EasyChair Preprint

№ 8342

Self-Perceptual Generative Adversarial Network for Synthetic Aperture Sonar Image Generation

Yuxiang Hu, Wu Zhang, Baoqi Li, Jiyuan Liu and Haining Huang

June 21, 2022

Self-Perceptual Generative Adversarial Network for Synthetic Aperture Sonar Image Generation

Yuxiang Hu[1,2], Wu Zhang[1], Baoqi Li[1], Jiyuan Liu[1], and Haining Huang[1]

[1] Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing, The Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China
{huyuxiang,zhangwu,libaoqi,liujiyuan,huanghaining}@mail.ioa.ac.cn
[2] University of Chinese Academy of Sciences, Beijing, 100049, China

Abstract. Due to the shortage of Synthetic Aperture Sonar (SAS) image datasets, the development of many underwater tasks is hindered. To tackle this problem, coupling optical rendering and image-to-image translation is a novel and feasible way. However, because of the big gap between simulated optical images and real SAS images, the performances of existing works are not desired and have plenty of room for improvement. In this letter, we introduce a Self-Perceptual Generative Adversarial Network (SPerGAN) which can controllably generate SAS images with high fidelity. It utilizes a kind of self-perceptual loss to generate high-quality and diverse SAS images. Moreover, we introduce a novel evaluation method of SAS image that accords closely with human cognition. To evaluate the performance of our method, we first compare it against recent outstanding image-to-image translation methods on qualitative and quantitative aspects. Then we make ablation studies to explore the effects of different cycle consistency loss and hyper-parameter. The results show that our method surpasses all existing methods and is able to generate diverse and realistic SAS images.

Keywords: SAS Image Generation · Generative Adversarial Network · Image-to-Image Translation.

1 Introduction

In many underwater tasks, such as Underwater Object Detection [1, 15, 28] and Seafloor Classification [14], there is an urgent demand for high-quality SAS image datasets. But unfortunately, collecting SAS image datasets cost so much time and manpower that few existing relevant datasets are available, which limits the development of the aforementioned underwater technology to a great extent. Therefore, it is necessary to find an efficient approach to acquire high-quality SAS images.

Usually, there are two paradigms for sonar image generation. One of them is generating sonar images by building an acoustic-imaging model which is similar to the optical-imaging model in computer graphics. Most of the existing relative

works [6, 16, 20, 26] are aimed to simulate forward-looking sonar images and they are very time-consuming and have poor performance. Another paradigm is coupling guide image synthesis and image-to-image translation to generate sonar images. This is a two-stage paradigm that generates guide images firstly, such as semantic image [12] or optical rendering image [19, 23, 25], then translates it to realistic sonar image. Thanks to the power of Generative Adversarial Networks (GANs) [5], it is able to generate far more realistic sonar images than the first paradigm. However, we find that these existing methods still cannot generate satisfying SAS images and control the content properly because of the limitation of their models in the second stage. For example, the results generated by Cycle-Consistent Adversarial Networks (CycleGAN) [30] are shown in Fig. 3. We can find that all orientations of shadow are the same. To acquire approving SAS images, a more powerful and robust image-to-image translation model should be applied.

Our image-to-image translation network is motivated by self-supervised learning and deep perceptual metric [13]. We argue that the $\mathcal{L}_1$ norm in cycle consistency loss is a strong constraint that minimizes the discrepancy between origin images and reconstruction images at pixel level. This strong constraint makes it difficult for the generator to learn about diverse representations so that generator can not control the content of generated SAS images according to guide images. Therefore, we can relax it to improve the performance of the generator. Recently, deep perceptual metric have been proved that it possesses superior performance than conventional low-level metrics [29], and it is not as strict as $\mathcal{L}_1/\mathcal{L}_2$ norm so we can leverage it to help the model relax. Usually, deep perceptual metric is based on a pre-trained network like VGG16 [24] which is trained in supervised training. However, most SAS image datasets are small and lack labels so it is hard to train an effective network. In recent years, self-supervised learning has been proven that it has great potential to learn the representation of data. In our work, we find that image-to-image translation itself can be seen as a pretext and the generator can learn rich representations in the training process. Therefore, we consider calculating deep perceptual metric by using the generator during the training. Based on that, we propose a new network called Self-Perceptual Generative Adversarial Network (SPerGAN). We demonstrate that SPerGAN surpasses the state-of-the-art methods in the SAS image generation task.

