Φ-GAN: Physics-Inspired GAN for Generating SAR Images Under Limited Data

Supplementary Material

1. Details of the physical model

1.1. PSC model

According to the geometric diffraction theorem, the ideal point scattering center (PSC) model can be articulated as a linear combination of N independent PSC points, as demonstrated in Equation 1:

$$\boldsymbol{E}(\boldsymbol{f},\boldsymbol{\phi}) = \sum_{i=1}^{N} A_i \cdot \exp(-j\frac{4\pi \boldsymbol{f}}{c}(x_i\cos\boldsymbol{\phi} + y_i\sin\boldsymbol{\phi})). \quad (1)$$

Here, $\boldsymbol{E}$ denotes the measurement in the frequency domain. A_i and $\exp(\cdot)$ represent the response intensity and phase information of the radar echo associated with the ith PSC, respectively. $j = \sqrt{-1}$ denotes the imaginary unit, while c signifies the velocity of propagation (i.e., light velocity). $\boldsymbol{f}$ and $\boldsymbol{\phi}$ denote the radar echo frequency and aspect angle, respectively. x_i and y_i denote the coordinates of each PSC in the range and azimuth directions, respectively.

The sparsity of radar echo signals necessitates the application of sparse signal theory for the analysis and extraction of PSC. The PSC model derived from this can be reformulated as follows:

$$\widetilde{\mathbf{r}} = \widetilde{\boldsymbol{\Psi}}(\boldsymbol{x},\boldsymbol{y})\mathbf{o} \quad (2)$$

Here, $\widetilde{r}$ denotes the vectorization of E, while $\mathbf{o}$ signifies the sparse coefficient vector corresponding to $\{A_i\}_{i=1}^{N}$ in Eq. 1. The symbol $\widetilde{\boldsymbol{\Psi}}$ represents a dictionary that encompasses the positional data of PSCs within the signal domain, corresponding to the exponent term in Equation 1. The dictionary can be formulated in columns as follows:

$$\begin{aligned} &\widetilde{\boldsymbol{\Psi}}(\boldsymbol{x},\boldsymbol{y}) = \widetilde{\Psi}_{:,1}, \widetilde{\Psi}_{:,2}, \cdots, \widetilde{\Psi}_{:,HW}, \\ &\widetilde{\Psi}_{:,HW} = vec(\exp(-j\frac{4\pi \boldsymbol{f}}{c}(x_H\cos\boldsymbol{\phi} + y_W\sin\boldsymbol{\phi})), \end{aligned} \quad (3)$$

where $\boldsymbol{f}$ and $\boldsymbol{\phi}$ are vectors evenly sampled P and Q times within their respective ranges. Here, $\boldsymbol{f}$ is within the interval $(f_c - B/2, f_c + B/2)$, and $\boldsymbol{\phi}$ is within the interval $(-\boldsymbol{\phi}_{syn}/2, \boldsymbol{\phi}_{syn}/2)$, where B denotes the bandwidth, f_c denotes the center frequency, and $\boldsymbol{\phi}_{syn}$ denotes the synthetic angle. $\boldsymbol{x}$ and $\boldsymbol{y}$ indicate vectors sampled H and W times in the range and azimuth directions at specified resolution intervals. Previous study [2] has demonstrated that executing an inverse Fourier transform on each column of $\widetilde{\boldsymbol{\Psi}}(\boldsymbol{x},\boldsymbol{y})$ facilitates the conversion of the dictionary from the frequency domain to the image domain, hence enabling the solution of the PSC in the image domain. The image domain representation of Equation 2 can be explained as follows:

$$\mathbf{r} = \boldsymbol{\Psi}\mathbf{o} \quad (4)$$

where $\boldsymbol{\Psi}$ represents the dictionary in the image domain and $\mathbf{r}$ denotes the input complex-valued image following vectorization. Consequently, $\mathbf{o}$ can be derived by resolving the optimization problem defined in Equation 5.

$$\hat{\mathbf{o}} = \arg\min_{\mathbf{o}} ||\boldsymbol{\Psi}\mathbf{o} - \mathbf{r}||_2 + \lambda||\mathbf{o}||_1. \quad (5)$$

1.2. Hyperparameters setting for the dictionary

Given that the targets often fall within the central 80×80 region, we construct the dictionary specifically for this area, setting the sampling numbers P, Q, H, and W all to 80, resulting in a fixed dictionary with dimensions of 6400×6400. More details of the hyperparameters relevant to the construction of the dictionary are presented in Table 1.

Hyperparameter	Value	Hyperparameter	Value
f_c	$9.599e9$	B	$0.591e9$
ϕ_{syn}	2.3°	c	$3e8$
x_{interval}	0.202	y_{interval}	0.202
H	80	W	80
P	80	Q	80

Table 1. The hyperparameters setting for the dictionary

1.3. Complex-valued SAR image processing

In the PSC model, complex-valued electromagnetic waves are processed to characterize the scattering behavior of targets. However, our method generates only amplitude data representing the strength of target echoes, without phase information. To facilitate end-to-end training, we append synthetic phase values, sampled from a uniform distribution over $[0, 2\pi)$, to the generator's output $\tilde{u}$, thereby synthesizing a complex-valued SAR image. In general, SAR phase data are uniformly distributed over $[0, 2\pi)$ in scenarios where no dominant coherence or geometric structure exists. This aligns with speckle theory and established statistical principles [1]. Consequently, this approach approximates SAR data with random, uniformly distributed phases, suitable for applications where phase coherence is not a primary concern. We also keep the amplitude information of the output of PSC reconstruction $\mathcal{F}_{\text{phy}}$ to obtain s and $\tilde{s}$, and the loss functions are calculated without phase information.

2. Datasets

We utilise three different SAR image datasets in our experiments: MSTAR, SAR-Airplane and OpenSARShip, in a variety of dataset sizes and target types. Some of the training samples are shown in Fig.1.

