GAN-Generated Ocean SAR Vignettes Classification

Omid Ghozatlou, *Student Member, IEEE*, Mihai Datcu, *Fellow, IEEE*, and Bertrand Chapron, *Fellow, IEEE*

***Abstract*— The use of deep learning (DL) in Earth observation technology has become essential. Even though DL models do remarkably well in classifying high-resolution satellite images and extracting semantic information, they frequently need a lot of training data, which can be costly and time-consuming to obtain. With an emphasis on ocean synthetic aperture radar (SAR) image analysis, this research investigates the use of synthetically generated data using generative adversarial networks (GANs) for data augmentation. Relying on GAN-based generated images for remote sensing applications requires a thorough assessment of the quality and authenticity of the generated images, as well as validation of the model's performance on real-world data. We assess the diversity and reliability of GAN-generated images by training a classification network on these images and evaluating their performance on real-world data. By comparing the classification accuracy in different experimental setups, we approximate the precision and recall for GANs performance.**

***Index Terms*— Data augmentation, deep neural network, generative adversarial networks (GANs), ocean pattern classification, synthetic aperture radar (SAR), synthetic image generation.**

I. Introduction

DEEP learning (DL) has emerged as a powerful tool in remote sensing image processing, playing a crucial role in Earth observation technology. However, implementing DL models for remote sensing image processing requires large amounts of training data, which can be expensive and time-consuming to acquire. To address this issue, researchers have proposed several solutions, including data augmentation techniques, transfer learning, and active learning [1].

One promising solution is the use of synthetic data generation using generative adversarial networks (GANs). GANs can provide more diverse and realistic samples compared with traditional data augmentation techniques. Synthetic data generated by GANs can be used to augment the training dataset, helping overcome the limitations of data scarcity and improve the performance of the DL model.

Received 7 June 2024; revised 28 August 2024; accepted 6 September 2024. Date of publication 24 September 2024; date of current version 3 October 2024. This work was supported by European Space Agency (ESA) through the "Explainable Synthetic aperture radar (SAR) Measurements for Wind Assessment With Artificial Intelligence" (ESAWAAI) Project, ESA, under Contract 4000142170/23/DT. *(Corresponding authors: Omid Ghozatlou; Mihai Datcu.)*

Omid Ghozatlou and Mihai Datcu are with the Center for Spatial Information (CEOSpaceTech), National University of Science and Technology POLITEHNICA of Bucharest (UNSTPB), 060042 Bucharest, Romania (e-mail: omid.ghozatlou@upb.ro; mihai.datcu@upb.ro).

Bertrand Chapron is with the Laboratoire d'Ocanographie Physique et Spatiale (LOPS), 29280 Brest, France, and also with IFREMER, 29280 Plouzané, France.

Digital Object Identifier 10.1109/LGRS.2024.3466970

Relying on GAN-based generated images for remote sensing applications requires a thorough assessment of the quality and authenticity of the generated images, as well as validation of the model's performance on real-world data. Selecting appropriate evaluation metrics is crucial for assessing the performance of GANs. However, traditional metrics, such as inception score (IS), Fréchet inception distance (FID) [2], and Kernel inception distance (KID) [3], may not fully capture the quality and diversity of the generated images for remote sensing applications. In this study, we discuss the reliability of GAN-based generated images and propose two approaches that approximate precision and recall for GAN performance.

A synthetic aperture radar (SAR) is crucial for advancing our understanding of the world's oceans, primarily due to its all-weather capability. Unlike optical sensors, which are affected by cloud cover and darkness, SAR operates effectively under any weather conditions, providing continuous data essential for studying dynamic ocean processes, such as ocean currents, wave patterns, and wind behavior.

In this study, we focus on ocean SAR image analysis and emphasize the need to evaluate the diversity and reliability of GAN-generated images. We train a classification network on these synthetic images and assess its performance on real-world data to ensure that the generated images are realistic and diverse enough to enhance model performance in practical applications. In addition, structural similarity index (SSIM) is exploited to check the correlation between generated images.

Our approach involves using a GAN to generate images, which are then used to train a classification network. To assess the reliability of the generated images, we utilize ResNet18 as a classifier and conduct two experimental setups. Each setup involves two balanced datasets: one of real images and the other of generated images. In the first setup, we train the classifier on real images and test it on generated images, while in the second setup, we do the reverse. By comparing classification accuracy across both setups, we approximate the precision and recall for GANs.

