

Research Article

Object-Level Remote Sensing Image Augmentation Using U-Net-Based Generative Adversarial Networks

Jian Huang ¹, Shanhui Liu ², Yutian Tang ³, and Xiushan Zhang ¹

¹School of Electrical Engineering, Naval University of Engineering, Hubei, China

²School of Information Science and Technology, Hubei University, Hubei, China

³School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Correspondence should be addressed to Yutian Tang; tangyt1@shanghaitech.edu.cn

Received 25 July 2021; Accepted 24 August 2021; Published 9 September 2021

Academic Editor: Yulin Wang

Copyright © 2021 Jian Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of deep learning in computer vision, semantic segmentation technology is constantly employed for processing remote sensing images. For instance, it is a key technology to automatically mark important objects such as ships or port land from port area remote sensing images. However, the existing supervised semantic segmentation model based on deep learning requires a large number of training samples. Otherwise, it will not be able to correctly learn the characteristics of the target objects, which results in the poor performance or even failure of semantic segmentation task. Since the target objects such as ships may move from time to time, it is nontrivial to collect enough samples to achieve satisfactory segmentation performance. And this severely hinders the performance improvement of most of existing augmentation methods. To tackle this problem, in this paper, we propose an object-level remote sensing image augmentation approach based on leveraging the U-Net-based generative adversarial networks. Specifically, our proposed approach consists two components including the semantic tag image generator and the U-Net GAN-based translator. To evaluate the effectiveness of the proposed approach, comprehensive experiments are conducted on a public dataset HRSC2016. State-of-the-art generative models, DCGAN, WGAN, and CycleGAN, are selected as baselines. According to the experimental results, our proposed approach significantly outperforms the baselines in terms of not only drawing the outlines of target objects but also capturing their meaningful details.

1. Introduction

With the continuous development of satellite remote sensing image technology, high-resolution satellite image makes the target segmentation technology of satellite image realized. In many fields, the segmentation of satellite image can help to collect information and collect information quickly. Different from other image segmentation, satellite image contains a large number of elements and is easily affected by weather and season, so it needs a large number of datasets for training; otherwise, the target model will have difficulty in learning the relevant feature distribution. Especially, for the ship remote sensing image, because the ship is usually in dynamic change, it is difficult to collect a lot of data from

the target, so it is necessary to augment the dataset. For semantic segmentation task, when training a model, paired data is needed, that is, an original image and an image with semantic tag. Therefore, we need to construct two corresponding images at the same time.

For traditional data augmentation methods such as Cut-Out [1], the input square region is randomly masked in the training process, which can improve the robustness and overall performance of the convolutional neural network. CutMix [2] generates a new training sample by randomly combining two trimmed training samples, which makes its performance better than CutOut. However, these methods generate new samples directly modified and then stitched at the image level, which means that the boundaries among

different objects cannot be clearly identified. Since boundaries are of vital importance in the semantic segmentation task, the above-mentioned methods are not suitable for augmenting samples targeting at semantic segmentation task.

In recent years, the concept of generative adversarial networks [3] (GAN) has become one of the most popular unsupervised algorithms. For instance, DCGAN [4] and Marta GAN [5] have been proposed to augment remote sensing images. However, due to the complexity and uncertainty presented in remote sensing images, it is difficult for GAN-based augmentation methods to learn the distribution characteristics of the target objects, resulting in unsatisfactory augmentation effect. For example, the resolution of the generated images is limited while most of meaningful details are missed. Moreover, GAN-based augmentation methods cannot be able to generate the paired semantic tag images which are critical to enable semantic segmentation task and usually annotated manually with high cost. Therefore, it is desired to propose an approach to augment remote sensing images by effectively tackling the complexity and reducing the annotation cost.

Recently, conditional GANs [6] are proposed, which is a variant of GANs and capable of performing the image translation task. Inspired by conditional GANs, we propose an approach consisting two main components including the semantic tag image generator and the translator to augment remote sensing images. Firstly, the target objects are extracted by learning original training samples and then reasonably composed to construct the semantic tag image. Secondly, the translator based on U-Net [7] GANs is responsible of transforming the generated semantic tag images into realistic-looking images (please refer to Section 4 for more details).

In this work, our contributions could be summarized as follows:

- (i) A framework based on U-Net GANs for remote sensing image augmentation is proposed in this paper.
- (ii) A new method to automatically generate semantic tag images is proposed with a set of heuristic generation rules and restriction rules.
- (iii) Comprehensive experiments are conducted on a public remote sensing image dataset while in-depth analysis is provided focusing on the comparison between our proposed approach and baselines.