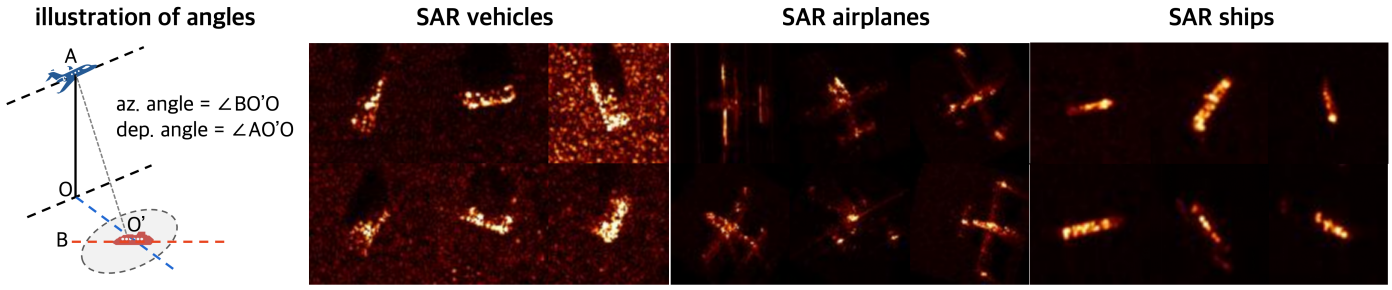

Figure 1. The illustration of azimuth angle and depression angle of SAR imaging geometry. Training samples of the three datasets used in the experiment.

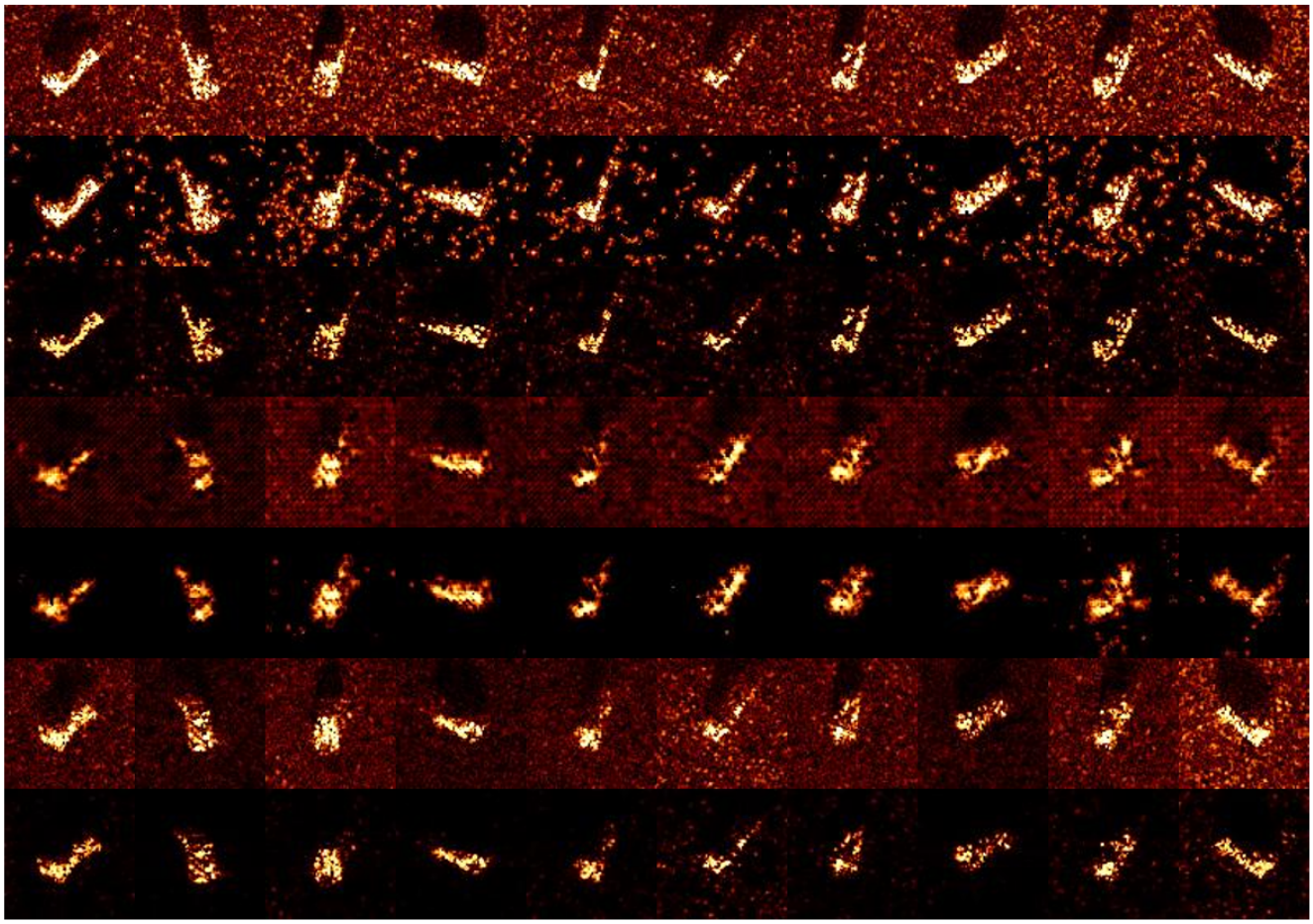

Figure 2. Visual comparisons between scattering point extraction method, namely Harris-Laplacian corner detection and PSC. From top to bottom: real images, HLC results of real images, PSC reconstruction results of real images, generated images by Φ-GAN(w/ HLC), HLC results of generated image by Φ-GAN(w/ HLC), generated images by Φ-GAN(w/ PSC), PSC reconstruction results of generated images by Φ-GAN(w/ PSC).

- **MSTAR** : The moving and stationary target acquisition and recognition (MSTAR) dataset is the most widely used dataset in the sar target generation task. MSTAR consists of ten-class vehicles collected at different viewing angles, at a resolution of 0.3m×0.3m, taken by an X-band SAR sensor. The azimuth angle ranges from 0° to 360°, with an approximate interval of 1° to 2°. The depression angles are 17° and 15°. The demonstrations of azimuth angle and depression angle can be found in Fig.1. The image size of MSTAR is 128×128. In our experiments, we uniformly

α	FID(↓)	KID(↓)
0.4	673.4226	1.0969
0.6	**132.5429**	**0.0974**
0.8	135.2127	0.1008

Table 2. Discuss α. β and γ fixed as 1,1, respectively.

β	FID(↓)	KID(↓)
1	**132.5429**	**0.0974**
5	925.4651	1.5387
10	1155.4966	2.2148

Table 3. Discuss β. α and γ fixed as 0.6,1, respectively.

γ	FID(↓)	KID(↓)
1	132.5429	0.0974
5	139.2050	0.1057
10	**130.2311**	**0.0861**

Table 4. Discuss γ. α and β fixed as 0.6,1, respectively.