II. Related Works

Synthetic SAR image generation has been an active area of research in recent years which has got the attention of remote sensing experts. By reviewing the literature, we realize many research works address SAR image generation for ship object generation [4], [5], for SAR target image generation [6], [7],

[8], and optical to SAR image translation [9], [10], [11] but not for ocean image pattern. Despite these articles are not related to ocean SAR image generation [12], they provide valuable insights into the use of DL techniques, specifically GANs, for generating synthetic SAR images with high fidelity and resolution. For example, [13] explores the use of GAN to generate new spotlight SAR images from a limited preexisting dataset for target detection. Also, [14] introduces a hierarchical GAN network model to generate SAR images, gradually improving the resolution and quality of the images through a multistage network. More advanced study [15] addresses the issue of insufficient large-scale training datasets in remote sensing by ensuring the distribution consistency between generated synthetic images and real images. It incorporates two novel designs: a noise modeling pipeline that simulates the physical process of noise production during image capture and a detection-oriented image harmonization model that reduces variations between the pasted foreground and background in synthetic images. Some studies [6], [7], [16] take into account the azimuth angles of the SAR data to generate precise SAR images with controllable azimuths. Zeng et al. [6] propose ATGAN that reframes generation as image-to-image translation, using a coarse-to-fine generator and spectral-normalized patch discriminator to produce precise SAR images.

In the domain of SAR image classification, particularly for ocean scenes, addressing the challenge of misclassification caused by the presence of speckles is of paramount importance. For instance, [17] introduces a novel DL model, developed from the ground up, designed to automatically classify oceanic and atmospheric phenomena in Sentinel-1 SAR wave mode (WV) vignettes. The innovative preprocessing technique applied in this research effectively mitigates speckle noise and enhances subtle features within the SAR images, facilitating the discrimination of these complex scenes.

In order to improve ocean SAR image classification accuracy, Ma et al. [18] proposed a novel technique called Kernel entropy component analysis (KECA). KECA captures complex nonlinear structures in SAR data by optimizing the selection of informative features. To address computational challenges associated with KECA, the study introduces KECANet, a multistage convolutional kernel network that improves classification accuracy while maintaining computational efficiency.

In summary, the classification of SAR images for ocean monitoring has been extensively explored in scholarly literature, underscoring the potential of SAR technology for various oceanographic applications. SAR provides several advantages, including high-resolution imaging and sensitivity to changes on the ocean surface. However, it also presents challenges, such as the necessity for domain expertise to create labeled datasets. Researchers can enhance the performance of SAR in ocean monitoring applications by employing data augmentation techniques. Incorporating these considerations, GANs can be effectively utilized to generate synthetic images for data augmentation, thereby improving the performance of DL models in remote sensing image processing.

III. Methodology

In this section, we present the methodology of StyleGAN2 with adaptive discriminator augmentation (ADA) [19], which extends the original StyleGAN [20] framework with a dynamic augmentation strategy for the discriminator during training. The StyleGAN2-ADA architecture consists of a generator and a discriminator network, both of which are composed of convolutional layers. The architecture is similar to StyleGAN2, with the primary difference being the introduction of the ADA mechanism to stabilize training when using limited data. The methodology encompasses the following steps and components.

A. Mapping Network

The mapping network is a fully connected (MLP) network that maps points in the latent space to an intermediate latent space. This network is responsible for controlling the style of the generated images.

B. Intermediate Latent Space

The intermediate latent space, also known as the style space, is used to control the style of the generated images at different levels of detail. The style space is injected into the generator model at multiple points, allowing for fine-grained control over the generated images.

C. Weight Demodulation

Weight demodulation is a technique introduced in StyleGAN2 that reconstructs the augmentation operation from the original StyleGAN. It is used to control the style of the generated images at different levels of detail.

D. Noise Injection

Noise is introduced as a source of variation at each point in the generator model. This noise helps create stochastic variations in the generated images, such as texture and fine details.

E. Adaptive Discriminator Augmentation

The ADA mechanism in StyleGAN2-ADA applies random augmentations to the input images during training, which helps prevent overfitting and stabilizes training when using limited data.

The generator and discriminator networks are trained using an optimization process that minimizes their respective loss functions, such as the Wasserstein loss with gradient penalty. For a given real sample (x) and a random noise vector z, we calculate the real score $D(x)$ and fake score $D(G(z))$, which are the results of discrimination on the real sample and fake sample, respectively.