The rest part of this paper is organized as follows. Concrete examples about remote sensing image augmentation and basics of both GAN and U-Net models are offered in Section 2. The remote sensing image augmentation problem is formally defined in Section 3. In Section 4, the methodology is illustrated in detail including the overall architecture of the proposed approach, semantic tag image generator, and remote sensing image translator. In Section 5, the experiments are conducted on a public remote sensing image dataset to validate the effectiveness of the proposed approach. Related works about existing remote sensing image augmentation solutions are discussed in Section 6 followed by the conclusion provided in Section 7 to summarize this work.

2. Preliminaries

2.1. Examples of Remote Sensing Image Augmentation. Different from the data augmentation task directly performed at the image scale, for semantic segmentation tasks, they mainly focus on differentiating target objects from the background. The reason that traditional image-level augmentation cannot well support semantic segmentation tasks is due to its inability to identify the features of target objects, or the boundary between target objects and the background. Therefore, it becomes the motivation for the work in this paper to propose an approach to augment the original images at the object level rather than the image level.

To further illustrate the difference between image-level and object-level remote sensing image augmentation, examples are provided as shown in Figure 1. The upper row of Figure 1 shows the typical augmentation operations such as crop, flipping, cutout, and stretch usually adopted in image-level augmentation. And the lower row of Figure 1 presents object-level augmentation operations including object remove, object flipping, cutout without destroying the integrity of the original object, and semantically reasonable object add. Obviously, object-level remote sensing image augmentation can better serve the semantic segmentation task by flexibly composing different objects into the newly generated images, compared with the image-level counterpart.

2.2. Basics of Generative Adversarial Networks. Generative adversarial networks (GANs) were introduced in 2014 [3] and widely applied to various application scenarios [5, 8, 9]. GAN is able to produce high-quality output images through the mutual game learning of (at least) two independent modules: the generative model and the discriminative model.

- (1) Generative model (aka Generator) has the goal of capturing the data distribution from training samples by receiving a random noise z and generating an image from that noise, which is denoted as $G(z)$
- (2) Discriminative model (aka Discriminator) has the task of telling if the current sample comes from the training set or from the Generator. Its input parameter is x , which may be extracted from the training sample or the “fake” sample generated by the Generator. Its output is 1 or 0, while 1 indicates that the Discriminator judges the sample as the real sample and 0 means that the Discriminator judges the sample as the fake sample.

The generative model G aims to learn a distribution p_g over data x , by building a mapping function from a prior noise distribution $p_z(z)$ to the data space, $G(z; \theta_g)$, where θ_g are the parameters of the model G , e.g. the weights of the multilayer perceptrons to implement G .

The discriminative model $D(x; \theta_d)$ is an independent module to be implemented as a binary classifier, which outputs a single scalar (i.e., 0 or 1) representing the probability that x came from the training set rather than p_g .

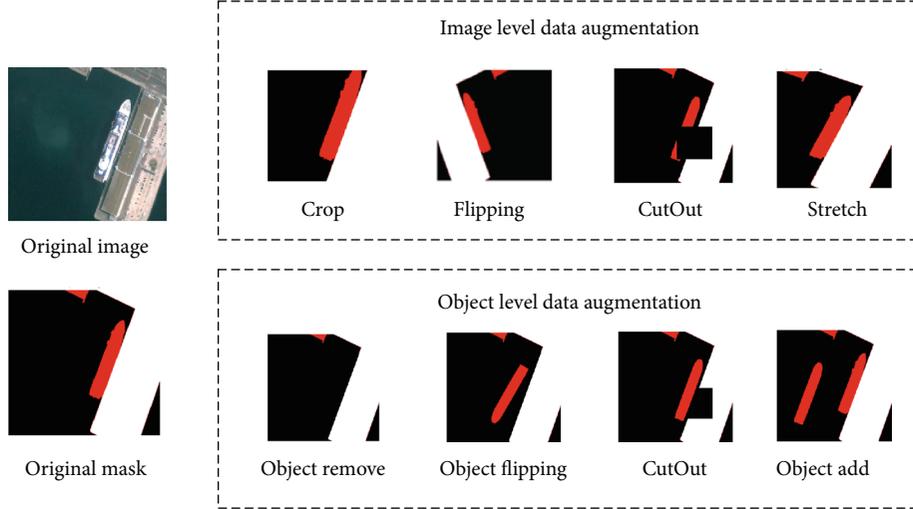


FIGURE 1: Examples of image-level and object-level remote sensing image augmentation.