Method	10% MSTAR		5% MSTAR	
	Uncertainty	Accuracy	Uncertainty	Accuracy
ACGAN	0.1961	74.9031±1.1192	0.2599	55.5629±0.6219
+RLC	0.2570	66.5649±1.5895	0.2901	55.1340±2.6723
+DIG	0.2576	66.2598±1.3447	0.3931	53.0391±0.8384
+Ours	**0.0832**	**89.3691**±1.0977	**0.1332**	**78.9361**±1.0541

Table 5. Additional quantitative results comparisons with other methods.

Method	10% MSTAR	5% MSTAR
without aug	75.67%	49.11%
Base	83.01%	70.84%
+RLC	79.63%	70.80%
+DIG	83.02%	65.79%
+Ours	**90.91%**	**81.75%**

Table 6. Few-shot SAR target recognition results.

sample images every 15° azimuth angle intervals from the images with a depression angle of 17° to obtain obtain 10% MSTAR dataset with 237 samples for 10-class. Similarly, the 5% MSTAR dataset with 121 samples for 10-class is acquired with sampling intervals of 30° azimuth angle. The rest samples with 17° depression angles are used for testing and evaluation of image quality.

- **SAR-Airplane**: The SAR-Airplane dataset contains two different types of airplane, namely Kodiak100 and AT-504 taken by a Ku-band SAR sensor in strip imaging mode, at a resolution of 0.05m with a depression angle of 45°. Each type of airplane consists of 72 images, which are resized to a uniform dimension of 80×80 pixels and the corresponding azimuth angle information is 0°, 5°, 10°,..., 355°. In our experiments, we uniformly choose instances of SAR-Airplane in 2 categories every 35° azimuth angle intervals to obtain the 14% training subset with 10 samples each category.
- **OpenSARShip**: The VV channel data of GRD products of OpenSARShip which was obtained by C-band Sentinel-1 satellites at a resolution of 20m×22m are used in our experiment. Considering the large range of ship sizes, only chips which have a size of larger than 80 × 80 pixels are chosen. All the chips are center-cropped to 80 × 80 pixels as inputs to the generator. Meanwhile, as FID and KID require sufficient real data as the reference dataset, samples from two categories including Cargo and Tanker are selected with 3358 and 403 samples, respectively. In our experiments, we only use 24 Cargo samples and 22 Tanker samples for training with uniform intervals of 15° azimuth angle.

  OpenSARShip dataset doesn't provide azimuth angle information, so we have an additional angle processing step. Since the x/y-axis of OpenSARShip images correspond to the range/azimuth direction, respectively, we obtain the azimuth angle according to the following procedure: **1)** Use SAM to segment the ship. **2)** Obtain the rotated bounding box (bbx) of the ship. **3)** Calculate the horizontal contained angle $\theta \in [0, \pi)$ of the rotated bounding box. **4)** Distinguish between the ship bow and stern to determine the final azimuth as θ or $\theta + \pi$.

3. Detailed analyses of Harris-Laplacian Corner (HLC) detection and our method

Harris-Laplacian Corner (HLC) detection method generates scale space by convolving the original image with Gaussian kernel functions of different scale parameters and searches for candidate corner points in each scale space layer. Inspired by the autocorrelation function of signal processing, the Harris Laplace detection algorithm uses the autocorrelation matrix M for corner detection. The eigenvalues of matrix M are the first-order curvature of the autocorrelation function. If both first-order curvature values of a point are large, it is a corner point. The LOG operator (Laplacian of Gaussian) is commonly used to detect local feature corners with scale invariance.

Compared to PSC, the HLC method is sensitive to noise, and when there is significant noise interference, the HLC method may not be able to accurately extract scattering centers. Furthermore, in the training process, the threshold is fixed which may be too high or too low to some certain images, resulting in poor extraction performance while PSC does not have this issue.

Fig.2 shows the visual results of the exacted HLC and PSC of real images and the corresponding generated images. It can be seen that speckle noise still remains in some of the HLC results extracted from the real image which may disturb the learning of targets while the extracted PSC only contains vital information of the targets. Additionally, compared to

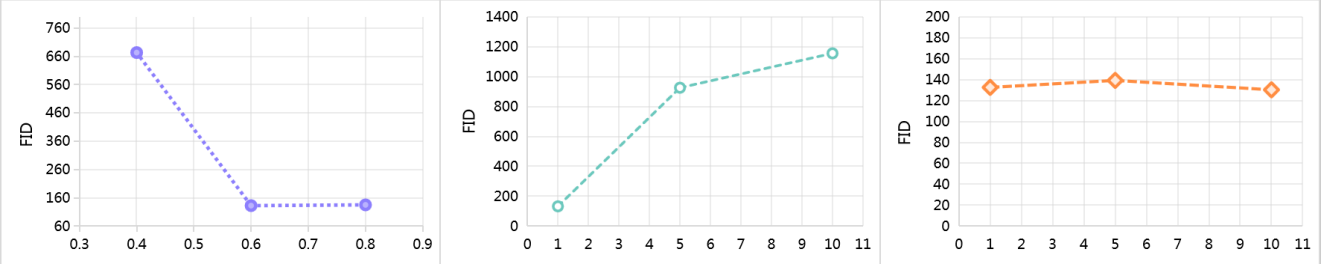

Figure 3. From left to right: FID for our methods with varying α, β and γ,respectively.