The loss function of the discriminator is defined as follows:

$$\text{Loss}_D = -\log(1 - \text{sigmoid}(D(G(z)))) - \log(\text{sigmoid}(D(x))) + L \quad (1)$$

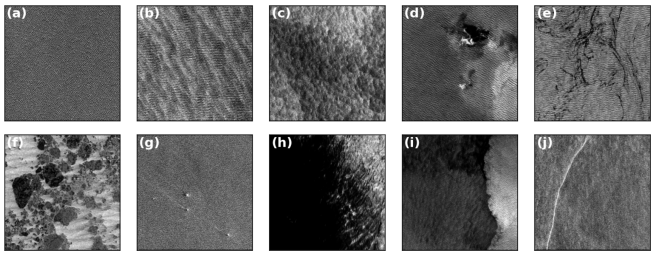

Fig. 1. Example of each class of TenGeoP-SARwv [21]. From (a)–(j) are pure ocean waves, wind streaks, micro convective cells, rain cells, biological slicks, sea ice, icebergs, low wind area, atmospheric front, and oceanic front.

where L is the Gradient penalty term and is defined as follows:

$$L = 0.5 * r_{1_\gamma} * \sum \nabla^2_{T_{\text{real}}} + 0.5 * r_{2_\gamma} * \sum \nabla^2_{T_{\text{fake}}} \tag{2}$$

where r_{1_γ} and r_{2_γ} are gradient penalty coefficients. $\nabla^2_{T_{\text{real}}}$ and $\nabla^2_{T_{\text{fake}}}$ are penalized gradient computations for real samples and generated samples, respectively. The loss function of the generator is the cross entropy of the generated samples, and it is calculated

$$\text{Loss}_G = -\log(\text{sigmoid}(D(G(z)))) + L_{\text{PPL}} \tag{3}$$

where L_{PPL} is path length regularization term.

IV. Experiments

This section describes the dataset, two experimental settings for classification, training detail, and evaluation measures utilized in this work.

A. Dataset

The dataset TenGeoP-SARwv, as provided by [21], comprises 37 560 SAR images sourced from Sentinel-1A WV acquisitions. These images represent ten distinct atmospheric and ocean-related physical phenomena. The ten distinct classes referred to in [21] are illustrated in Fig. 1 along with their corresponding class names. From Fig. 1(a) to (j) are pure ocean waves, wind streaks, microconvective cells, rain cells, biological slicks, sea ice, icebergs, low wind area, atmospheric front, and oceanic front.

B. Experimental Setups

We create two balanced datasets. The first dataset, D_1, consists of 500 real samples of each class, and the second dataset, D_2, consists of 500 generated images for each class. Images in D_2 are generated randomly with different random seed numbers. In the first setup, we train the classifier on real images using D_1 and test on D_2. This experiment will assess the quality of generated images. Higher accuracy means the classifier, which is trained on real images, can correctly classify the generated images. In other words, it means the generated images have high quality, and intraclass diversity is enough good. In the second setup, D_2 has been used for training and D_1 for testing. This setup approximates the recall of GANs and assesses the diversity of generated images. Higher accuracy in this setup shows that generated images have high diversity. The classifier used in both setups is ResNet18.

TABLE I
Comparison of FID Scores and Classification Accuracies for Each Class

	Class	FID	ACC_0	ACC_1	ACC_2
a	Pure Ocean Waves	2.34	99.06	99.9	86.6
b	Wind Streaks	3.51	98.4	97.16	72
c	Micro Convective Cells	9.93	96.28	92.5	46.8
d	Rain Cells	10.54	99.9	98.44	60
e	Biological Slicks	9.19	99.9	99.16	88.8
f	Sea Ice	18.76	96.04	95.51	73
g	Iceberg	15.69	98.2	97.36	14.8
h	Low Wind Area	14.86	97.99	92.94	94.8
i	Atmospheric Front	9.74	94.8	95.33	52.6
j	Oceanic Front	16.6	95.51	83.2	22.6
	Average	11.116	97.608	95.15	61.2

C. Training Details

In each experiment, StyleGAN2-ADA was trained separately for each class. The training starts using weight random initialization with a fixed random seed value. Each vignette is resized to 256^2 and converted from uint16 to uint8 with a range of [0–255] in portable network graphic (PNG) format. For the style-based generator, we use leaky rectified linear unit (ReLU) with $\alpha = 0.2$ and an equalized learning rate for all layers. The mapping network consists of eight fully connected layers, and the dimensionality of all input and output activations is 512. The classifier, ResNet18, has 18 layers, including convolutional layers and skip connections. It starts with a 7×7 convolutional layer followed by four layers of convolutional blocks, each containing two residual blocks. The optimizer for the classifier and GANs is Adam with $\beta_1 = 0$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. The batch size is set to 32 by considering the computational power and shared memory of the GPU. All experiments are conducted on a single NVIDIA Quadro GV100 GPU using the PyTorch framework with CUDA support.