Then, both models are jointly trained to play the following two-player min-max game as defined in the following equation until they reach the Nash Equilibrium:

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))]. \quad (1)$$

2.3. Basics of U-Net. In the last decade, deep learning models have been universally applied to different application scenarios such as time series analytics [10], rate adaptation [11], and edge computing [12]. As a dedicated deep learning model, U-Net has made great successes on the semantic segmentation task of medical images [7]. Therefore, it is chosen as the basic component of the model employed in this work for object-level remote sensing image augmentation. U-Net is named due to its symmetrical structure looks like an upper letter U as shown in Figure 2. For an input image with the size $N \times N$, U-Net will firstly conduct the 3×3 convolution for twice as shown in the upper left component of Figure 2. And then, the max pooling will be executed to downsize the output of the last layer of the upper component to fit the size of the first layer of the lower component, which is shown as the red downward arrow. The above procedure will be iterated for 4 times until the bottom component is reached, which plays the role as the conjunction of the left and right edges of U-Net. On the right edge of U-Net, the 3×3 convolution will be done within each component. Different from the operations among components of the left edge, the 2×2 upconvolution is conducted to restore the downsized sample to its original size. Moreover, the gray horizontal arrow represents the copy and crop operations on the output of the last layer of the left edge component that is taken as part of the input of the first layer of the corresponding right edge component. This feature is regarded as one of the factors making U-Net so successful on semantic segmentation of medical images, which embeds multi-scale grain of the input image into the learning process.

3. Problem Formulation

The object-level remote sensing image augmentation problem could be formulated as follows.

The following are given:

- (i) $I = \{p\}^{W \times H}$, where I is a remote sensing image, p is a pixel, and W and H mean the width and height of the image
- (ii) $T = \{t_n\}^{W \times H}$, where T is a semantic tag image and t is a pixel with the tag n
- (iii) $S = \{I_k, T_k\}$, where S is the training set with K paired original image I_k and its semantic tag image T_k .

Assume that

- (i) The pixels p and t at the same position of the paired I_k and T_k indicate the same object
- (ii) Each object identified in T_k consists a set of pixels which are spatially connected
- (iii) There exists at least one mapping function from T_k to I_k , which not only draws the outlines of objects of T_k but also captures their meaningful details to be presented in I_k .

The objective is as follows:

- (i) By learning the mapping function from T_k to I_k , it aims to generate a set of synthetic remote sensing images $A = \{I_m, T_m\}$ which are of higher diversity and reasonably realistic looking.

In the following section, we would like to introduce an approach to augment the remote sensing images at the object level by leveraging the generative adversarial network architecture based on U-Net.

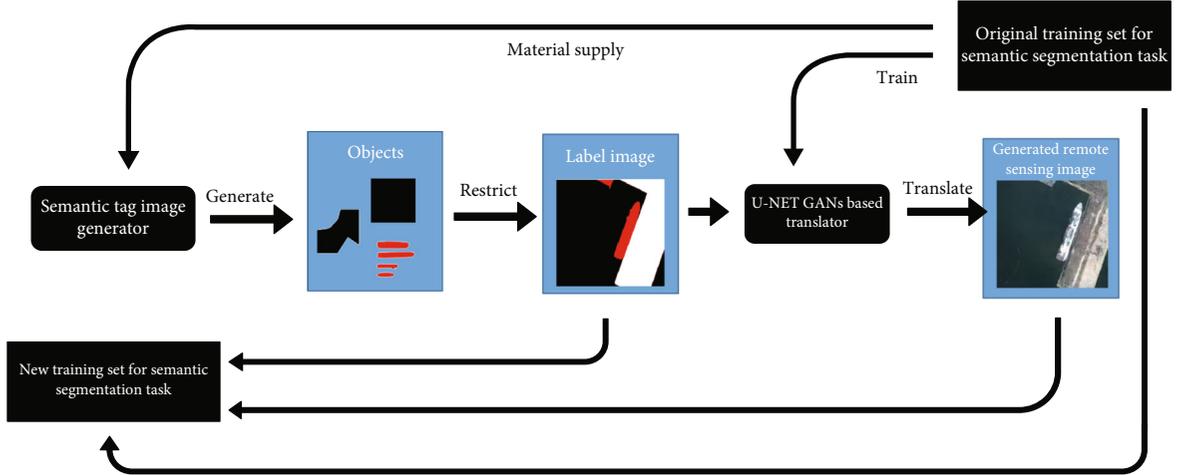


FIGURE 3: Overall architecture of the proposed approach to object-level remote sensing image augmentation.

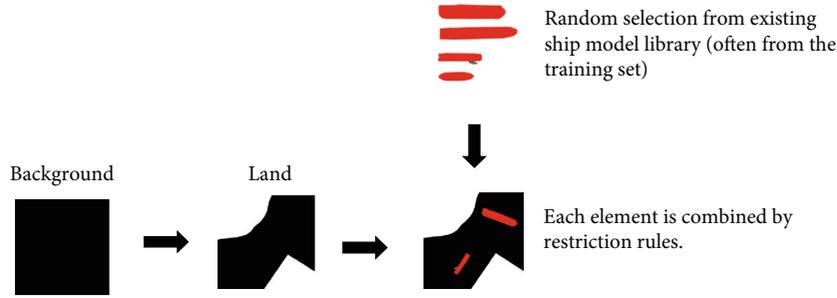


FIGURE 4: Procedure for generating semantic tag image.