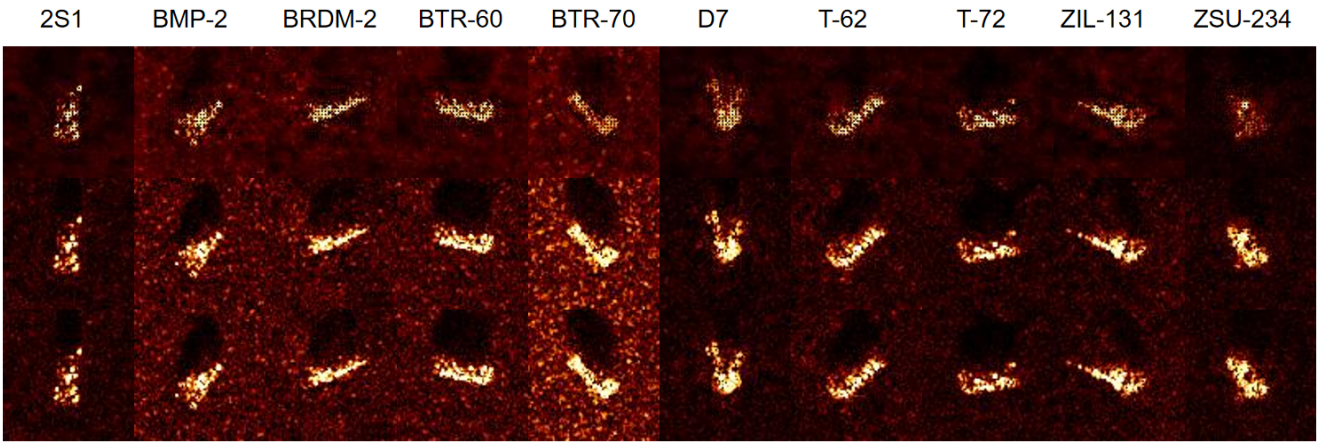

Figure 4. From top to bottom: visual comparisons between generated images with $\alpha = 0.4$, $\alpha = 0.6$ and $\alpha = 0.8$, respectively with fixed $\beta = 1$ and $\gamma = 1$.

HLC which are points with significant variations in any two perpendicular directions, PSC is more sparse and can offer more important physical information. In that case, it is easier for the network to learn crucial physical features under limited data under the guidance of our proposed method.

4. Hyperparameter ablation studies

We discuss the hyperparameter in Fig.4, 5, 6 and Table 2, 3, 4. The results show that too small α and too big β will impose too strong constrain on the physical parameters, perhaps disregarding other image features, which adversely affects the realism of the generated images. In addition, the model is not sensitive to the value of γ. As can be seen from the Fig.4, 5, too strong constrain makes the network fail to capture the distribution of the background, thus reducing the realism of the generated images.

5. Additional evaluation results by X-Fake

Table 5 compares the Prob-Eva accuracy and uncertainty of our method with those of the RLC and DIG methods in terms of utility. The results show that the RLC and DIG methods are not well-suited for this task, as both their Prob-Eva accuracy and uncertainty are worse than the baseline. In contrast, our method significantly outperforms the baseline, demonstrating notable improvements in both Prob-Eva accuracy and uncertainty. The results of Fig.7 further prove the advantage of maintaining the physical consistency of our method. Fig.7 (c) and Fig.7 (h) illustrate the differences between the generated SAR images and their corresponding counterfactual explanations, with missing details marked in red and redundant components highlighted in blue. A comparison with Fig.7 (c), (f), and (h) reveals that the baseline method struggles to clearly generate certain PSC points in the SAR images, as indicated by the red points in Fig.7 (c). In contrast, our method is able to generate these PSC points more accurately and with greater clarity, as evidenced by the improved results in our approach.

6. Evaluation on few-shot SAR target recognition

We further analyze the effectiveness of the proposed Φ-GAN by performing a SAR target recognition task with few-shot

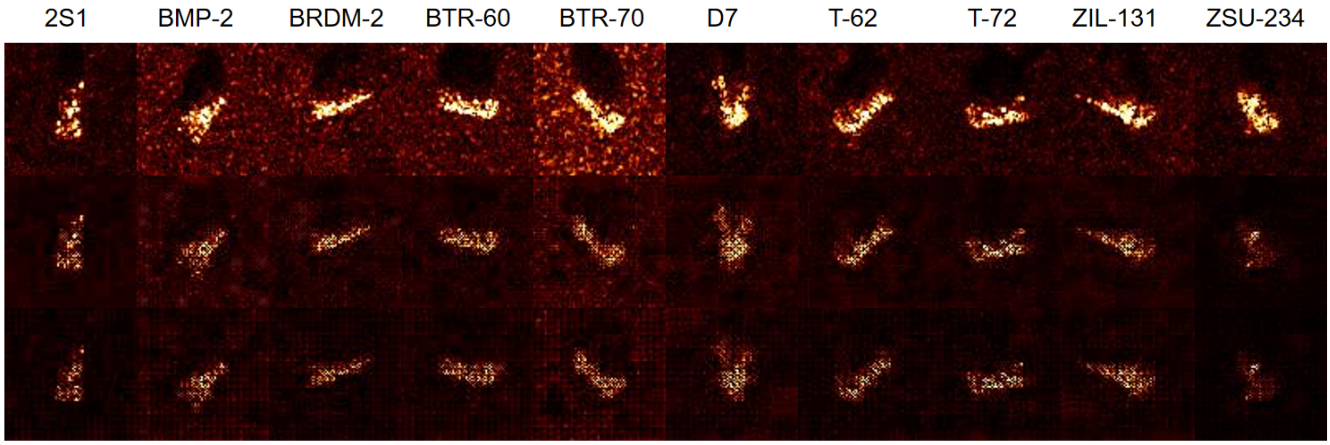

Figure 5. From top to bottom: visual comparisons between generated images with $\beta = 1$, $\beta = 5$ and $\beta = 10$, respectively with fixed $\alpha = 0.6$ and $\gamma = 1$.

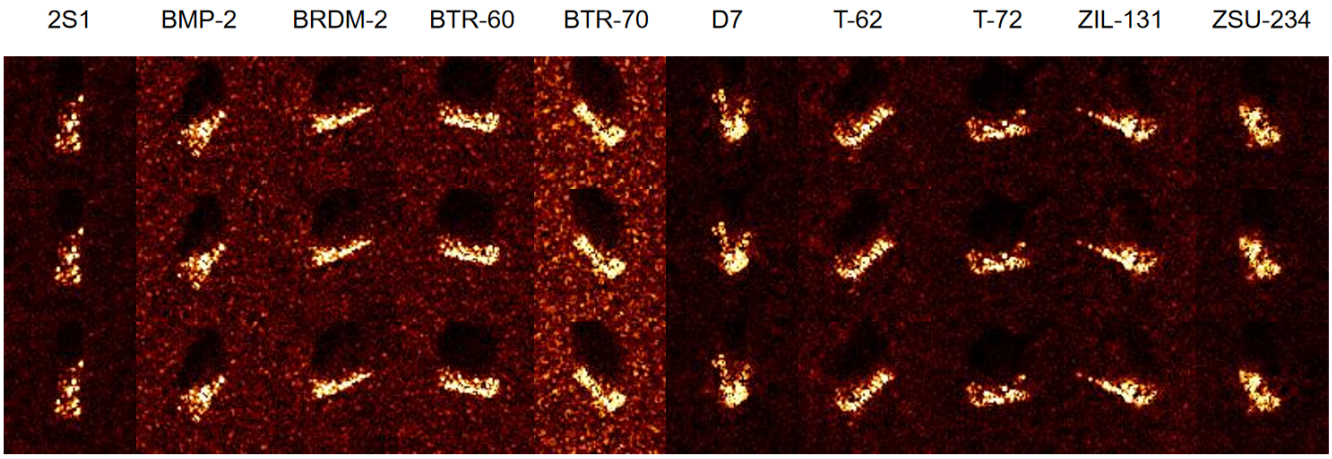

Figure 6. From top to bottom: visual comparisons between generated images with $\gamma = 1$, $\gamma = 5$ and $\gamma = 10$, respectively with fixed $\alpha = 0.6$ and $\beta = 1$.

