


## Article

# Using Vehicle Synthesis Generative Adversarial Networks to Improve Vehicle Detection in Remote Sensing Images

Kun Zheng , Mengfei Wei, Guangmin Sun, Bilal Anas and Yu Li \*Faculty of Information Technology, Beijing University of Technology, No.100, Pingleyuan Road,  
Beijing 100124, China

\* Correspondence: yuli@bjut.edu.cn; Tel.: +86-67391526

Received: 28 July 2019; Accepted: 29 August 2019; Published: 4 September 2019



**Abstract:** Vehicle detection based on very high-resolution (VHR) remote sensing images is beneficial in many fields such as military surveillance, traffic control, and social/economic studies. However, intricate details about the vehicle and the surrounding background provided by VHR images require sophisticated analysis based on massive data samples, though the number of reliable labeled training data is limited. In practice, data augmentation is often leveraged to solve this conflict. The traditional data augmentation strategy uses a combination of rotation, scaling, and flipping transformations, etc., and has limited capabilities in capturing the essence of feature distribution and proving data diversity. In this study, we propose a learning method named Vehicle Synthesis Generative Adversarial Networks (VS-GANs) to generate annotated vehicles from remote sensing images. The proposed framework has one generator and two discriminators, which try to synthesize realistic vehicles and learn the background context simultaneously. The method can quickly generate high-quality annotated vehicle data samples and greatly helps in the training of vehicle detectors. Experimental results show that the proposed framework can synthesize vehicles and their background images with variations and different levels of details. Compared with traditional data augmentation methods, the proposed method significantly improves the generalization capability of vehicle detectors. Finally, the contribution of VS-GANs to vehicle detection in VHR remote sensing images was proved in experiments conducted on UCAS-AOD and NWPU VHR-10 datasets using up-to-date target detection frameworks.

**Keywords:** vehicle detection; remote sensing; deep learning; generative adversarial network; data augmentation

## 1. Introduction

Fast and robust vehicle detection in remote sensing images has potential applications in traffic surveillance, emergency management, and economic analysis. Moreover, the location and density information of vehicles is an essential data source for building intelligent transportation systems. However, accurate and robust vehicle detection from remote sensing images has been a challenging task for many years.

Traditional vehicle detection methods rely on handcrafted features extracted from sliding windows with different scales. For instance, Shao et al. (2012) [1] extracted Harr features and local binary patterns from images and classified vehicles with a support vector machine. Kluckner et al. (2007) [2] and Thuckner et al. (2013) [3] applied the histogram of oriented gradients and integral channel features to train the AdaBoost classifier. However, these methods are highly reliant on manually designed features and they cannot adequately cope with large variations in targets and backgrounds. Recently,

the convolutional neural network (CNN) has been applied to aerial image object detection, and it achieved promising results. For instance, Faster R-CNN (Region-Convolutional Neural Networks) [4,5] and You Only Look Twice (YOLT) [6] have been applied to various benchmarks.

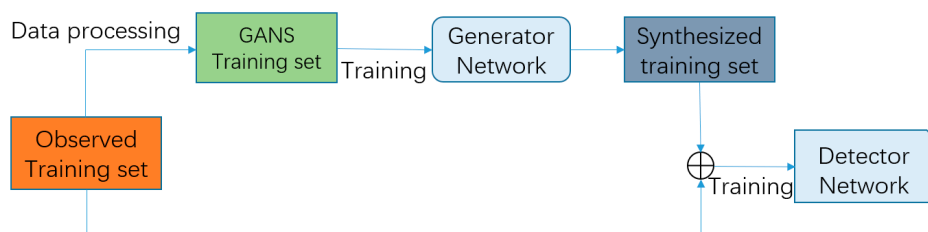
Many approaches have been proposed for vehicle detection [7–10], yet the robustness problem remains unsolved. Based on CNN models, recent works [11–13] have achieved good detection performances on several benchmarks. Hu et al. (2019) [14] proposed a scale-insensitive convolutional neural network that could detect vehicles with a large variance of scales. Gao et al. [15] presented an end-to-end vehicle detection model that unifies and combines the deep and shallow feature maps, followed with deformable convolution and region of interest (RoI) pooling. It achieved good performance on vehicle target detection in dense areas.

Built on a massive amount of training datasets, these models can achieve significant performance improvement over previous baselines. However, the performance of CNN-based vehicle detectors is heavily dependent on the quality and quantity of annotations of the training data. The imaging configuration of remotes sensing data is variant (incidence angle, distance, sensor type, etc.), which leads to a significant difference between targets in remote sensing images and those from other images. Meanwhile, labeling ground truth bounding boxes for vehicle locations in aerial images requires considerable human efforts. Therefore, it is extremely crucial to design approaches that only rely on limited labeled images.

To alleviate the demand of massive sample annotation in the deep learning era, weakly supervised methods were proposed. Li et al. proposed a weakly supervised deep learning method that only used scene-level tags, which took advantage of both the information from separate scene categories and mutual cues between scene pairs [16]. Zhao et al. proposed a multiscale image block-level fully convolutional neural network that only requires class labels without bounding boxes [17]. However, these methods still require a large number of image samples.