By applying the above introduced generation rules and restriction rules to the generation procedure of tag images, proper tag images with typical objects flexibly placed could be trivially obtained. Afterwards, given the generated tag images, we propose a translator based on U-Net GANs to transform each tag image to a synthetic remote sensing image by learning pixel-wise details from the original training set. And the second use of the tag image is to take them as the ground truth for training and testing the semantic segmentation model. The design details of the translator are provided in Section 4.3.

4.3. U-Net GAN-Based Translator. Given the generated tag image, the translator is responsible in transforming it to a synthetic but realistic-looking remote sensing image. In this paper, we employ the generative adversarial networks as the reference for implementing the translator. As shown in Figure 5, the translator consists two key components including a generative model Generator and a discriminative model Discriminator. The design of the Generator conforms to that of U-Net introduced in Section 2.3. And the design of Discriminator is a FCN model as proposed by [13]. The training procedure for U-Net GAN-based translator follows the steps stated as below.

- (i) Step 1: Generator takes the semantic tag image as the input which is denoted as z
- (ii) Step 2: Generator generates a new remote sensing image denoted as IMG' which tries to deceive the Discriminator for judging IMG' as the real remote sensing image
- (iii) Step 3: Discriminator makes every endeavour to figure out whether IMG' is real or fake according to original images IMG from the training set
- (iv) Step 4: the above three steps will be executed in an iterative manner until the Nash Equilibrium is reached.

The optimization goal of U-Net GAN-based translator consists two parts. The Generator denoted as G for short needs to learn a distribution p_g over output images x by building a mapping function $G(z; \theta_g)$ from the given tag image distribution $p_z(z)$ to the original image representation space. θ_g are the parameters of the Generator, i.e., the weights of U-Net implementing Generator in this paper. And the Discriminator denoted as D for short is implemented as a binary classifier, which outputs a single scalar

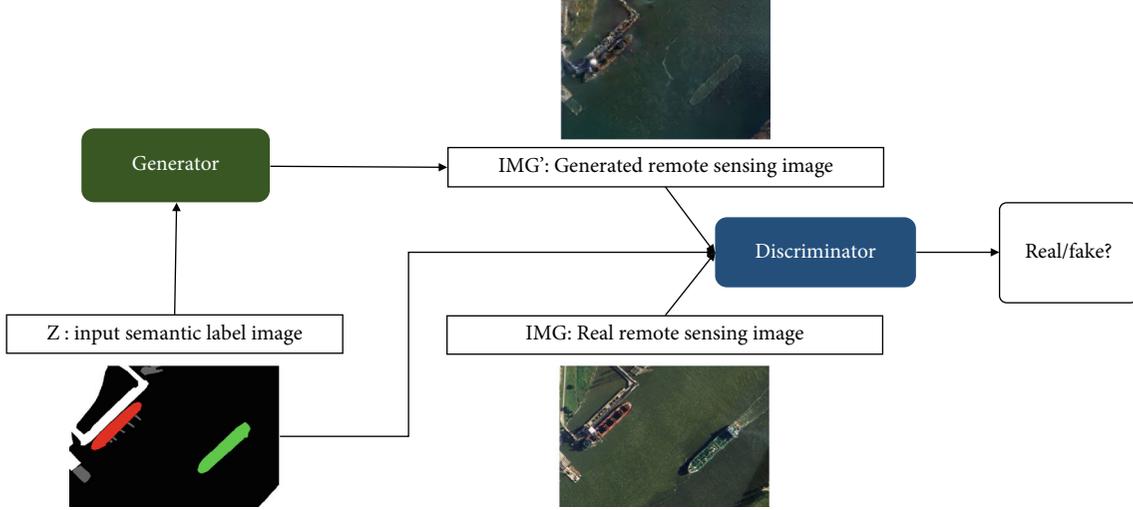


FIGURE 5: Model architecture of U-Net GAN-based translator.

representing the probability that output images x of the Generator came from the training set rather than p_g .

The loss function $L(G)$ of the Generator G could be mathematically defined as shown in the following equation.

$$L(G) = E_{z \sim p_z(z)} [\log (1 - D(z, G(z)))]. \quad (2)$$

The loss function $L(D)$ of the Discriminator D could be mathematically defined as shown in Equation (3).

$$L(D) = E_{x \sim p_{data}(x)} [\log D(z, x)] \quad (3)$$

Generally, the optimization objective of the translator could be defined as shown in the following equation:

$$\min_G \max_D V(D, G) = L(D) + L(G). \quad (4)$$

In order to let the generator better learn the details of target images, it is beneficial to integrate the traditional GANs' optimization objective with an extra loss such as the smooth $L1$ distance. The smooth $L1$ distance $L_{smooth L1}(G)$ is mathematically defined as shown in the following equation:

$$L_{smooth L1}(G) = \begin{cases} E_{z,x} [0.5 * (x - G(z))^2] & \|(x - G(z))\| < 1 \\ E_{z,x} [\|(x - G(z))\| - 0.5] & \text{otherwise.} \end{cases} \quad (5)$$