On the other hand, most SAS image generation works utilize Frechet Inception Distance score (FID score) [9] to evaluate the performance of the model. However, using this index for evaluation may be biased as big gap between a natural image and a SAS image. Therefore, we propose a novel evaluation index called Frechet SwAV Distance score (FSD score) where replacing Inception V3 [27] with ResNet [8] trained by using SwAV [2] on real SAS image datasets. As for SAS images, it is in line with human judgment.

To summarize, the main contributions of our work can be listed as follows:

1. We propose a novel image-to-image translation network called SPerGAN for SAS image generation.

2. We propose a novel index called FSD score for evaluating the performance of the model in the SAS image generation task.
3. Our experiments show that the proposed index is more suitable than the common index like FID for evaluation and the SPerGAN surpasses the state-of-the-art methods in the SAS image generation task.

2 Related Work

2.1 Two-stage Paradigm for SAS Image Generation

To the best of our knowledge, the earliest work using the two-stage paradigm to generate realistic SAS images is Chen et al. [3]. They try to utilize Conditional Generative Adversarial Nets (cGAN) [22] and style transfer based method [4] to generate photo-realistic sonar images respectively. Recently, Jiang et al. [12] utilize Photoshop-like tools to label segmentation map firstly, then combine Pix2PixHD architecture with SPADE block to translate segmentation map to realistic SAS image. Reed et al. [23] firstly combine POV-Ray and preprocessing to generate simulated SAS images and then use Wasserstein-GAN with gradient penalty (WGAN-GP) [7] to improve SAS realism. Our work is similar to [23], but there are substantial differences. Firstly, in the first stage, we simulate very simple optical images which only contain object and shadow information, instead of considering the texture of the object and seabed like [23]. Secondly, we utilize a more powerful image-to-image translation network for generating realistic SAS images.

2.2 Image-to-Image Translation

Recently, a parametric approach using CNNs architecture is proposed by Gatys et al. [4]. This method translates image pair each time requiring training once again. In 2017, Isola et al. [11] propose Pix2Pix networks, which utilize cGAN to achieve paired image-to-image translation. Furthermore, they invent a framework named CycleGAN [30] which can apply to unpaired image-to-image translation. This elegant framework consists of two pairs of GANs, each pair is responsible for one direction of image translation, and it cleverly takes advantage of cycle consistency loss as a content constraint to guarantee the mapping from input to output. In our work, we follow the basic architecture of CycleGAN, but we discard $\mathcal{L}_1$ norm cycle consistency loss and use a self-perceptual metric to achieve better results.

2.3 Perceptual Metric

As an effective way to evaluate the similarity between two images, perceptual metrics are widely used in image reconstruction, image-to-image translation, and image synthesis. In early works, researchers mainly utilize hand-crafted metrics for evaluating the quality of images, such as PSNR, SSIM, MS-SSIM and FSIM.

In 2016, Johnson et al. [13] propose a kind of perceptual loss that calculate the distance between two images' feature map extracted by VGG16, and they demonstrate that it has remarkable performance in style transfer and super-resolution. After that, Zhang et al. [29] make a comparative survey on the performance of classic and deep perceptual metrics, they find that deep features outperform all classic metrics by large margins. Recently, many works integrate perceptual metrics into GANs. In [10], SSIM and perceptual loss by using VGG19 are applied to cycle consistency loss for improving the quality of generated images. Differentiating from all the above works, we do not adopt the ImageNet pre-trained model to calculate the perceptual loss for cycle consistency, but creatively utilize inherent generators of CycleGAN to calculate it.