Augmenting the labeled dataset is a more direct and practical approach. There are some common data augmentation methods that can be used to extend datasets over time, such as flip, rotation, color jittering, random crop, etc. Montserrat et al. [18] applied some regular data augmentation methods where linear and nonlinear transforms were done on the training data to create “new” training images for object detection. Oliveira et al. [19] generated smooth face image variations to improve age estimation based on the detection of fiducial points on the face. However, all of the above methods are restricted to very straightforward transformations or are based heavily on simple image transformations and have a limited performance on improving the robustness of the model due to the restricted diversity of samples.

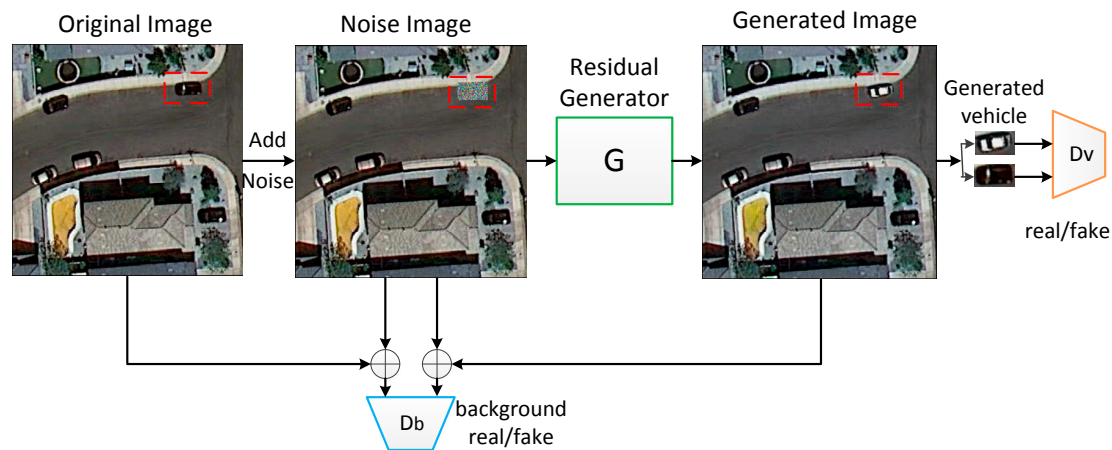
Based on Generative Adversarial Nets (GANs) [20] proposed by Goodfellow et al., labeled data samples could be generated with larger diversity while maintaining the core characteristics of the original data samples [21]. The overall flowchart of data augmentation with GANs is shown in Figure 1. In a GAN, the generator and classifier networks are jointly learned based on the observed training dataset, and the observed and synthesized data samples are combined to train the detector network.



**Figure 1.** Flowchart of Generative Adversarial Nets (GANs)-based data augmentation.

In this study, a data augmentation framework called Vehicle Synthesis Generative Adversarial Networks (VS-GANs) is proposed, which utilizes GANs to generate labeled data samples for the vehicle detection task (Figure 2). The highlights of this study include the following:

- Our proposed model is capable of generating clear and photorealistic vehicle images in remote sensing images and can fit the background well in real images.
- The data generated by our model can be combined with real datasets to train CNN-based detectors. This data augmentation step can improve both detection performance and robustness compared with the baseline method.
- The proposed method is convenient to apply within the training process of the CNN-based detector.



**Figure 2.** The key framework of using VS-GANs for vehicle generation. It learns to smoothly synthesize vehicles in background images through the multiple discriminators ( $D_b$  and  $D_v$ ) network.

## 2. Vehicle Generation Model

### 2.1. Generative Adversarial Networks

GANs have been widely used to improve the training stability and the quality of data generation [22–25]. It has also served in many other applications, such as image translation [26–30], super resolution [31], and image inpainting [32–34]. Liu et al. [35] proposed a DP-GAN (Domain Priori GAN) for obtaining super-resolution vehicle license plate images, which achieved pixel-level detail recovery.

Some researchers have also applied GANs for object detection. For example, Inoue et al. [36] proposed an object detection method that used labeled datasets (such as PASCAL VOC) to transfer training unlabeled datasets (such as cartoon datasets) based on cycle-GAN. Li et al. [37] proposed a Perceptual Generative Adversarial Network (Perceptual GAN) model that improved small object detection through narrowing the representation difference between small objects and large ones. Zhu et al. [26] added cycle consistency loss to the original GAN, enabling the model to conduct translations without paired training examples and task-specific designed functions. To synthesize the vehicles in the noise boxes, we adopted the training method of cycle-GAN and proposed a relatively simple architecture with one generator and two discriminators.