data. Specifically, different GAN models are utilized to augment the small sample set, and the augmented datasets are used to train multiple SAR target recognition models for evaluation. The recognition results are summarized in Table 6. The results demonstrate that the augmented data generated by our Φ-GAN significantly improves the performance of the recognition model on the 5% MSTAR dataset. This indicates that Φ-GAN effectively enhances learning by enriching the limited training samples with high-quality synthetic data.

7. Limited data training discussion

We compare the generated results under different limited training samples, i.e., 10%, 5%, and 3% MSTAR (8 samples per class). The results are recorded in Table 7. It demonstrates that our method is effective under 5% and 10% MSTAR dataset. However, when the the amount of data further decreases, it is hard for the $\mathcal{D}_{\text{phy}}$ to capture the discriminative features of the scattering points, thus failing to give positive guidance to the generator. Our future work will focus on developing more effective physics-aware regularization methods to address the challenges posed by increasingly rigorous data-scarce scenarios in SAR image generation.

8. Encoding more angle conditions

The definition of azimuth angle (AA) and depression angle (DA) is shown in Fig.1. Azimuth angle is encoded as the

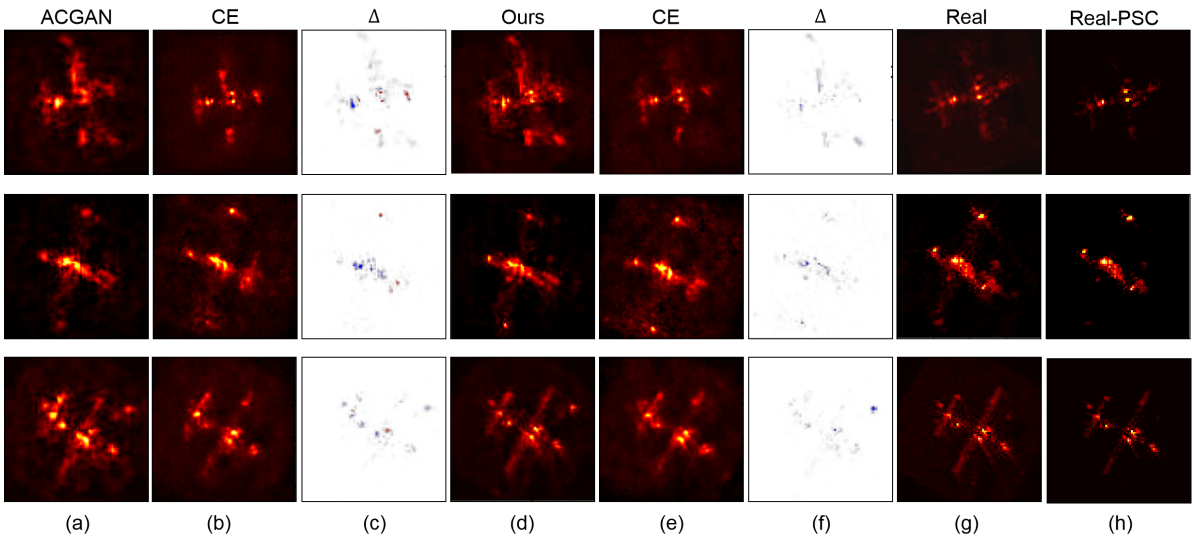

Figure 7. We evaluate the generated SAR-Airplane images by using the "X-Fake" method. (a) and (d) are generated images by ACGAN and ACGAN+ours, respectively. (b) and (e) are the counterfactual explanation results of (a) and (d), respectively. (c) and (f) are the differences between the generated SAR images and their counterfactual explanations, where the red contents illustrate the missing details of the generated images. (g) and (h) are the real SAR images with the given class and nearby azimuth angle conditions, and their PSC reconstruction results, respectively.

Method	10% MSTAR					5% MSTAR					3% MSTAR				
	VIF(↑)	FSIM(↑)	GMSD(↓)	FID(↓)	KID(↓)	VIF(↑)	FSIM(↑)	GMSD(↓)	FID(↓)	KID(↓)	VIF(↑)	FSIM(↑)	GMSD(↓)	FID(↓)	KID(↓)
Base	0.0386	0.7432	0.1510	290.0484	0.4548	0.0315	0.7291	0.1583	340.2440	0.4713	**0.0281**	**0.7243**	**0.1632**	**354.4560**	**0.5667**
+Ours	**0.0781**	**0.7622**	**0.1385**	**87.2719**	**0.0414**	**0.0547**	**0.7441**	**0.1485**	**130.2311**	**0.0861**	0.0217	0.6915	0.1745	705.7391	1.1700

Table 7. Limited data training discussion.

main condition because: **1)** Azimuth angle significantly influences the EM features of ground targets. **2)** Some datasets lack complete angle information (no depression angle information). **3)** Φ-GAN is independent of input conditions and can be applied to multiple angles. As a result, we also add some results of encoding both azimuth angle and depression angle in Tab.8 (10%/5% MSTAR), showing Φ-GAN still works well and makes improvements.

Method	AA (10%)	AA+DA (10%)	AA (5%)	AA+DA (5%)
ACGAN	290.0484	310.3221	340.2440	316.0579
+Ours	**87.2719**	**104.7421**	**130.2311**	**148.1926**

Table 8. FID results of encoding AA and DA as conditions.

9. Extrapolation results

Three extrapolation results where training data (10 per class) is restricted to AA ranges of [0,90], [45,135] and [45,225], respectively, are shown in Fig.8. We calculate the FID within every 15 degree range. The results of extrapolation are worse than interpolation but our method still shows superiority. The phenomenon is common in GAN-based method due to its unawareness of the 3D info. of the target. Combining 3D-GAN or NeRF could address this issue which will be studied in the future.