D. Evaluation Metrics

In this letter, we calculate FID, which captures the similarity of generated images to real ones better than the IS. FID measures the distance between generated data and real data at the feature level and is defined as follows:

$$\text{FID} = \|m - m_w\|^2 + \text{Tr}\left[C + C_w - 2(CC_w)^{\frac{1}{2}}\right] \tag{4}$$

where m and m_w are the means of real image features and generated image features, respectively. C and C_w are the covariance matrices of the real image and generated image, respectively. Tr is the trace of the matrix. The results are provided by FIDs using 50 000 images drawn randomly from the training set and report the lowest distance encountered throughout training. It is worth noting that a lower value of FID demonstrates better performance. In addition to FID, we calculate the classification accuracy of each class. We note the first setup's accuracy, Auc_1, and the second setup's classification accuracy, Auc_2.

V. Results

Fig. 2 shows six randomly generated images for each class using StyleGAN2-ADA. For comparison, Fig. 1 displays

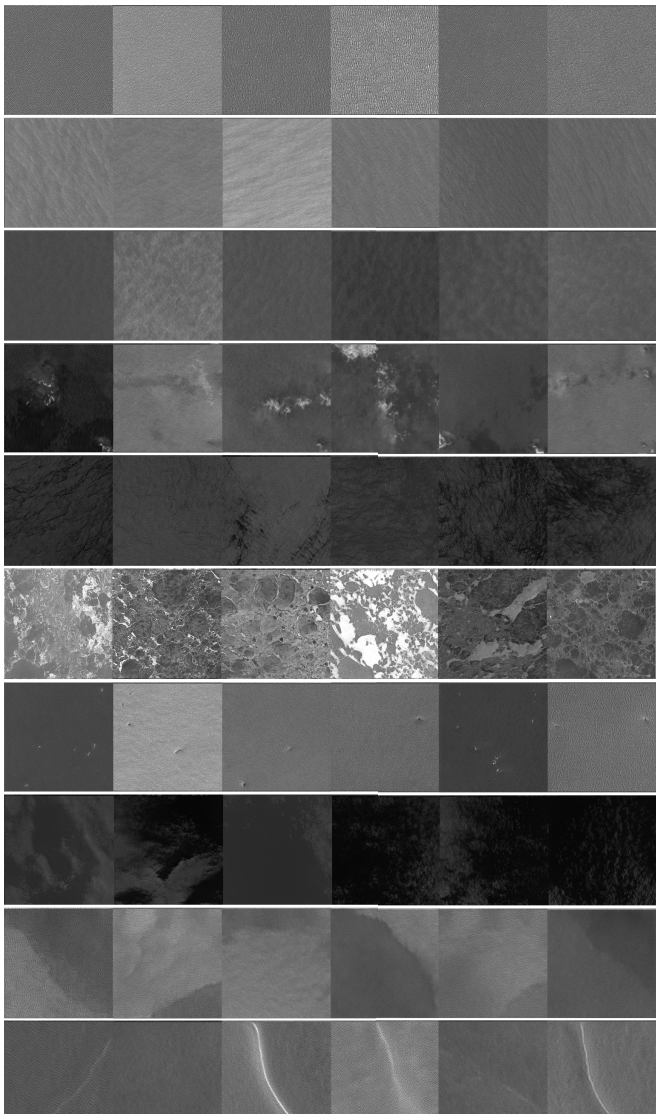

Fig. 2. Presenting six randomly generated images for each class in order (images are more bright for better visualization).

one real sample from each class. The generated images are remarkably realistic, closely resembling their real counterparts and making it challenging to differentiate between them. The model successfully created high-quality images across all classes, from simple pure ocean waves to complex sea ice textures. The sea ice images in the sixth row are particularly noteworthy, showcasing intricate features and varying contrasts that closely match real sea ice samples. This demonstrates the model's ability to generate diverse and high-fidelity images across different textural complexities. To thoroughly assess the model's performance in terms of fidelity and diversity, we use both quantitative measures, such as FID and a classifier-based approach. The latter involves evaluating classification accuracy in two distinct experimental setups.