Compared with image inpainting works [32–34], which aim to fill the randomly removed monochromatic patches in the original image, our framework fills the bounding box area with noise rather than monochromatic blocks to generate patches with diverse shapes and colors. The work by Liu et al. [35] exploited a similar GAN with two discriminators for image inpainting to learn more context information of the surrounding pixels. In contrast, we passed the vehicle patch cropped from the generated output into the discriminator and used it in the model to generate vehicles with different appearances.

## 2.2. Vehicle Synthesis-GAN

The vehicle generation model in this paper is inspired by the structure of the original GANs. In the training stage, as shown in Figure 2, we cover the vehicle in the ground truth image with random noise and then send the noise-added image into the generator. There are two adversarial losses between the generator and both discriminators:  $D_v$  for discrimination vehicles and  $D_b$  for background learning.

### 2.2.1. The Structure of the VS-GANs Model

#### (a) Structure of the generator

The purpose of the generator is to learn a function mapping relation  $F: x \rightarrow y$ , where  $x$  is the input noise image and  $y$  is the ground truth image. As shown in Figure 3, the encoder-decoder structure in deep learning is adopted in the experiment. In order to alleviate the gradient disappearance caused by the design of a too deep model and increase the semantic information of deep features, the basic residual block structure [38] is added to the generator structure. When the noise image  $x$  is fed into the generator, a series of convolution layers is used to reduce the size and increase the depth of the output feature map. This process captures the deep features of the image. These features are extracted through a set of residual modules to extract more abundant semantic information. Then the final feature map is extracted through the decoder composed of deconvolution structures.

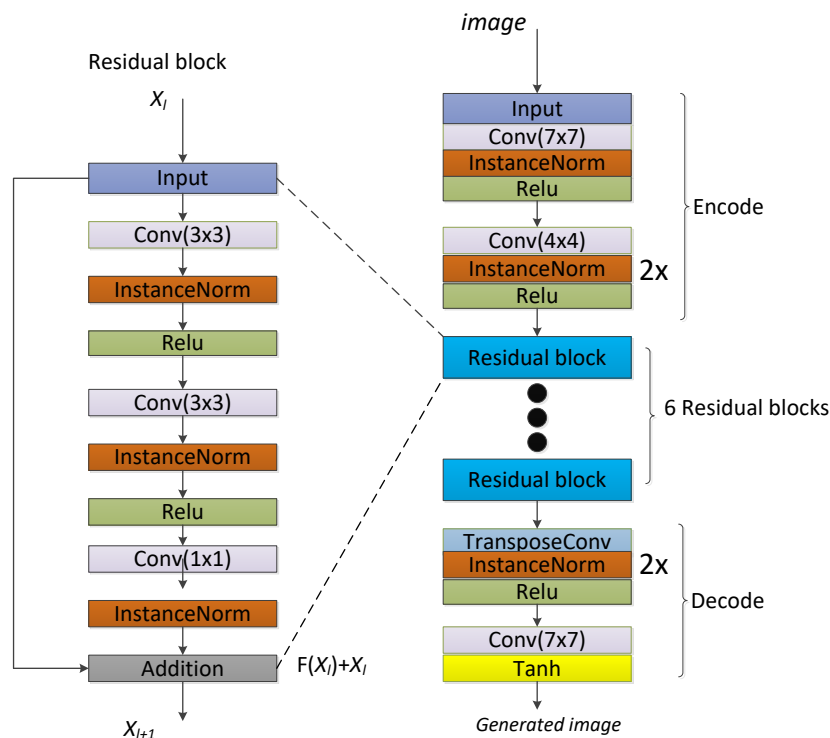
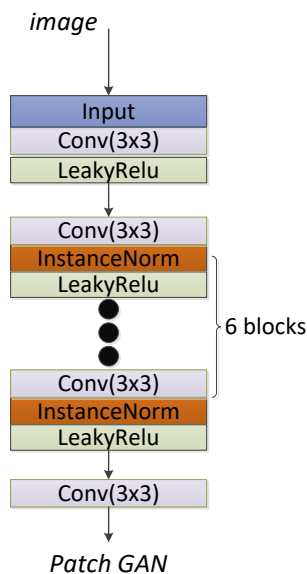


Figure 3. Structure of the VS-GANs generator.

#### (b) Structure of the background discriminator

Discriminator  $D_b$  is used to learn background context information. The task for our model is not only to generate a vehicle sample but also to integrate the generated vehicle sample into the background image smoothly. Therefore, the model is required to learn the background information of the image, such as the surrounding environment and lighting conditions. In this work, we use  $D_b$  to distinguish real background pairs from generated background pairs. The positive pairs are the noise image  $x$  and the ground truth image  $y$ , and the generated pairs are the concatenation of the noise image  $x$  and the generated image. As shown in Figure 4, the structure of  $D_b$  follows the design