Meanwhile, the role of the Generator G has changed to not only fool the Discriminator D but also approach multiscale grains of the ground-truth images guided by the new optimization objective as defined by the following equation:

$$\min_G \max_D V(D, G) = L(D) + L(G) + L_{smooth L1}(G) \quad (6)$$

5. Evaluation

5.1. Experimental Settings. A public dataset called HRSC2016 [14] is adopted to evaluate the effectiveness of the proposed U-Net GAN-based approach to remote sensing image augmentation. In the dataset HRSC2016, all the images are collected from six famous harbors with the resolutions ranging from 0.4 m to 2 m. The image sizes vary from 300 to 1500 while most of them are larger than 1000×600 . The training set contains 436 images including 1207 samples, and the validation set contains 181 images including 541 samples, respectively. The test set contains 444 images including 1228 samples.

As for the baselines, we are going to compare our approach with two typical types of augmentation methods including the geometric transformation methods and generative models. Specifically, four types of transformations including Scaling, Flipping, CutOut, and CutMix will be tested for evaluation. Moreover, three generative models including WGAN [15], DCGAN, and CycleGAN [16] will also be evaluated under the same evaluation settings. The key hyperparameter settings of the baseline models and our proposed model are shown in Table 1.

All the experiments are conducted on a Windows 10 64-bit server equipped with one Intel Xeon CPU at 3.7 GHz and 64 GB main memory at 2666 MHz. All the generative models are trained on one NVIDIA GeForce RTX 2080Ti GPU of which the dedicated memory is 11 GB. And the deep learning framework to support the implementation and training of generative models is tensorflow 2.3 library in the Python 3.8 environment.

5.2. Experimental Results and Analysis. Firstly, we would like to compare the performance of geometric transformation methods with the approach proposed in this work. As shown in Figure 6, given a pair of the original image and its tag

TABLE 1: Key hyperparameter settings of baselines and our model.

Model	Key hyperparameter settings
DCGAN	Epochs = 300, batch size = 16, activation Func. = $\tan h$ Optimizer = Adam, loss Func. = binary cross entropy
WGAN	Epochs = 300, batch size = 16, activation Func. = $\tan h$ Optimizer = Adam, loss Func. = binary cross entropy
CycleGAN	Epochs = 300, batch size = 16, activation Func. = $\tan h$ Optimizer = Adam, loss Func. = binary cross entropy, lambda = 10
Ours	Epochs = 300, batch size = 16, activation Func. = $\tan h$ Optimizer = Adam, loss Func. = binary cross entropy, LAMBDA = 100

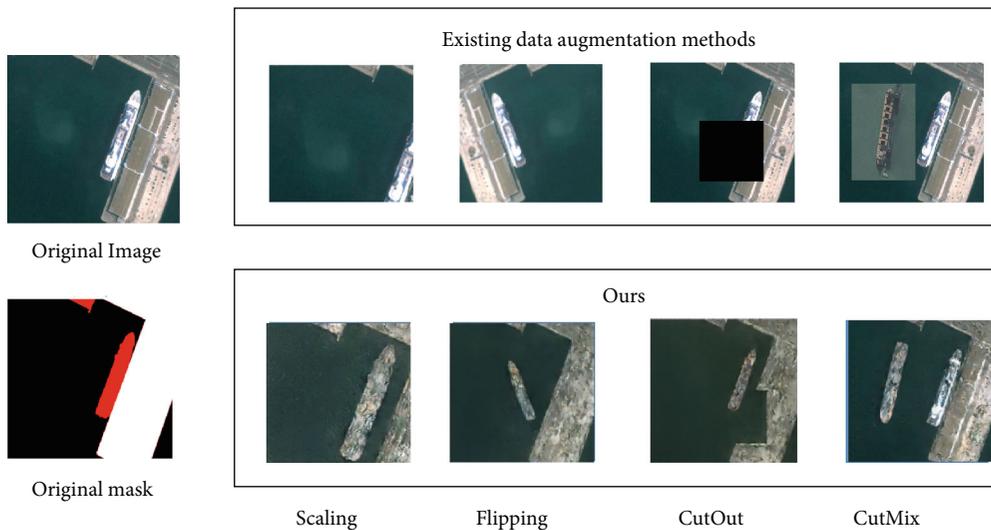


FIGURE 6: Comparison with existing geometric transformation methods.

image, the upper row lists the augmented images after Scaling, Flipping, CutOut, and CutMix by leveraging traditional geometric transformation methods while the lower row shows the augmented images generated by our proposed approach. In the “Scaling” case, the target object (i.e., the ship) is partially cut while it is simultaneously scaled with the whole image by our approach. In the “Flipping” case, our approach turns the vertical flipping as done by traditional geometric transformation methods. In the “CutOut” case, the augmented image has a very inharmonious black region which is smartly processed by our proposed approach. More interesting, our approach patches the black region with the water surface while maintains the integrity of the ship. At last, in the “CutMix” case, the traditional geometric transformation methods simply place a rectangle patch containing another ship with considering the semantic consistency between the patch and the original image. In contrast, our proposed approach places the newly added ship to the proper area of the original image with its background seamlessly wired. According to the above analyzed cases, it is clear that our proposed approach significantly outperform the traditional geometric transformation methods in terms of

maintaining object integrity, diversity, and background harmony for the augmented remote sensing images.