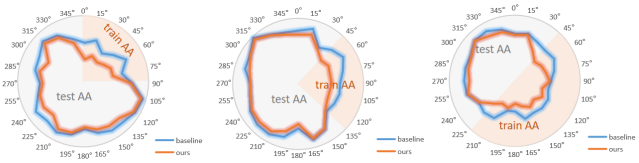

Figure 8. Results of generation in extrapolation scenarios.

10. Additional visualization

Fig.9 shows examples from 2S1 and T72 with movable barrels, where Φ-GAN generates clearer barrel features at specific angles than the baseline. Recognition performance for 2S1 and T72 improved by 35.04% and 43.87%, surpassing the average improvement of 29.95% in Tab.6, demonstrating Φ-GAN's effectiveness in generating superior features across various angles and enhancing recognition accuracy.

Fig.10 illustrates the evolution of FID and KID scores for the generated SAR images throughout the model's training process. As the number of training epoch increases, both

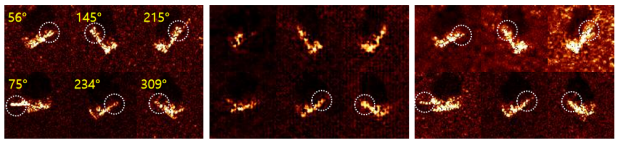

Figure 9. Row1-2: 2S1, T72. Left: GT. Middle: Baseline. Right: Φ-GAN.

FID and KID scores decrease more rapidly, ultimately converging to lower scores in our method. This suggests that our approach is more efficient in capturing image features, leading to faster convergence and improved performance in generating high-quality SAR images.

The following Fig.11,12,13,14 provide images generated on 5% and 10% MSTAR, 1% OpenSARShip and 14% SAR-Airplane datasets with different approaches. The comparison highlights the enhancement in image quality achieved with the application of Φ-GAN.

References

[1] Debora Chan, Juliana Gambini, and Alejandro C Frery. Entropy-based non-local means filter for single-look sar speckle reduction. *Remote Sensing*, 14(3):509, 2022. 1

[2] Dongwen Yang, Wei Ni, Lan Du, Hongwei Liu, and Jiadong Wang. Efficient attributed scatter center extraction based on image-domain sparse representation. *IEEE Transactions on Signal Processing*, 68:4368–4381, 2020. 1

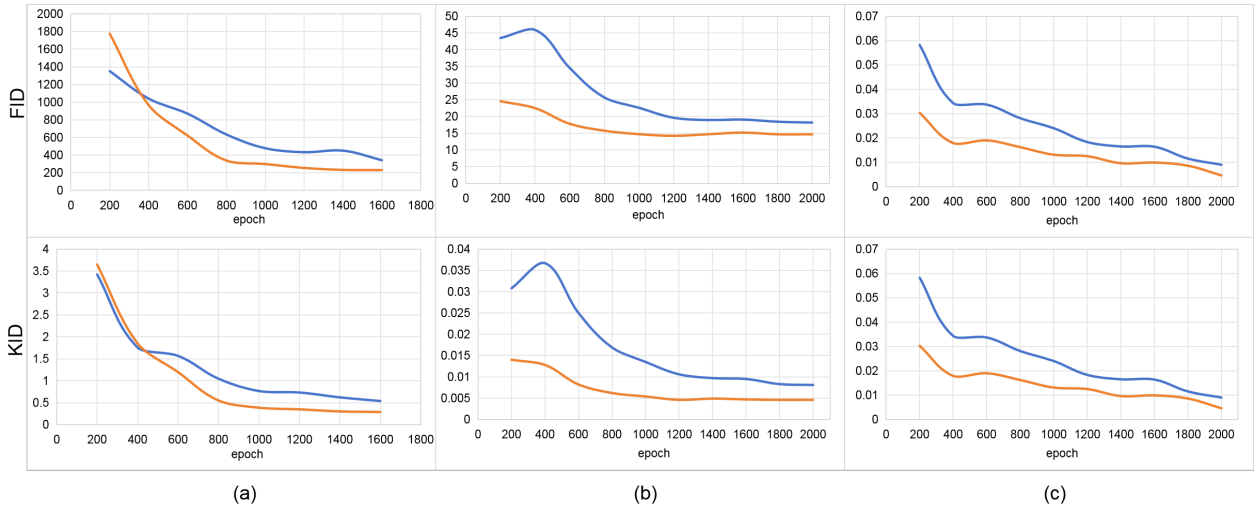

Figure 10. The trends of FID and KID of generated SAR images during model training (from 200th epoch to 2000th epoch).