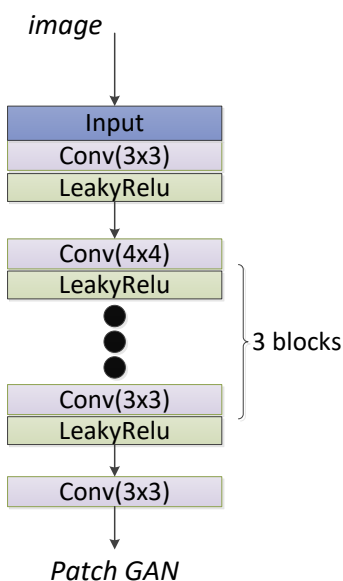
of the DCGAN(deep convolutional GAN) [25], but the following adjustments are made: (1) the first convolution structure is adjusted so that it can accept six input channels, (2)  $D_b$  has a similar structure with PatchGAN [28], in which  $D_b$  is used to distinguish the true and false parts of each area of  $N \times N$  pixels (in this experiment,  $N$  is 19), and (3) in order to adapt to the design pattern of the PatchGAN and to generate better quality images, the least-squares loss function is selected as the loss function of the discriminator  $D_b$ .



**Figure 4.** Structure of the VS-GANs background discriminator.

(c) Structure of the vehicle discriminator

The discriminator  $D_v$  is used to distinguish the positive vehicles from the negative vehicles. The positive vehicles are real vehicles in the ground truth image, and the negative vehicles are fake vehicles generated by the generator at the location of the noise box. The discriminator  $D_v$  can drive the generator to learn the mapping relationship between noise  $z$  and real vehicles. The structure of the discriminator  $D_v$  is shown in Figure 5. The main structure uses a six-layer convolutional neural network followed by a LeakyReLU layer.



**Figure 5.** Structure of the VS-GANs vehicle discriminator.

### 2.2.2. Loss Function of VS-GANs

As shown in Figure 2, our model consists of two adversarial learning procedures,  $G \leftrightarrow D_b$  and  $G \leftrightarrow D_v$ . The adversarial learning procedure between  $G$  and  $D_b$  can be formulated as follows:

$$\mathcal{L}(G, D_b) = E_{y \sim p_{gt.image}(y)} [(D_b(y) - 1)^2] + E_{x, z \sim p_{noise.image}(x, z)} [(D_b(G(x, z)))^2] \quad (1)$$

where  $x$  is the image with the noise box and  $y$  is the ground truth image. We replace the original GAN loss by the least-squares loss that was used in LSGAN (Least squares GAN) [22].

In order to generate a realistic vehicle using  $G$  in the noise box  $z$  of input image  $x$ , another adversarial training process is carried out between  $G$  and  $D_v$ :

$$\begin{aligned} \mathcal{L}(G, D_v) = & -E_{y_v \sim p_{vehicle}(y_v)} [D_v(y_v)] + E_{z \sim p_{noise}(z)} [D_v(G(z))] \\ & + \lambda E_{z \sim p_{noise}(z)} [(\|\nabla_z D_v(z)\|_2 - 1)^2] \end{aligned} \quad (2)$$

where  $z$  is the noise box in image  $x$ . And  $y_v$  is the cropped vehicle from ground truth image  $y$ , and the gradient penalty (GP) strategy used in WGAN (Wasserstein GAN) [39] is taken to balance the training procedure.

In this paper, we use  $\ell_1$  loss to balance the difference between the synthesized image and the ground truth image:

$$\mathcal{L}_{\ell_1}(G) = E_{x, z \sim p_{noise.image}(x, z), y \sim p_{gt.image}(y)} [\|y - G(x, z)\|_1]. \quad (3)$$

The final loss function is calculated by adding the previously defined losses. The  $\lambda$  is a hyper-parameter to control  $\ell_1$  loss:

$$\mathcal{L}(G, D_b, D_v) = \mathcal{L}(G, D_b) + \mathcal{L}(G, D_v) + \lambda \mathcal{L}_{\ell_1}(G). \quad (4)$$

## 3. Experimental Results

The experiments mainly include two parts: In the first, the generated vehicle samples with VS-GANs were compared with those derived from other data augmentation methods. In the second, vehicle detection was conducted in order to test the contribution of the proposed method.

UCAS-AOD data [40] were used to train the VS-GANs, and the quality of the synthesized images was evaluated. Vehicles generated with VS-GANs were compared with those generated by traditional data augmentation methods and other GANs. To analyze the influence of the data augmentation, we used both the real and synthesized data to train the YOLOv3 [41] and RetinaNet [42] detectors and evaluated the performance. Vehicles generated at both “proper” and “random” locations were considered to examine the effect of vehicle location on the training of the detector. The experiments were based on the PyTorch platform and accelerated with Tesla M40 Graphics Processing Units (GPUs).