3 Approach

Our goal is to generate realistic SAS images in a controllable way by leveraging computer graphics and deep learning. We follow the two-stage paradigm, our framework consists of two components: (1) an optical renderer that is able to generate guide images according to scene settings, (2) a powerful and robust image-to-image translation network that is responsible to translate guide image to realistic SAS image. Note that we simulate very simple guide images which only contain object and shadow information in the first stage, and our emphasis is on the second stage.

3.1 Optical Rendering

To acquire guide images, we utilize Blender, a free and open source software with ray-tracing-based optical renderer. By using it, users are able to build scenes conveniently in accordance with specified settings and directly acquire high-quality optical images through the built-in optical renderer.

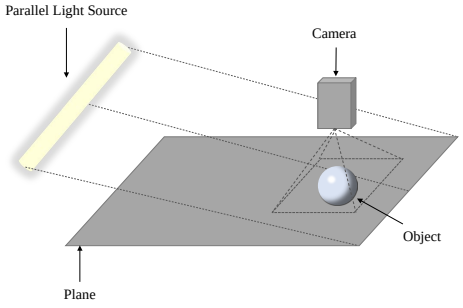

Fig. 1. A brief sketch of scene setting. The plane is infinite and the parallel light source consists of many dense point light sources in line. We use this way to model the scene and simulate guide images.

In our work, we build a scene containing four elements: (1) a parallel light source, (2) a camera, (3) a sphere object and (4) an infinite plane. For simulating alike shadow of objects in real SAS image, we use a simple setting: placing the camera directly above the object and using a parallel light source to produce SAS-like shadow. The parallel light source consists of many dense point light sources in line. By adjusting the relative location between parallel light source and object, we can simulate different shadow effects related to the positional relationship between sonar and object. A brief sketch of our modeling is shown in Fig. 1 and generated examples are shown in the first column of Fig. 3. It should be noted that our optical image simulation method is different from [23]. In [23], they have considered textures of objects and seabed in this part which is much more complex and time-consuming than ours.

3.2 Self-Perceptual Generative Adversarial Network

CycleGAN has achieved astonishing performance in image-to-image translation. Given two domain datasets X and Y, training samples can be denoted as $\{x_i\}_{i=1}^{N}$ where $x_i \in X$ and $\{y_j\}_{j=1}^{M}$ where $y_j \in Y$. We denote the distribution of x and y as p_x and p_y respectively. G and F are a pair of generators, G is responsible for translation $X \to Y$ and F is responsible for translation $Y \to X$. Moveover, there are two adversarial discriminators D_X and D_Y, where D_X distinguish between x and $F(y)$, D_Y distinguish between y and $G(x)$. The full objective of CycleGAN can be expressed as:

$$\begin{aligned}\mathcal{L}\left(G, F, D_X, D_Y\right) = \mathcal{L}_{GAN}\left(G, D_Y, X, Y\right) + \mathcal{L}_{GAN}\left(F, D_X, Y, X\right) \\ + \lambda_1 \mathcal{L}_{cyc}(G, F) + \lambda_2 \mathcal{L}_{identity}(G, F)\end{aligned} \tag{1}$$

where $\mathcal{L}_{GAN}\left(G, D_Y, X, Y\right)$ and $\mathcal{L}_{GAN}\left(F, D_X, Y, X\right)$ are least-squares adversarial loss [21], and $\mathcal{L}_{identity}(G, F)$ is identity mapping loss which is beneficial to color preservation of input image. $\mathcal{L}_{cyc}(G, F)$ is cycle consistency loss which keeps the reconstruction image close to the input image, it can be expressed as:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_x}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_y}[\|G(F(y)) - y\|_1] \tag{2}$$

Although CycleGAN is able to generate promising results, it seems can not change the content of generated image according to the guide image. We argue that $\mathcal{L}_1$ norm in original cycle consistency loss may be too strict and it limits the diversity of generated images. Consequently, we get rid of $\mathcal{L}_1$ norm and propose a novel self-perceptual loss to ensure the diversity of generated images.