Before presenting the final results of two experiments, to check the actual correlation between real and generated images and correlation among generated images in D_2, we calculate SSIM for each class. Fig. 3 compares SSIM between real and generated images, as well as the SSIM for generated images only. The results show that the average

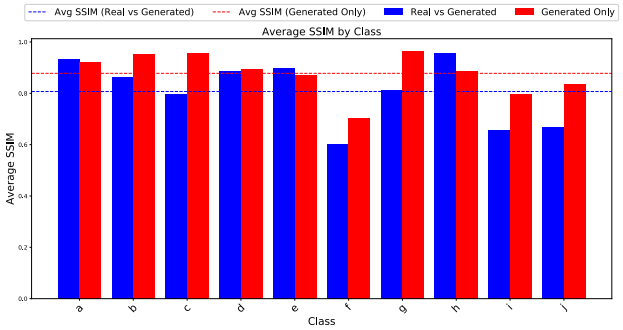

Fig. 3. SSIM between real and generated images, as well as the SSIM for generated images only.

SSIM for real versus generated images across all classes is 0.8351, which evaluates the visual quality of generated images. On the other hand, to evaluate the internal consistency of generated images or to analyze how different generations compare against each other, we calculate SSIM for generated images only with an average of 0.8692. This indicates that the generated images, on average, have a higher structural similarity when compared with other generated images rather than to the real images and demonstrates the lack of diversity of generated images, while generated images provide high visual quality.

However, it is important to note that the SSIM values vary across individual classes. For example, class "*d*" (rain cells) has a relatively high SSIM of 0.8844 for real versus generated images, but a lower SSIM of 0.8691 for generated images only. In contrast, class "*c*" (microconvective cells) has a lower SSIM of 0.7969 for real versus generated images but a higher SSIM of 0.9564 for generated images only. These findings suggest that the performance of the generative model may be class-dependent, and further investigation is needed to understand the factors contributing to these differences. We further explore this assessment by employing the classification-generated images.

Table I compares FID scores and classification accuracies for each class of ocean SAR images. The FID scores provide insight into the quality of generated images, with lower scores indicating better quality. Pure ocean waves and wind streaks have the lowest FID scores (2.34 and 3.51, respectively), suggesting high-quality generated images for these classes. In contrast, sea ice and oceanic front have the highest FID scores (18.76 and 16.6), indicating lower quality of generated images.

The classification accuracies are presented in three columns: ACC_0, ACC_1, and ACC_2. ACC_0 represents the accuracy on the original unbalanced dataset [21], showing consistently high performance across all classes (average 97.608%). ACC_1, representing accuracy when training on real images and testing on generated ones, is also high (average 95.15%), indicating that generated images closely resemble real ones in most cases. This comparison provides a more comprehensive view of the fidelity of the generated synthetic data by demonstrating how the same classification model performs on both real and synthetic data.

However, ACC_2, which represents accuracy when training on generated images and testing on real ones, shows a

significant drop (average 61.2%). This suggests that while generated images may look realistic, they often lack the diversity needed to train a classifier that generalizes well to real data. The disparity is particularly pronounced for classes, such as iceberg (14.8%) and oceanic front (22.6%), indicating that these classes pose significant challenges for the GAN in terms of generating diverse, representative samples.

Interestingly, there is not always a clear correlation between FID scores and classification accuracies. For instance, low wind area has a relatively high FID (14.86) but achieves the highest ACC_2 (94.8%). This suggests that FID alone may not be sufficient to evaluate the quality and diversity of generated images for classification tasks.

In conclusion, while the GAN performs well in generating high-quality images for most classes (as evidenced by low FID scores and high ACC_1), it struggles to capture the full diversity of real images, particularly for complex classes, such as iceberg and oceanic front. This is reflected in the lower ACC_2 scores, suggesting room for improvement in generating more diverse and representative synthetic images for ocean SAR data augmentation.

VI. Conclusion

In this study, we utilized StyleGAN2-ADA to generate synthetic ocean pattern images using SAR data from Sentinel-1A WV acquisitions. We assessed the reliability of the generated images by employing ResNet18 as a classifier in two experimental setups, each with balanced datasets of real and generated images. The consistency of generated images in the created dataset is assessed by SSIM. In addition, we calculated FID of the generated images and compared it with classification accuracy. Our findings lead to the following conclusions.