In another set of experiments, we compare the performance of different generative models including DCGAN, WCGAN, CycleGAN, and our proposed approach. As shown in the first two rows of Table 2, DCGAN and WGAN only accept random noise as the input and can hardly generate meaningful output images. The most competitive generative model is CycleGAN of which the generation results are shown in the third row of Table 2. Obviously, CycleGAN is able to generate the outline for each type of target objects (i.e., the water surface, port land, and ships). However, if we zoom in the images generated by CycleGAN, it is found that almost no detail of the target objects is captured. And our proposed approach does not only draw the outline of multiple objects but also capture their much more details in the generated images.

Furthermore, the detailed training process of each generative model is shown in Table 3. It is observed that DCGAN and WGAN are not able to generate images rather than random noise until Epoch 200. And even after Epoch 200, DCGAN and WGAN just capture some very vague features which cannot be clearly identified. CycleGAN and our

TABLE 2: Comparison with existing generative models.

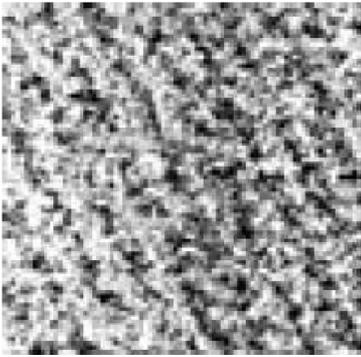
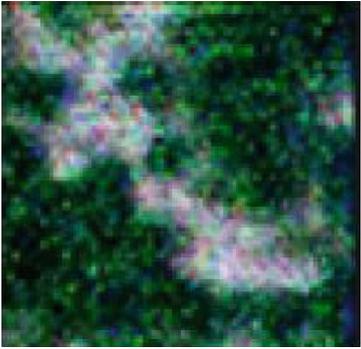
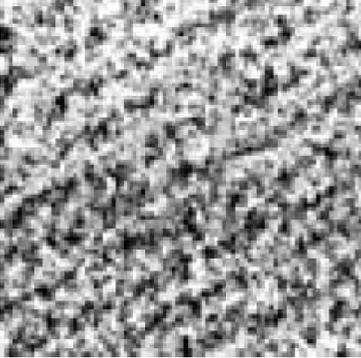
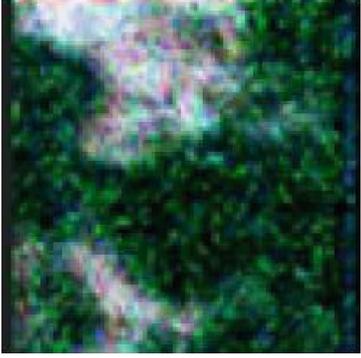
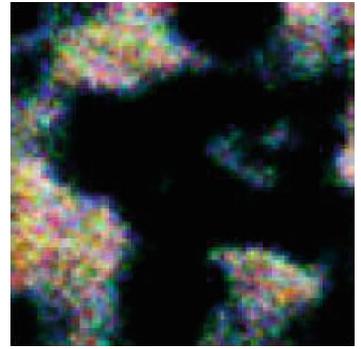
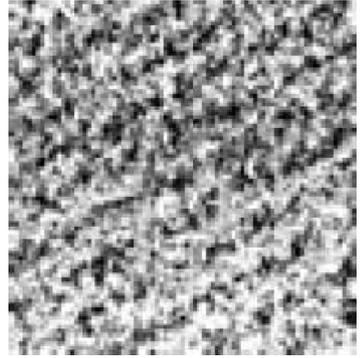
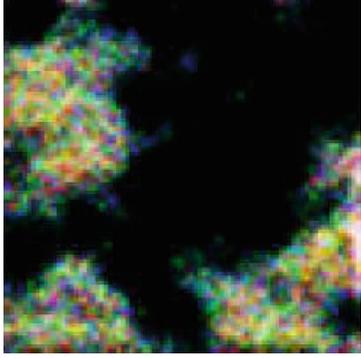
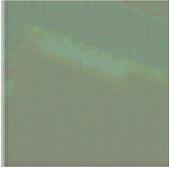
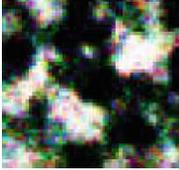
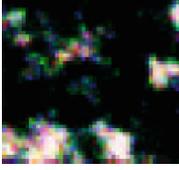
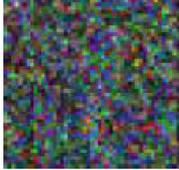
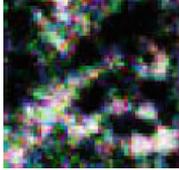
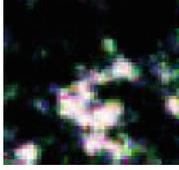
Model	Input1	Output1	Input2	Output2
DCGAN				
WGAN				
CycleGAN				

TABLE 2: Continued.