As illustrated in Fig. 2, we take the outputs of the last residual block before the upsampling layer as feature maps, and we denote the feature extractors as ϕ_1 and ϕ_2 for G and F respectively. Moreover, we denote $G(x)$ and $F(y)$ as $\hat{y}$ and $\hat{x}$ respectively, denote $F(G(x))$ and $G(F(y))$ as x' and y' respectively. We utilize ϕ_1 and ϕ_2 to extract the feature maps of the input image and reconstruction image, then calculate the $\mathcal{L}_1$ distance between two feature maps as the self-perceptual loss. The self-perceptual loss can be expressed as:

$$\mathcal{L}_{sp}(G, F) = \mathbb{E}_{x \sim p_x}[\|\phi_2(x') - \phi_2(x)\|_1] + \mathbb{E}_{y \sim p_y}[\|\phi_1(y') - \phi_1(y)\|_1] \tag{3}$$

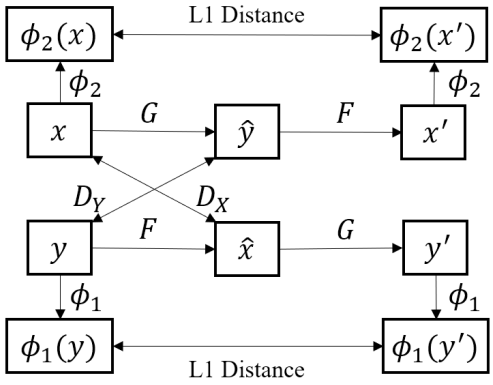

Fig. 2. The architecture of SPerGAN. We denote the part before the first upsampling layer of generator F as ϕ_2 and the similar part of generator G as ϕ_1. As for mapping $x \to \hat{y} \to x'$, we utilize feature extractor ϕ_2 to calculate the feature maps of input x and reconstruction x', and then calculate the $\mathcal{L}_1$ distance between two feature maps as self-perceptual loss for this mapping. As for another mapping $y \to \hat{x} \to y'$, we utilize extractor ϕ_1 to calculate the self-perceptual loss in a similar way.

Note that G and F used in self-perceptual loss are identical with the G and F used in the training process, which means our method does not need to train a pair of generators in advance. The full objective of SPerGAN can be expressed as:

$$\begin{aligned}\mathcal{L}\left(G, F, D_X, D_Y\right) = \mathcal{L}_{GAN}\left(G, D_Y, X, Y\right) + \mathcal{L}_{GAN}\left(F, D_X, Y, X\right) \\ + \lambda_1 \mathcal{L}_{sp}(G, F) + \lambda_2 \mathcal{L}_{identity}(G, F)\end{aligned} \tag{4}$$

where λ_1 in our most experiments is set 10, λ_2 use the default value 0.5. In follow-up experiments, we demonstrate that our method can not only tackle the problem of CycleGAN but also improve the quality of generated SAS images.

3.3 Evaluation for SAS Image Generation

Until now, there is no reliable evaluation method for the SAS image generation task. In recent SAS image generation works [23], [12], Frechet Inception Distance score (FID score) [9] is widely used due to its simplicity and robustness. However, there are two problems if we use FID to evaluate models directly: (1) FID utilizes ImageNet pre-trained model as feature extractor, but in fact, there is a big gap between optical image and SAS image, whether the features extracted from SAS images can effectively represent the statistical property of SAS images? (2) If we train the extractor from scratch on the SAS dataset which is usually very small, it easily leads to overfitting and the pre-trained extractor is unable to output representative features. Due to the aforementioned problems, we propose a novel evaluation method named Frechet SwAV Distance score (FSD score), it is suitable to measure the similarity between two SAS images by calculating the statistical property of their feature vectors. Different from FID, it utilizes a

ResNet pre-trained on the SAS image dataset by using a self-supervised learning method named SwAV to calculate the feature maps of SAS images.