### 3.1. Datasets

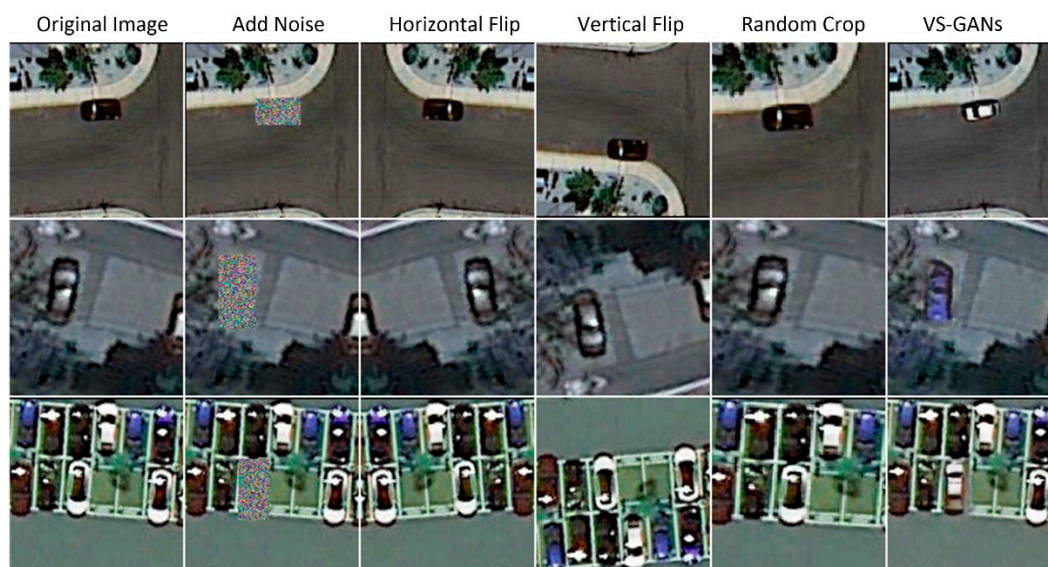
UCAS-AOD [40] is a dataset for object detection in remote sensing images and contains 510 images with 7114 vehicle samples. All images were collected from Google Earth and manually labeled. Compared to the other benchmark DOTA [43], USAC-AOD has higher resolution and a relatively similar scale and ratio, which is more suitable to train GANs.

### 3.2. Qualitative Analysis of the Generated Vehicle Samples

In Figure 6, we compare VS-GANs with three other basic data augmentation methods frequently used in the training procedure of CNN-based detectors. It intuitively shows the difference between VS-GANs and other data augmentation methods. The original images are in the first column, and the second column shows the noise boxes added to the location where the vehicle is about to be



generated. The third column to fifth columns show the results derived by some basic data augmentation methods. We can see that the appearance of the vehicles is not changed and that they only have a linear transformation. The last column shows that the VS-GANs-based data augmentation resulted in a significant change in appearance. Moreover, the new vehicles generated by VS-GANs in the images share the characteristics of the original vehicles.



**Figure 6.** Comparison of VS-GANs with several data augmentation methods.

We generated noise within the bounding boxes in images from the UCAS-AOD and NWPU VHR-10 datasets. There were some labeled vehicles that are too small or partially blocked by other objects. Therefore, only those vehicles with a width between 20 and 70 pixels and a length–width ratio between 1.3 and 2.2 were selected. Then, we cropped the  $300 \times 300$  patches around the chosen vehicles from the original images. We obtained 3409 images and randomly selected 10% of them as the testing dataset, and used the others as the training data for VS-GANs.

We added the pre-trained generator to the CNN-based training procedure to validate the effectiveness of VS-GANs. Next,  $300 \times 300$  patches around the vehicles were cropped from the original images with a certain probability, and noise boxes were filled to cover the real vehicles in those patches. Our well-trained generator could synthesize vehicles within those noise boxes.