Model	Input1	Output1	Input2	Output2
Ours				

TABLE 3: Detailed training process of baselines and our model.

Model	0	5	10	15	20	100	200	300
Ours								
CycleGAN								
DCGAN								
WGAN								

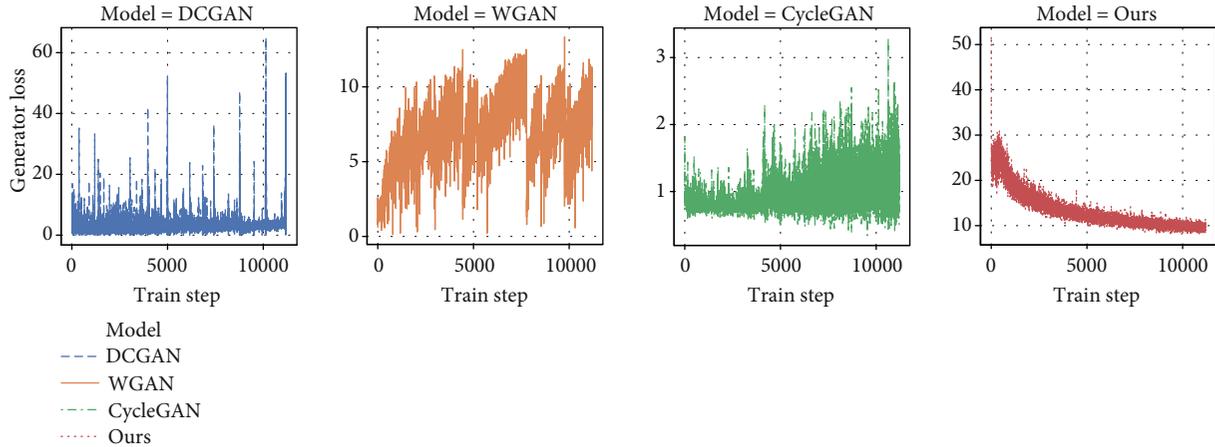


FIGURE 7: Generator loss of baselines and our model over train steps.

proposed approach start the training with the similar pure color input. Both models are able to capture the outline of each object as early as Epoch 5. But as the training epoch elapsed, CycleGAN is unable to capture more details about outlined objects while our proposed approach gradually adds more details to those objects. And finally, the output image generated by our proposed approach at Epoch 300 shows the highest visible similarity with the ground-truth image among all the generative models. According to the experimental results listed above, in summary, it is validated that our proposed approach to remote sensing image augmentation significantly outperforms the baselines including the traditional geometric transformation methods and generative models.

Last but not the least, the generator loss of all baselines and our model over each train step is shown in Figure 7 so that we can observe the learning behavior of the generator of all baseline models and our proposed model. And it can be clearly observed that the loss value of DCGAN's generator fluctuates from the very beginning of the training process till the last train step. For WGAN, the loss value of its generator becomes higher and higher over train steps. It indicates the fact that the generator of both DCGAN and WGAN can hardly converge on the experimental dataset, and thus, no meaningful image could be generated. As for CycleGAN, it performs better than DCGAN and WGAN by showing a converging trend during the training process. However, its loss value has a high deviation which probably means its generator cannot learn complex features from original images in a stable manner. At last, when we analyze the loss value of the generator of our proposed model, it presents a much better converging trend over train steps than baseline models. And this is a strong evidence to confirm the superiority of our proposed model over baselines in the task of object-level remote sensing image augmentation.

6. Related Work

Data augmentation is the technique to augment original training samples by generating new samples. The existing data augmentation techniques can be roughly divided into

the following two categories: (1) geometric transformation methods, which generate new samples by performing various geometric operations on original samples, and (2) generative models, which generate new samples by learning discriminative features of original samples and utilizing their labels.

Geometric transformation methods have been widely used, including random cropping, horizontal flipping, and color enhancement [17], which can improve the robustness of translation and reflection and illumination objects, respectively. Random scaling, random rotation, and affine transformation are also widely used in data augmentation scenarios [1]. Moreover, CutOut and CutMix [2] are also employed to augment new samples by learning features from original samples. In general, geometric transformation methods are usually applied to solve either the class imbalance problem or the limited sample problem. According to previous studies [1, 2, 17], the above-mentioned methods have been proved to be fast, reproducible, and reliable. And their implementation is relatively simple, which can be easily generalized to the currently popular deep learning framework. However, these methods can only perform image-level transformation, which means they only change the depth or scale of the image after generation. Although image-oriented tasks such as image classification benefit from geometric transformation methods, they are not capable of improving object-oriented tasks such as the semantic image segmentation.