For our FSD, we denote the feature extraction process as Φ, as for two SAS images x and y, the evaluation metric can be expressed as:

$$FSD(x, y) = \|\mu_x - \mu_y\| + Tr(\Sigma_x + \Sigma_y - 2\sqrt{\Sigma_x \Sigma_y}) \tag{5}$$

where μ_x and μ_y are expectation of feature maps $\Phi(x)$ and $\Phi(y)$ respectively, Σ_x and Σ_y are covariance of $\Phi(x)$ and $\Phi(y)$ respectively, Tr is the trace of matrix.

4 Experiments

In this section, we first evaluate the qualitative and quantitative performance of our method by comparing it with other advanced image-to-image translation methods. Then we make some ablation studies to explore the effects of different cycle consistency loss and self-perceptual coefficient λ_1.

4.1 Settings

Datasets We build and adopt an Optical Simulated Image to SAS image dataset (OSim-SAS) in our experiments. The optical image part consists of 32 images (256×256) simulated by using Blender, including sphere objects with different positions and orientations. The SAS image part consists of 8 images (212×212) acquired in the lake trial, all images contain sphere objects.

Baseline In our experiments, we compare our SPerGAN against five recent works in total. Firstly, we choose three unpaired image-to-image translation methods including CycleGAN [30], DRIT [17] and DRIT++ [18]. Secondly, we add a typical neural style transfer method [4] to the comparison experiment. Thirdly, we find that [23] has achieved promising results for SAS image generation by using WGAN-GP [7] based network, so it also compares with our method.

Training Details For all experiments, we utilize Adam solver with a batch size of 1, $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate is set to 0.0002 and all networks are trained from scratch. We keep the learning rate unchanged for the first 600 epochs and decay it to zero linearly over the next 400 epochs. Besides, we apply 9 residual blocks architecture used in CycleGAN as our generator and PatchGANs as our discriminator.

4.2 Qualitative Results

As illustrated in Fig. 3, we observe that Gatys, Reed, DRIT and DRIT++ are unable to generate satisfying SAS images. Gatys' method only learns coarse

background texture but not object characteristics, the results are not realistic. Reed's method can not learn any useful information. DRIT and DRIT++ have learned background texture and object characteristics to some extent, but they lead to offset of the object position that can be seen clearly in the fourth row of DRIT and DRIT++ in Fig. 3. CycleGAN has the best performance in baselines, but it does not learn the relation between the object and shadow so that it is unable to change the orientation of shadow according to the guide image. By contrast, our proposed method has superior learning ability and it can generate realistic SAS images with true relation of object and shadow.

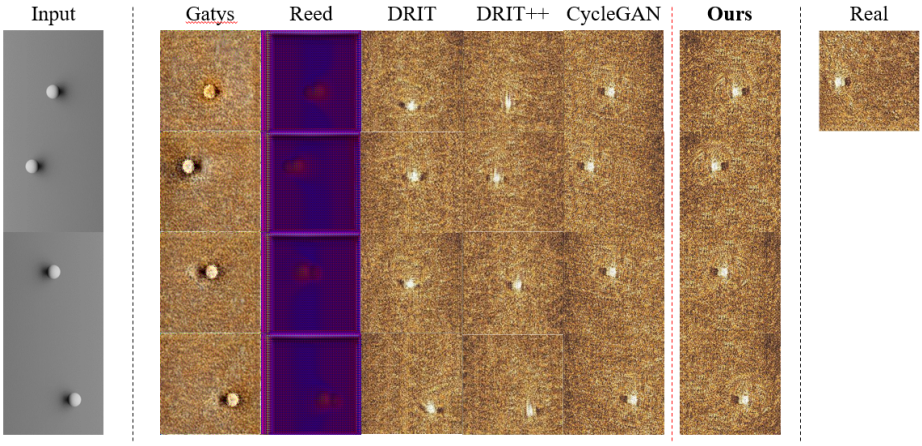

Fig. 3. Qualitative results of different methods. From left to right: input, neural style transfer network, WGAN-GP based network, DRIT, DRIT++, CycleGAN, our method trained on OSim-SAS dataset, and real SAS image (ground truth).