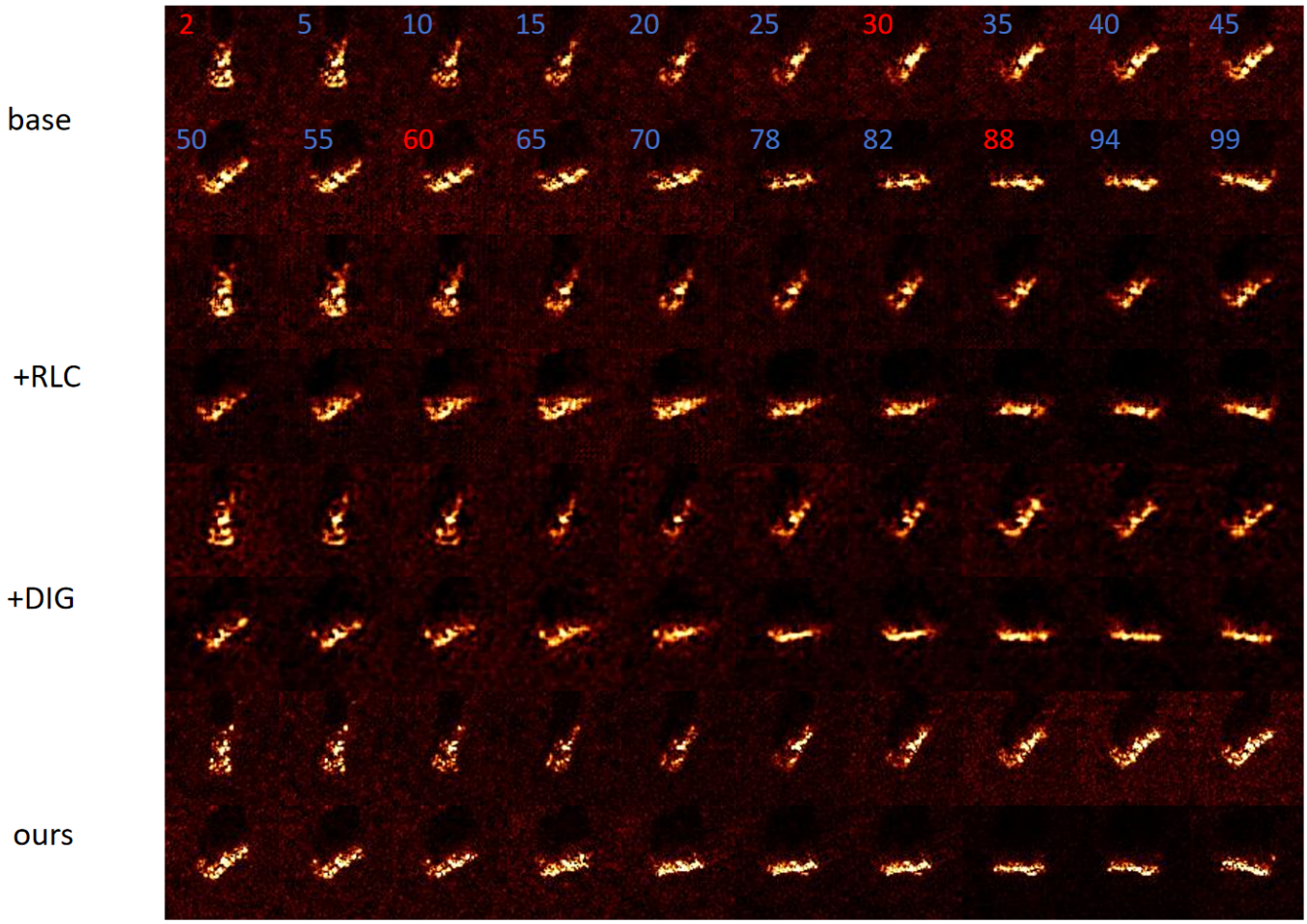

Figure 11. Generated images using different approaches on 10% MSTAR dataset. The angles near seen angles are marked in red and the unseen angles are in blue.

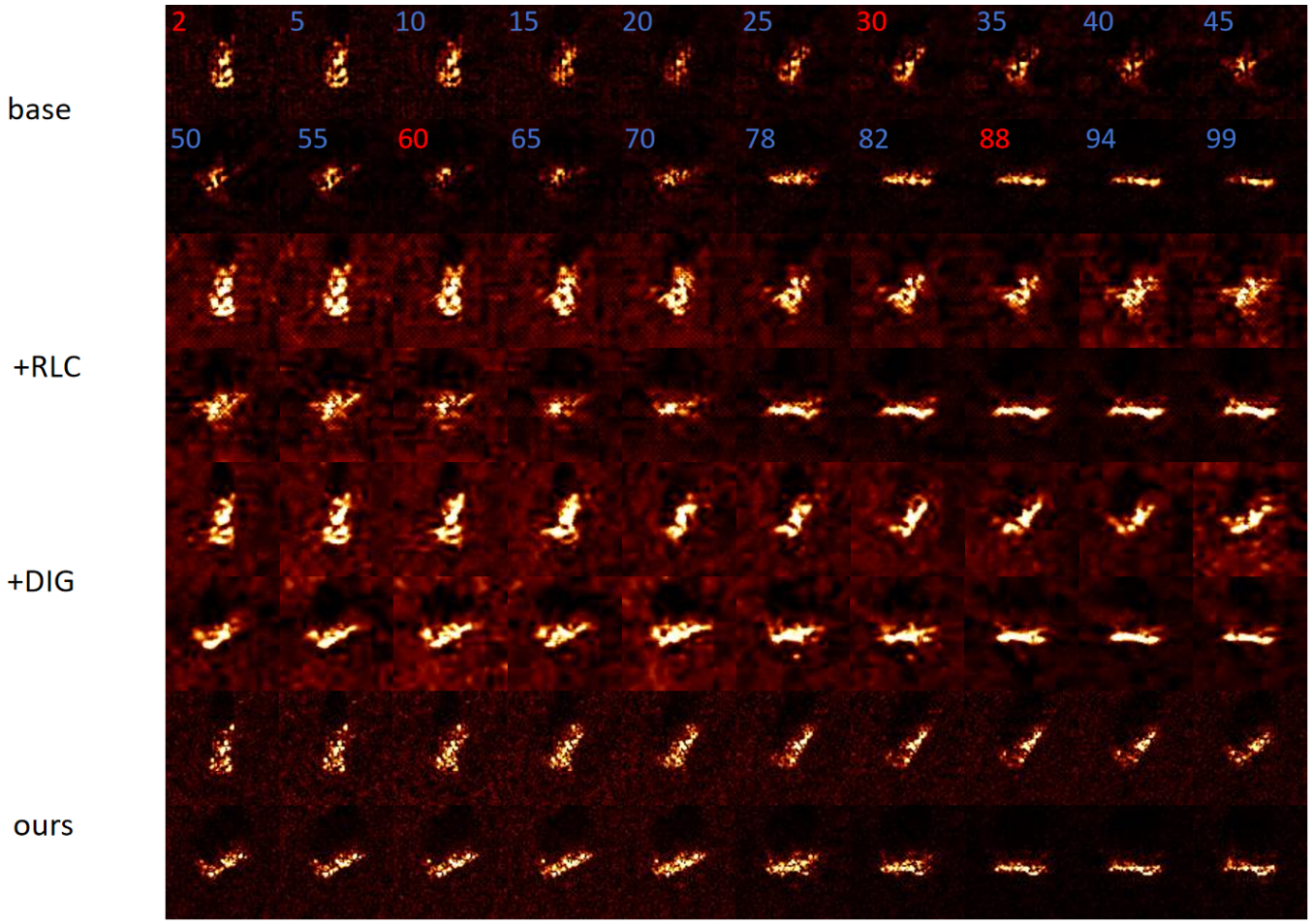

Figure 12. Generated images of 2S1 using different approaches on 5% MSTAR dataset.The angles near seen angles are marked in red and the unseen angles are in blue.