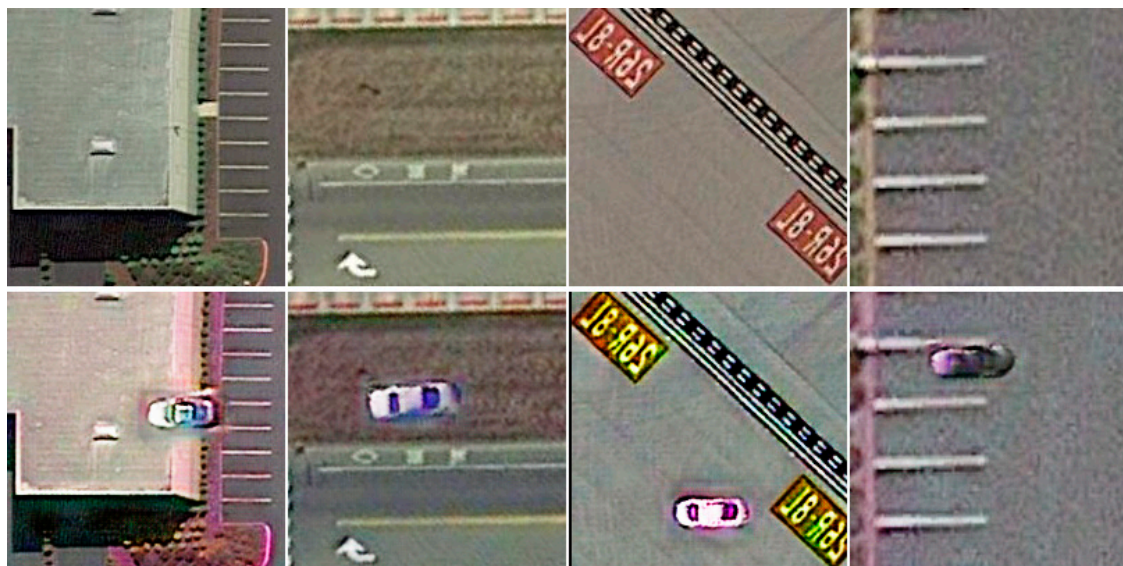
Several commonly used GANs were compared to our proposed method. In our experiment, we found that using the least-squares GAN [22] for  $D_b$  is useful for learning the background context. VS-GANs achieved the best picture quality when applying the least-squares loss for the adversarial learning  $G \leftrightarrow D_b$  and keeping the GAN loss stable for  $G \leftrightarrow D_v$ . The reasons why the two discriminators  $D_b$  and  $D_v$  led to different optimal GAN losses in our network were as follows. (1) For  $D_v$ , since we applied the PatchGAN trick and the gradient penalty strategy used by WGAN-GP [39], the least-squares GAN with least-squares loss obtained a larger error than the regular GAN loss, which made the generator more sensitive to every pixel as compared to that in the regular GAN. Thus, the generator was forced to learn too much detailed information about vehicles instead of capturing their more general characteristics. (2) The discriminator  $D_b$  can benefit from the least-squares loss when learning the background context information. We expect the generator to smoothly learn the background information from the vehicles' surrounding pixels.

The generated and real vehicles are shown in Figure 7. We can see that VS-GANs can generate vehicles with a clear shape and a more photorealistic appearance compared with some existing GAN methods. It should be noted that to achieve such a result we only used 8850 vehicles in 3409 images. Generating samples only at the position of existing vehicles may limit the variety of samples; therefore,

we also generated vehicles at random positions. As shown in Figure 8, some of the generated vehicles are at unrealistic positions.



**Figure 7.** Comparisons between VS-GANs and other GAN methods in vehicle generation.



**Figure 8.** The synthesized vehicles in the negative images. The original images are on the top, and the corresponding synthesized images are on the bottom.

### 3.3. Vehicle Detection Experiments

Data augmentation based on VS-GANs was applied to train the vehicle detector with the framework of YOLOv3 and RetinaNet on datasets UCAS-AOD and NWPU VHR-10. In the first experiment, we randomly put noise boxes at the location of the vehicle to generate diverse vehicles on the training dataset, as shown in Figure 2. The experimental datasets were cropped from original UCAS-AOD data and NWPU VHR-10 data to a size of  $416 \times 416$  pixels. In total, there were 3083 training images and 326 testing images. First, YOLOv3 and RetinaNet were used to train the detector based on the training dataset. Then, the data samples generated by VS-GANs were applied. All the derived detectors were tested on 326 testing images and the average precision (AP) was calculated when all the models converged. As shown in Table 1, although YOLOv3 and RetinaNet achieved a relatively good performance (92.91% and 84.28%, respectively), combining this data augmentation with the vehicle detector significantly improved the vehicle detector performances (to 96.12% and 90.78%, respectively).



**Table 1.** Experimental results of data augmentation based on VS-GANs.

Source of Vehicle Samples	YOLOv3	RetinaNet
UCAS + NWPU (8850 vehicles)	92.91%	84.28%
+VS-GANs augmentation	96.12%	90.78%

Examples of synthetic vehicles generated by the existing GANs and the proposed VS-GANs are shown in Figure 9. It can be observed that vehicles generated by VS-GANs can be smoothly and realistically integrated into the background of the original image. To prove how the synthetic images can help boost the performance of vehicle detection between the existing GANs and VS-GANs, we used the vehicle samples to train the YOLOv3 detector. The accuracy was derived based on 5-fold cross-validation. As shown in Table 2, adding more generated data samples achieved higher detection accuracy, and the proposed VS-GANs were shown to outperform other vehicle-generating algorithms. The standard deviation (listed in the bracket) of 5-fold cross-validation is relatively small. Hence, the analysis result is comparable and can be referenced.