4.3 Quantitative Results

To demonstrate the reasonability of FSD, we firstly compare the scores of five baselines in different evaluation methods. The result is illustrated in Table 1. The sole difference of the three evaluation methods in Table 1 is using different extractors. In particular, SwAV in the table means that we utilize the extractor model trained on ImageNet by using SwAV. The lower the score, the better performance of the model.

Table 1. Quantitative comparison of different image-to-image translation methods in different evaluation methods.

	Gatys	Reed	DRIT	DRIT++	CycleGAN	Ours
FID	0.570	0.237	0.154	0.223	**0.040**	0.200
SwAV	**2.020**	27.22	34.57	34.31	17.98	26.48
FSD	6.962	928.0	7.174	5.322	5.169	**3.302**

Combining Table 1 with Fig. 3, we observe that FID gives a lower score to Reed than Gatys which means Reed's method has better performance than Gatys'. However, it is obviously wrong. SwAV thinks that the SAS image generated by Gatys has the best quality with a score of 2.020 and Reed is superior to DRIT/DRIT++, these are also not in line with human judgment. By contrast, our proposed evaluation method FSD has more reasonable results. Moreover, we find that score of Reed is much higher than others, which means FSD has better discrimination ability. In this way, our SPerGAN gets the lowest score (3.302) which demonstrates that it is able to generate more realistic SAS images.

4.4 Ablation Studys

Cycle Consistency Loss In this part, we only change cycle consistency loss. As shown in Fig. 4, utilizing SSIM and MS-SSIM as cycle consistency loss can not learn basic textures of SAS image. $\mathcal{L}_1$ norm and perceptual loss based on VGG16 can not adjust the shadow orientation. Besides, the latter leads to a false shadow shape. By contrast, our method is a better choice. Table 2 illustrates quantitative results of different cycle consistency loss and our method gets the lowest score (5.749).

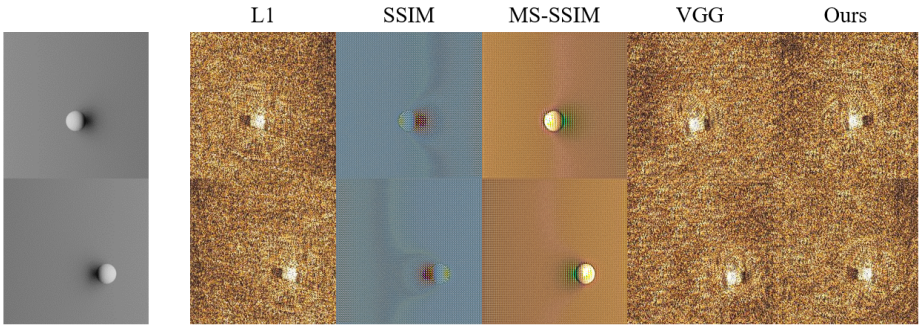

Fig. 4. Qualitative results of ablation study of cycle consistency loss.