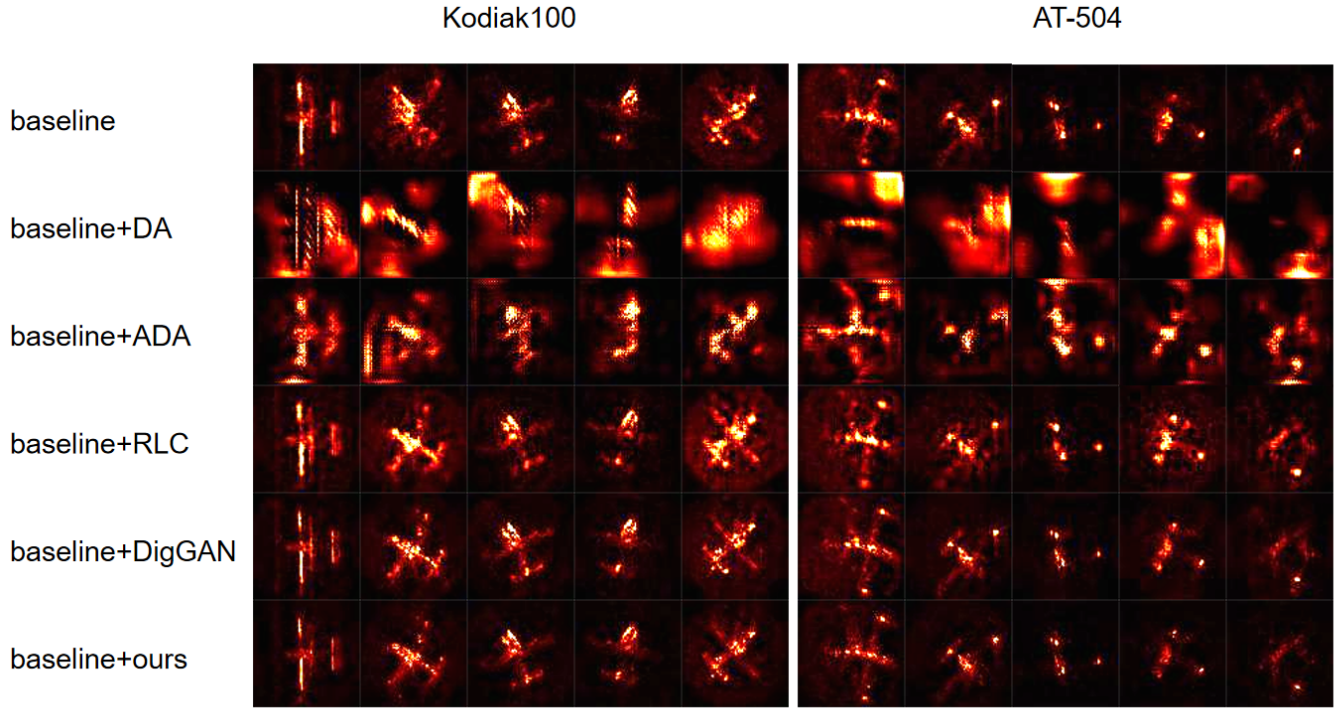

Figure 13. Generated images using different approaches on 14% SAR-Airplane dataset.

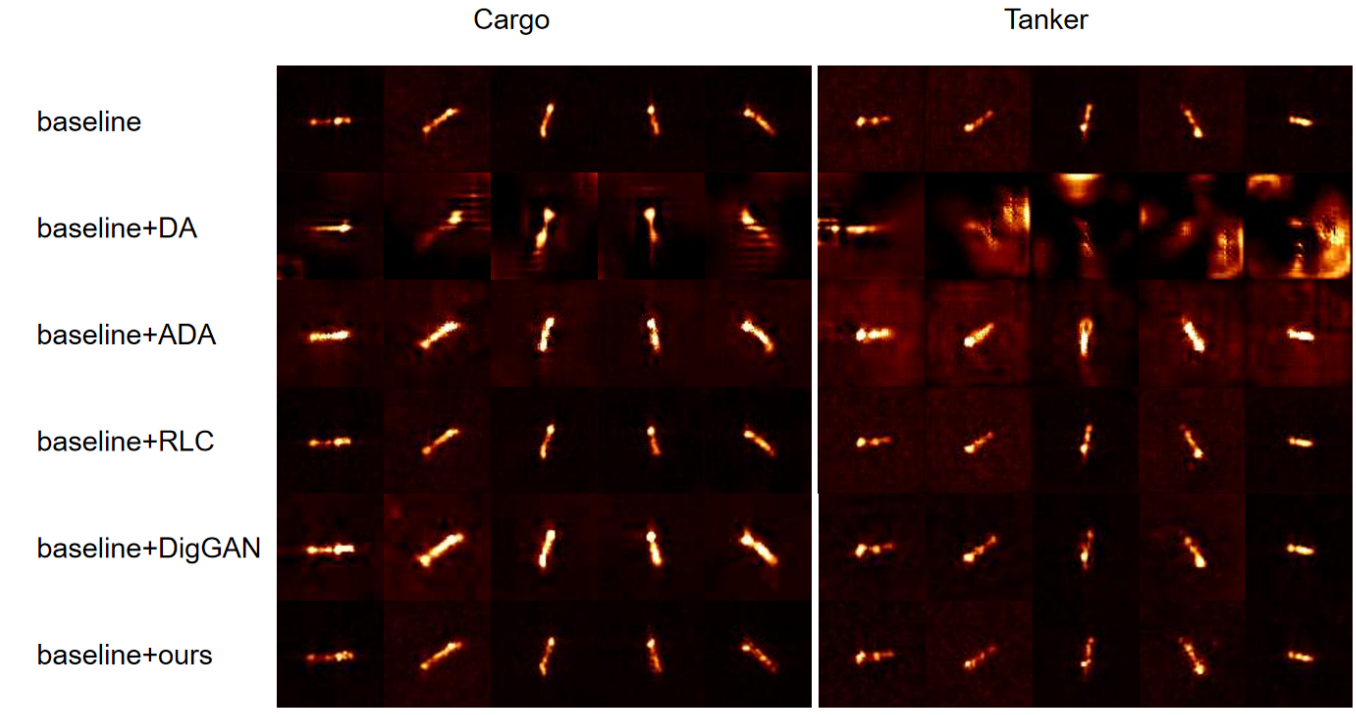

Figure 14. Generated images using different approaches on 1% OpenSARShip dataset.