**Figure 9.** Examples of synthetic vehicles generated by the existing GANs and VS-GANs.

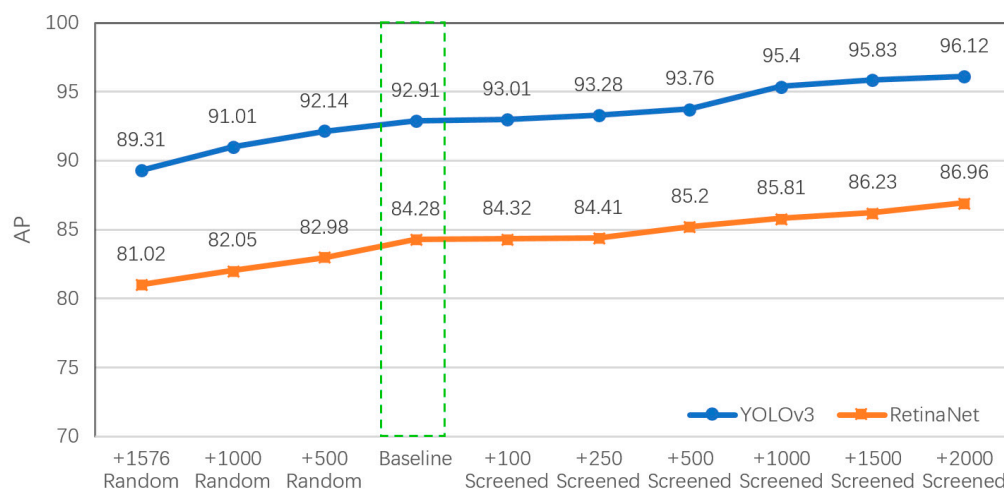
**Table 2.** Performance of vehicle detector based on the existing GANs and VS-GANs.

Source of Vehicle Samples	DCGAN	LSGAN	WGAN-GP	VS-GANs
UCAS + NWPU (8850 vehicles)	92.91% (0.66) *			
+125 images (1000 synthesized vehicles)	93.84% (0.46)	93.65% (0.65)	94.16% (0.34)	95.40% (0.33)
+250 images (2000 synthesized vehicles)	94.17% (0.41)	94.77% (0.38)	95.44% (0.18)	96.12% (0.21)

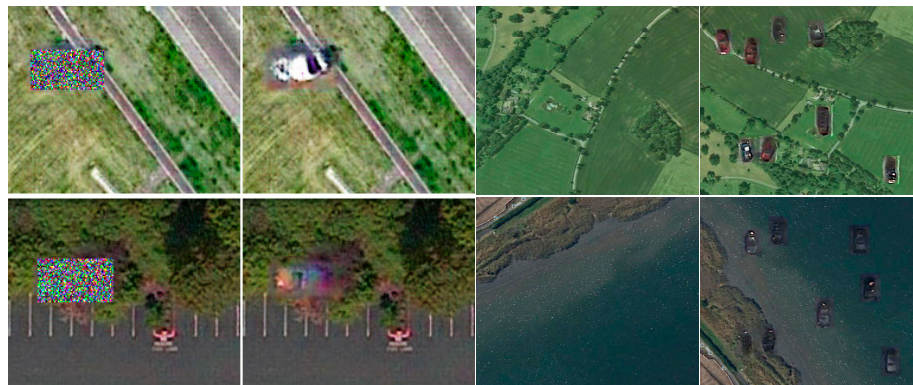
\* The mean value of 5-fold cross-validation is listed, followed by standard deviation in the brackets.

The locations of noise boxes were chosen to be where vehicles were already present in the first experiment. In the second experiment, we also selected some images without vehicles from the UCAS-AUO dataset and randomly put noise boxes into these images (as shown in Figure 6). It can be observed that if the noise boxes are located at the wrong locations, the generator may fail to generate proper vehicle samples due to an unrealistic background, as shown in Figure 9.

To give more insight on the effect of the number of samples, synthesized vehicles were gradually added into training samples of VS-GANs. As shown in Figure 10, the baseline APs for YOLOv3 and RetinaNet were 92.91% and 84.28%, respectively (highlighted in the green box). With adding the screened vehicle samples from 100 to 2000, the performance gradually increased to 96.12% and 86.96% for YOLOv3 and RetinaNet, respectively. However, when adding random vehicle samples, the performance of YOLOv3 and RetinaNet dropped gradually to 89.31% and 81.02%. Moreover, by adding 2000 augmented data samples, the growth tendency of the AP was not yet stopped, which suggests that the performance can further increase. Hence, a higher upper-bound of the proposed method could be expected.

**Figure 10.** Performance of detectors by adding different numbers of VS-GAN-generated samples.

It was clearly observed that adding samples at proper positions can improve the vehicle detection results, while adding samples at improper positions has an adverse effect on the vehicle detection performance. This is because there may be some water or forest areas in the images, which could lead to unsuccessful vehicle generation, as shown in Figure 11. Unsuccessful samples from the generator have large differences compared with real vehicles, which is not conducive for vehicle detector training.



**Figure 11.** The synthesized vehicles in the negative images. There is a poor generation result if the noise boxes are placed at the wrong locations.

In the experiments of this paper, most of the trained, detected, and generated vehicles were small vehicles such as cars, SUVs, and minivans due to their abundance in data samples. Given sufficient samples, the proposed algorithm could also work on buses, trucks, and other types of vehicles. Figure 12 shows different types of vehicles detected by the proposed method.