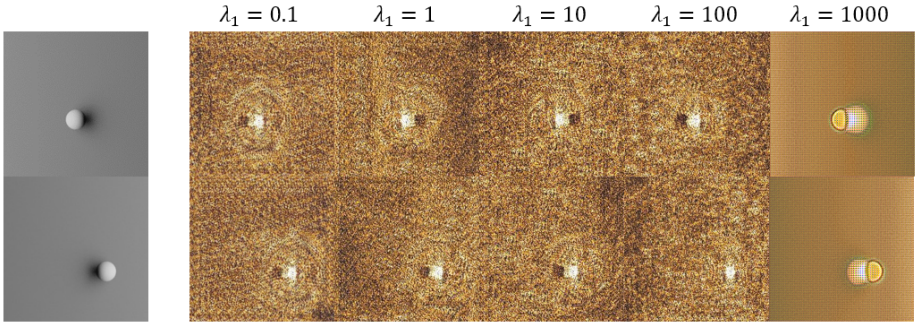

Fig. 5. Qualitative results of ablation study of self-perceptual coefficient.

Table 2. Quantitative results of ablation study of cycle consistency loss.

	$\mathcal{L}_1$	SSIM	MS-SSIM	VGG	Ours
FSD	7.108	11.51	92.55	6.500	**5.749**

Self-Perceptual Coefficient In this part, we only change the self-perceptual coefficient λ_1 in equation 4. The qualitative and quantitative results are shown in Fig. 5 and Table 3 respectively. It shows that the λ_1 should keep in a proper range. Too small or too large both lead to degradation of the model performance. Small λ_1 will generate artifacts around the object which can be seen in the column of $\lambda_1 = 0.1$ and large λ_1 makes it is hard for the model to learn the useful representations of SAS image.

Table 3. Quantitative results of ablation study of self-perceptual coefficient.

	0.1	1	10	100	1000
FSD	10.77	**6.231**	6.833	8.482	84.42

5 Conclusions

We utilize a two-stage paradigm to generate realistic SAS images. The first stage is using optical rendering to acquire guide images, and in the second stage, we translate them to realistic SAS images. As for the second stage, we propose a novel image-to-image translation network——SPerGAN. It can adapt to changes in the guide image and generates diverse and high-quality SAS images. To evaluate the performance of the network, we propose a novel evaluation index called Frechet SwAV Distance score (FSD score) which is in line with human judgment. In all experiments, our method performs better results than recent popular methods.

Acknowledgments. This work was supported by Institute of Acoustics, Chinese Academy of Sciences, under a project entitled, "Intelligent Classification of Underwater Objects in Sonar Images".

References

1. Berthomier, T., Williams, D.P., Dugelay, S.: Target localization in synthetic aperture sonar imagery using convolutional neural networks. In: OCEANS 2019 MTS/IEEE SEATTLE. pp. 1–9. IEEE (2019). https://doi.org/10.23919/OCEANS40490.2019.8962774
2. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Advances in Neural Information Processing Systems **33**, 9912–9924 (2020)

3. Chen, J.L., Summers, J.E.: Deep neural networks for learning classification features and generative models from synthetic aperture sonar big data. In: Proceedings of Meetings on Acoustics 172ASA. vol. 29, p. 032001. Acoustical Society of America (2016). https://doi.org/10.1121/2.0000458
4. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016). https://doi.org/10.1109/CVPR.2016.265
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
6. Gu, J.H., Joe, H.G., Yu, S.C.: Development of image sonar simulator for underwater object recognition. In: 2013 OCEANS-San Diego. pp. 1–6. IEEE (2013). https://doi.org/10.23919/OCEANS.2013.6741048
7. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. Advances in neural information processing systems **30** (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
9. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
10. Hwang, J., Yu, C., Shin, Y.: Sar-to-optical image translation using ssim and perceptual loss based cycle-consistent gan. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC). pp. 191–194. IEEE (2020). https://doi.org/10.1109/ICTC49870.2020.9289381
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017). https://doi.org/10.1109/CVPR.2017.632
12. Jiang, Y., Ku, B., Kim, W., Ko, H.: Side-scan sonar image synthesis based on generative adversarial network for images in multiple frequencies. IEEE Geoscience and Remote Sensing Letters **18**(9), 1505–1509 (2020). https://doi.org/10.1109/LGRS.2020.3005679
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016). https://doi.org/10.1007/978-3-319-46475-6-43
14. Kohntopp, D., Lehmann, B., Kraus, D., Birk, A.: Seafloor classification for mine countermeasures operations using synthetic aperture sonar images. In: OCEANS 2017-Aberdeen. pp. 1–5. IEEE (2017). https://doi.org/10.1109/OCEANSE.2017.8084752
15. Köhntopp, D., Lehmann, B., Kraus, D., Birk, A.: Classification and localization of naval mines with superellipse active contours. IEEE Journal of Oceanic Engineering **44**(3), 767–782 (2018). https://doi.org/10.1109/JOE.2018.2835218
16. Kwak, S., Ji, Y., Yamashita, A., Asama, H.: Development of acoustic camera-imaging simulator based on novel model. In: 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC). pp. 1719–1724. IEEE (2015). https://doi.org/10.1109/EEEIC.2015.7165431
17. Lee, H.Y., Tseng, H.Y., Huang, J.B., Singh, M., Yang, M.H.: Diverse image-to-image translation via disentangled representations. In: Proceedings