Despite the many successes of generative adversarial network (GAN) and its numerous variants, there are still a lot of challenging issues such as mode collapse [8] and generation quality [18, 19]. Objectaug [20] is one kind of generative models for object-level data augmentation. It decomposes the image into separate objects and backgrounds using semantic tags and applies augmentation on background and objects individually. Objectaug can effectively enhance the boundaries between the target objects and the background. However, its core data augmentation method is still based on the traditional geometric transformation, which limits the diversity of generated samples. Conditional adversarial nets [6] are proposed to handle both unimodal and multimodal samples by extending the original GAN to its

conditional variant. As for other GAN-based models like DCGAN [4] and WGAN [15], they are not capable of generating new samples with visually similar features as those of the original samples due to the lack of properly guided input. Different from DCGAN and WGAN, CycleGAN [16] incorporates additional information with the original input which greatly enhances the quality of the generated samples. However, the main drawback of CycleGAN is its unpaired training process which limits its further performance improvement.

By taking drawbacks of the aforementioned data augmentation methods into accounts, in this paper, it motivates us to design and implement a new approach to augmenting existing dataset by generating diverse and high-quality samples at the object level.

7. Conclusion

In this paper, we study the object-level remote sensing image augmentation problem. In Section 3, the problem formulation is provided in a formal format to facilitate the understanding of the target problem. Then, an approach composed of the semantic tag image generator and the U-Net GAN-based translator is proposed in Section 4 to illustrate in detail how we can achieve object-level remote sensing image augmentation. To validate the effectiveness of the proposed approach, comprehensive experiments are conducted on a public dataset HRSC2016. With experimental results carefully examined and analyzed in Section 5.2, our proposed approach shows the promising performance by not only drawing the outline of different objects but also capturing their meaningful details.

Data Availability

The dataset used to support the evaluation of the proposed approach is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the Infrastructure Construction Project of Naval University of Engineering (No. HCX-MN-0600006).

References

- [1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, <https://arxiv.org/abs/1409.1556>.
- [2] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "Cut-Mix: regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6022–6031, Seoul, Korea (South), 2019.
- [3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.
- [4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, <https://arxiv.org/abs/1511.06434>.
- [5] D. Lin, K. Fu, Y. Wang, G. Xu, and X. Sun, "MARTA GANs: unsupervised representation learning for remote sensing image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2092–2096, 2017.
- [6] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, <https://arxiv.org/abs/1411.1784>.
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., pp. 234–241, Springer International Publishing, Cham, 2015.
- [8] W. Li, L. Fan, Z. Wang, C. Ma, and X. Cui, "Tackling mode collapse in multi-generator GANs with orthogonal vectors," *Pattern Recognition*, vol. 110, p. 107646, 2021.
- [9] W. Li, L. Xu, Z. Liang et al., "Sketch-then-edit generative adversarial network," *Knowledge-Based Systems*, vol. 203, p. 106102, 2020.
- [10] C. Ma, X. Shi, W. Zhu, W. Li, X. Cui, and H. Gui, "An approach to time series classification using binary distribution tree," in *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 399–404, Shenzhen, China, 2019.
- [11] W. Xu, C. Ma, S. Guo, and H. Zhou, "Efficient rate adaptation for 802.11af TVWS vehicular access via deep learning," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Waikoloa, HI, USA, 2019.
- [12] C. Ma, X. Shi, W. Li, and W. Zhu, "Edge4tsc: binary distribution tree-enabled time series classification in edge environment," *Sensors*, vol. 20, no. 7, p. 1908, 2020.
- [13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, MA, USA, 2015.
- [14] Z. Liu, L. Yuan, L. Weng, and Y. Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pp. 324–331, Porto, Portugal, 2017.
- [15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *34th International Conference on Machine Learning (ICML)*, pp. 214–223, Sydney, Australia, 2017.
- [16] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [18] W. Li, W. Ding, R. Sadasivam, X. Cui, and P. Chen, "His-GAN: a histogram-based GAN model to improve data generation quality," *Neural Networks*, vol. 119, pp. 31–45, 2019.
- [19] W. Li, Z. Liang, P. Ma, R. Wang, X. Cui, and P. Chen, "Hausdorff GAN: Improving GAN generation quality with Hausdorff metric," *IEEE Transactions on Cybernetics*, pp. 1–13, 2021.
- [20] J. Zhang, Y. Zhang, and X. Xu, "Objectaug: object-level data augmentation for semantic image segmentation," 2021, <https://arxiv.org/abs/2102.00221>.