation error of the spin shaft (degrees)	Estimation error of the spin speed (10^{-2}rad/s)
Moment 1	1.24	0.39	1.88	0.31
Moment 2	1.08	1.29	2.16	0.12
Moment 3	0.36	0.92	2.30	0.23
Moment 4	1.37	1.25	1.63	0.30
Moment 5	0.85	1.12	1.72	0.22
Moment 6	0.83	0.33	1.50	0.14
Moment 7	0.97	1.17	1.23	0.08
Moment 8	1.49	1.65	0.80	0.15
Moment 9	0.38	0.95	0.47	0.06
Moment 10	1.12	1.38	0.38	0.11

[4] Boley C., and Michael B., "Satellite mega-constellations create risks in Low Earth Orbit, the atmosphere and on Earth.", Scientific Reports 11.1: 10642, 2021.

[5] Kucharski D., Kirchner G., Koidl F., and et al., "Attitude and spin period of space debris Envisat measured by satellite laser ranging.", IEEE Transactions on Geoscience and Remote Sensing, 52(12): 7651-7657, 2014.

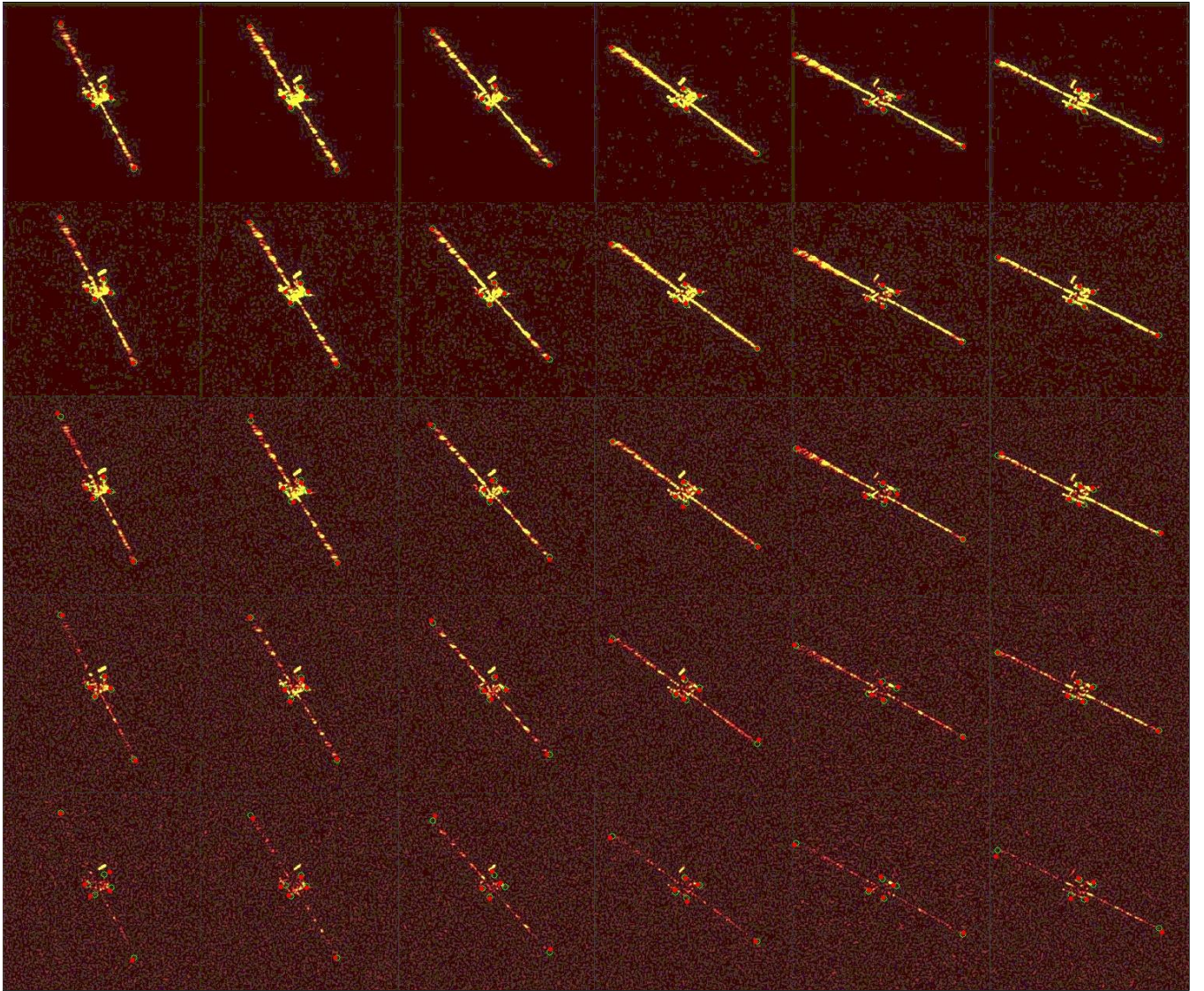[6] Pittet J., Šilha J., and Schildknecht T., "Spin motion determination of