of the European conference on computer vision (ECCV). pp. 35–51 (2018). https://doi.org/10.1007/978-3-030-01246-5-3
18. Lee, H.Y., Tseng, H.Y., Mao, Q., Huang, J.B., Lu, Y.D., Singh, M., Yang, M.H.: Drit++: Diverse image-to-image translation via disentangled representations. International Journal of Computer Vision **128**(10), 2402–2417 (2020). https://doi.org/10.1007/s11263-019-01284-z
19. Liu, D., Wang, Y., Ji, Y., Tsuchiya, H., Yamashita, A., Asama, H.: Cyclegan-based realistic image dataset generation for forward-looking sonar. Advanced Robotics **35**(3-4), 242–254 (2021). https://doi.org/10.1080/01691864.2021.1873845
20. Mai, N.T., Ji, Y., Woo, H., Tamura, Y., Yamashita, A., Asama, H.: Acoustic image simulator based on active sonar model in underwater environment. In: 2018 15th International Conference on Ubiquitous Robots (UR). pp. 775–780. IEEE (2018). https://doi.org/10.1109/URAI.2018.8441870
21. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Smolley, S.P.: On the effectiveness of least squares generative adversarial networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(12), 2947–2960 (2019). https://doi.org/10.1109/TPAMI.2018.2872043
22. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
23. Reed, A., Gerg, I.D., McKay, J.D., Brown, D.C., Williamsk, D.P., Jayasuriya, S.: Coupling rendering and generative adversarial networks for artificial sas image generation. In: OCEANS 2019 MTS/IEEE SEATTLE. pp. 1–10. IEEE (2019). https://doi.org/10.23919/OCEANS40490.2019.8962733
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint (2014). https://doi.org/arXiv:1409.1556
25. Sung, M., Kim, J., Lee, M., Kim, B., Kim, T., Kim, J., Yu, S.C.: Realistic sonar image simulation using deep learning for underwater object detection. International Journal of Control, Automation and Systems **18**(3), 523–534 (2020). https://doi.org/10.1007/s12555-019-0691-3
26. Sung, M., Lee, M., Kim, J., Song, S., Song, Y.w., Yu, S.C.: Convolutional-neural-network-based underwater object detection using sonar image simulator with randomized degradation. In: OCEANS 2019 MTS/IEEE SEATTLE. pp. 1–7. IEEE (2019). https://doi.org/10.23919/OCEANS40490.2019.8962403
27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016). https://doi.org/10.1109/CVPR.2016.308
28. Williams, D.P.: The mondrian detection algorithm for sonar imagery. IEEE Transactions on Geoscience and Remote Sensing **56**(2), 1091–1102 (2017). https://doi.org/10.1109/TGRS.2017.2758808
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018). https://doi.org/10.1109/CVPR.2018.00068
30. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017). https://doi.org/10.1109/ICCV.2017.244