1) FID effectively evaluates the quality of generated images but is limited in quantifying their diversity, as evidenced by low average AUC-2 (61.2%) despite low average FID (11.116). This indicates that realistic images do not necessarily equate to diverse images.
2) A qualitative and quantitative comparison revealed that generating diverse images requires sufficiently varied datasets for training GANs.
3) GANs appear to face a tradeoff between quality and diversity. While they can produce high-quality images for dominant modes, they may struggle to maintain quality across the entire data distribution, resulting in lower quality for less represented modes.

References

[1] O. Ghozatlou, M. Datcu, A. Focsa, M. H. Conde, and S. L. Ullo, "A review and a perspective of deep active learning for remote sensing image analysis: Enhanced adaptation to user conjecture," *IEEE Geosci. Remote Sens. Mag.*, vol. 12, no. 3, pp. 125–148, Sep. 2024.

[2] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30. Red Hook, NY, USA: Curran Associates, 2017, pp. 1–12.

[3] C. Miao, J. Wu, J. Chen, S. Xiong, L. Wang, and Q. Wang, "Image generation of traditional Chinese window grilles based on generative adversarial networks," in *Proc. Int. Conf. Culture-Oriented Sci. Technol. (CoST)*, Aug. 2022, pp. 232–236.

[4] S.-J. Lee and K.-J. Lee, "Efficient generation of artificial training DB for ship detection using satellite SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11764–11774, 2021.

[5] L. Li, C. Wang, H. Zhang, and B. Zhang, "SAR image ship object generation and classification with improved residual conditional generative adversarial network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[6] Z. Zeng, X. Tan, X. Zhang, Y. Huang, J. Wan, and Z. Chen, "ATGAN: A SAR target image generation method for automatic target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 6290–6307, 2024.

[7] C. Wang et al., "SAR target image generation method using azimuth-controllable generative adversarial network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9381–9397, 2022.

[8] Y. Sun et al., "Attribute-guided generative adversarial network with improved episode training strategy for few-shot SAR image generation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1785–1801, 2023.

[9] H. Shi, B. Zhang, Y. Wang, Z. Cui, and L. Chen, "SAR-to-optical image translating through generate-validate adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[10] J. Bao, W. Yu, K. Yang, C. Liu, and T. J. Cui, "Improved few-shot SAR image generation by enhancing diversity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3394–3408, 2024.

[11] Y. Zhao, T. Celik, N. Liu, and H.-C. Li, "A comparative analysis of GAN-based methods for SAR-to-optical image translation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[12] O. Ghozatlou, M. Datcu, and B. Chapron, "GAN-based ocean pattern SAR image augmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2023, pp. 4056–4059.

[13] S. R. M. Bhamidipati, C. Srivatsa, C. Kanakapura Shivabasave Gowda, and S. Vadada, "Generation of SAR images using deep learning," *Social Netw. Comput. Sci.*, vol. 1, no. 6, p. 355, Nov. 2020.

[14] H. Huang, F. Zhang, Y. Zhou, Q. Yin, and W. Hu, "High resolution SAR image synthesis with hierarchical generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 2782–2785.

[15] W. Zhang et al., "Physics guided remote sensing image synthesis network for ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4700814.

[16] J. Zhai, X. Dang, F. Chen, X. Xie, Y. Zhu, and H. Yin, "SAR image generation using structural Bayesian deep generative adversarial network," in *Proc. Photon. Electromagn. Res. Symp.-Fall (PIERS-Fall)*, Dec. 2019, pp. 1386–1392.

[17] J. Anil Raj, S. M. Idicula, and B. Paul, "Novel deep learning-based technique for Sentinel-1 ocean SAR vignettes classification," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 5088–5091.

[18] M. Ma, C. Bai, S. Zhang, L. Qian, and H. Yan, "KECANet: A novel convolutional kernel network for ocean SAR scene classification with limited labeled data," *Frontiers Mar. Sci.*, vol. 9, Sep. 2022, Art. no. 935600.

[19] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, Red Hook, NY, USA: Curran Associates, 2020, pp. 12104–12114.

[20] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

[21] C. Wang et al., "A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode," *Geosci. Data J.*, vol. 6, no. 2, pp. 105–115, 2019.