Fig. 10: Extraction results of Rosetta in different noise conditions.(The first row: SNR = 10 dB; The second row: SNR = 5 dB; The third row: SNR = 0 dB; The fourth row: SNR = -5 dB; The fifth row: SNR = -10 dB.)

the Envisat satellite through laser ranging measurements from a single pass measured by a single station.", Advances in space research 61(4): 1121-1131, 2018.

[7] Gómez N., and Walker S., "Earth's gravity gradient and eddy currents effects on the rotational dynamics of space debris objects: Envisat case study.", Advances in Space Research 56(3): 494-508, 2015.

[8] Lin H., and Zhao C., "An estimation of Envisat's rotational state accounting for the precession of its rotational axis caused by gravity-gradient torque.", Advances in Space Research, 61(1):182-188, 2018.

[9] Peng J., Xu W., Yan L., and et al., "A pose measurement method of a space noncooperative target based on maximum outer contour recognition.", IEEE Transactions on Aerospace and Electronic Systems, 56(1):512–526, 2020.

[10] Zhong W., Wang J., Ji W., and et al., "The attitude estimation of three-axis stabilized satellites using hybrid particle swarm optimization combined with radar cross section precise prediction.", Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 230(4):713-725, 2016.

[11] Lyu J., Zhong W., Liu H., and et al., "Novel approach to determine spinning satellites' attitude by RCS measurements.", Journal of Aerospace Engineering, 34(4):04021023, 2021.

[12] Sommer S., Rosebrock J., Cerutti-Maori D., and L. Leushacke, "Temporal analysis of Envisat's rotational motion.", in Proc. Eur. Conf. Space Debris, 2017.

[13] Lemmens S., and Krag H., "Sensitivity of automated attitude determination form ISAR radar mappings," in Proc. Adv. Maui Opt. Space Surveillance. Technol. Conf., pp. 1–12, 2013.

[14] Rosebrock J., "Absolute attitude from monostatic radar measurements of rotating Objects," IEEE Transactions on Geoscience and Remote Sensing, 49(10):3737–3744, 2011.

[15] Zhou Y., Xie P., Li C., and et al., "Automatic Dynamic Estimation of On-orbit Satellites through Spaceborne ISAR Imaging." IEEE Transactions on Radar Systems, 2023.

[16] Suwa K., Wakayama T., and Iwamoto M., "Three-dimensional target geometry and target motion estimation method using multistatic ISAR movies and its performance," IEEE Transactions on Geoscience and Remote Sensing, 49(6):2361–2373, 2011.

[17] Zhou Y., Zhang L., Cao Y., "Dynamic estimation of spin spacecraft based on multiple-station ISAR images.", IEEE Transactions on Geoscience and Remote Sensing, 58(4): 2977-2989, 2019.

[18] Kou P., Liu Y., Zhong W., and et al. "Axial attitude estimation of spacecraft in orbit based on ISAR image sequence.", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 7246-7258, 2021.

[19] Sharma S., D'Amico S., "Neural network-based pose estimation for noncooperative spacecraft rendezvous.", IEEE Transactions on Aerospace and Electronic Systems, 56(6): 4638-4658, 2020.

[20] Duan J., Ma Y., Zhang L., and et al., "Abnormal Dynamic Recognition

of Space Targets From ISAR Image Sequences With SSAE-LSTM Network.", IEEE Transactions on Geoscience and Remote Sensing, 2023.

[21] Guthrie B., Kim M., Urrutxua H., and et al.," Image-based attitude determination of co-orbiting satellites using deep learning technologies.", Aerospace Science and Technology, 120: 107232, 2022.

[22] Qiao S., Zhang H., Meng G., and et al., "Deep-Learning-Based Satellite Relative Pose Estimation Using Monocular Optical Images and 3D Structural Information.", Aerospace, 9(12): 768, 2022.

[23] Zhou Y., Zhang L., Cao Y., and et al., "Optical-and-radar image fusion for dynamic estimation of spin satellites.", IEEE Transactions on Image Processing, 29: 2963-2976, 2019.

[24] Zheng S., Shi W., Liu J., and et al., "Multisource image fusion method using support value transform," IEEE Transactions on Image Processing, 16(7):1831–1839, 2007.

[25] Xu Y., Ma P., Ng E. and et al., "Fusion of WorldView-2 Stereo and Multitemporal TerraSAR-X Images for Building Height Extraction in Urban Areas.", IEEE Geoscience and Remote Sensing Letters, 12(8):1795-1799, 2015.

[26] Xu, W. and Wang, G., "A domain gap aware generative adversarial network for multi-domain image translation.", IEEE Transactions on Image Processing, 31:72-84, 2021.

[27] Monitoring the re-entry of the chinese space station Tiangong-1 with TIRA, FGAN Lab, Berlin, German, Mar. 2018.

[28] Goodfellow I., Pouget-Abadie J., Mirza M., and et al. "Generative Adversarial Nets.", Communications of the ACM, 63(11):139-144, 2020.

[29] Radford A., Metz L., and Chintala S., "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". arXiv e-prints, 2015.

[30] Mirza M., and Osindero S., "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784, 2014.

[31] Chen X., Duan Y., Houthooft R., and et al., "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets.", Advances in neural information processing systems 29, 2016.

[32] Odena A., Olah C., Shlens J., "Conditional image synthesis with auxiliary classifier gans.", International conference on machine learning. PMLR: 2642-2651, 2017.

[33] Liang W., Liu Z., and Liu C., "Dawson: A domain adaptive few shot generation framework.", arXiv preprint arXiv:2001.00576, 2020.

[34] Cloutre L., and Demers M.,"FIGR: Few-shot Image Generation with Reptile.", 10.48550/arXiv.1901.02199. 2019.

[35] Antoniou A., Storkey A., and Edwards H., "Data augmentation generative adversarial networks.", arXiv preprint arXiv:1711.04340, 2017.

[36] Hong Y., Niu L., Zhang J., and et al., "Matchinggan: Matching-based few-shot image generation.", 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020: 1-6.

[37] Gu Z., Li W., Huo J., and et al., "LoFGAN: Fusing Local Representations for Few-shot Image Generation.", 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 8443-8451, 2021.

[38] Hong Y., Niu L., Zhang J., and et al., "F2gan: Fusing-and-filling gan for few-shot image generation.", Proceedings of the 28th ACM international conference on multimedia. 2535-2543, 2020.

[39] Ronneberger, O., Fischer, P. and Brox, T., "Unet: Convolutional networks for biomedical image segmentation". In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, Part III 18pp234-241, 2015.

[40] Auer S., Hinz S., and Bamler R., "Ray-tracing simulation techniques for understanding high-resolution SAR images," IEEE Transactions on Geoscience and Remote Sensing, 48(3):1445–1456, 2010.

[41] Kulpa K.S., Samczyński P., Malanowski M., and et al., "An advanced SAR simulator of three-dimensional structures combining geometrical optics and full-wave electromagnetic methods.", IEEE Transactions on Geoscience and Remote Sensing, 52(1):776-784, 2014.

[42] Wang F., Eibert T., and Jin Q., "Simulation of ISAR imaging for a space target and reconstruction under sparse sampling via compressed sensing.", IEEE Transactions on Geoscience and Remote Sensing, 53(6):3432–3441, 2015.

[43] Boag A., "A fast physical optics (FPO) algorithm for high-frequency scattering." IEEE Transactions on Antennas and Propagation, 52(1)197–204, 2004.

[44] Xu Q., Huang G., Yuan Y., and et al., "An empirical study on evaluation metrics of generative adversarial networks.", arXiv preprint arXiv:1806.07755, 2018.

[45] Heusel M., Ramsauer H., Unterthiner T., and et al., "Gans trained by a two time-scale update rule converge to a local nash equilibrium.", Advances in neural information processing systems, 30. 2017.

[46] Zhang, R., Isola, P., Efros, A., and et al., "The unreasonable effectiveness of deep features as a perceptual metric.", In Proceedings of the IEEE conference on computer vision and pattern recognition, 586-595, 2018.

[47] Isola P., Zhu J., Zhou T. and et al., "Image-to-image translation with conditional adversarial networks.", In Proceedings of the IEEE conference on computer vision and pattern recognition, 1125-1134, 2017.

[48] Zhu J., Park T., Isola P., and et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks.", Proceedings of the IEEE international conference on computer vision. Venice, 2223-2232, 2017.

[49] Zhu J., Zhang R., Pathak D., and et al., "Toward multimodal image-to-image translation.", Advances in neural information processing systems, 30, 2017.

[50] Chen R., Huang W., Huang B., and et al., "Reusing discriminators for encoding: Towards unsupervised image-to-image translation.", Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle. WA, USA, 8168-8177, 2020.

[51] Mao Q., Hsin-Ying L., and et al., "Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 1429-1437, 2019.