**Figure 12.** Examples of different types of vehicles detected by the proposed method.

#### 4. Conclusions

This study proposed a learning method called VS-GANs, which can quickly generate high-quality annotated vehicles from remote sensing data with limited annotated vehicle samples. The synthesized vehicles can benefit the training of CNN-based vehicle detectors. The VS-GANs model can also help boost the performance of vehicle detection if the generated vehicles are placed at the right locations of images without vehicles, which demonstrates its ability to transfer knowledge. The proposed model could be useful in other related tasks in which only limited datasets are available. The key findings of this study can be summarized as follows:

- (1) The vehicle samples generated by VS-GANs hold characteristics similar to those of real vehicle samples while providing diverse details.
- (2) Adding synthesized samples in the training process significantly improves the performance of the vehicle detector.
- (3) The location of the generated vehicle sample has to be appropriately selected to avoid introducing an unrealistic background.

The method proposed in this study has great potential to improve the performance of vehicle detectors. The generated data samples have been proved to significantly enlarge the variety of data samples and improve the performance of target detection algorithms. The discriminator trained by VS-GANs can be applied to distinguish vehicles in backgrounds. Moreover, the technique proposed in this study could be used in other target detection/identification tasks in which only limited datasets are available.



**Author Contributions:** Conceptualization, Kun Zheng, Mengfei Wei and Guangmin Sun; Methodology, Kun Zheng, Mengfei Wei and Yu Li; Validation, Mengfei Wei and Yu Li; Formal Analysis, Mengfei Wei; Investigation, Kun Zheng, Mengfei Wei and Yu Li; Writing-Original Draft Preparation, Kun Zheng and Mengfei Wei; Writing-Review & Editing, Anas Bilal and Yu Li; Visualization, Mengfei Wei; Supervision, Guangmin Sun and Yu Li; Funding Acquisition, Kun Zheng.

**Funding:** This research was partially funded by the National Key Research and Development Program of China (Grant No. 2016YFB0501501), the Natural Science Foundation of China (Grant No. 41706201), and the International Research Cooperation Seed Fund of Beijing University of Technology (Grant No. 2018-B1).

**Acknowledgments:** The authors would like to thank anonymous reviewers for their constructive comments and suggestions, which greatly improved the quality of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Shao, W.; Yang, W.; Liu, G.; Liu, J. Car detection from high-resolution aerial imagery using multiple features. In Proceedings of the Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 4379–4382. [CrossRef]
2. Kluckner, S.; Pacher, G.; Grabner, H.; Bischof, H.; Bauer, J. A 3D teacher for car detection in aerial images. In Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8. [CrossRef]
3. Tuermer, S.; Kurz, F.; Reinartz, P.; Stilla, U. Airborne vehicle detection in dense urban areas using HoG features and disparity maps. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2327–2337. [CrossRef]
4. Tianyu, T.; Shilin, Z.; Zhipeng, D.; Huanxin, Z.; Lin, L. Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining. *Sensors* **2017**, *17*, 336. [CrossRef]
5. Yang, M.Y.; Liao, W.; Li, X.; Rosenhahn, B. Deep Learning for Vehicle Detection in Aerial Images. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018.
6. Van Etten, A. You Only Look Twice: Rapid Multi-Scale Object Detection in Satellite Imagery. Available online: <https://arxiv.org/abs/1805.09512> (accessed on 20 August 2019).
7. Zhao, T.; Nevatia, R. Car detection in low resolution aerial images. *Image Vis. Comput.* **2003**, *21*, 693–703. [CrossRef]
8. Line, E.; Lars, A.; Hans, K. Classification-based vehicle detection in high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 65–72. [CrossRef]
9. Zheng, H.; Pan, L.; Li, L. A morphological neural network approach for vehicle detection from high resolution satellite imagery. In Proceedings of the International Conference on Neural Information Processing, Hong Kong, China, 3–6 October 2006. [CrossRef]
10. Qu, T.; Zhang, Q.; Sun, S. Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks. *Multimed. Tools Appl.* **2016**, *76*, 21651–21663. [CrossRef]
11. Li, H.; Fu, K.; Yan, M.; Sun, X.; Sun, H.; Diao, W. Vehicle detection in remote sensing images using denoising-based convolutional neural networks. *Remote Sens. Lett.* **2017**, *8*, 262–270. [CrossRef]
12. Chauhan, R.; Ghanshala, K.K.; Joshi, R.C. Convolutional Neural Network (CNN) for Image Detection and Recognition. In Proceedings of the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 15–17 December 2018.
13. Zhai, S.; Cheng, Y.; Feris, R.; Zhang, Z. Generative Adversarial Networks as Variational Training of Energy Based Models. Available online: <https://arxiv.org/abs/1611.01799> (accessed on 20 August 2019).
14. Hu, X.; Xu, X.; Xiao, Y.; Chen, H.; He, S.; Qin, J.; Heng, P.-A. SINet: A Scale-Insensitive Convolutional Neural Network for Fast Vehicle Detection. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 1010–1019. [CrossRef]
15. Gao, X.; Li, H.; Zhang, Y.; Yuan, M.; Zhang, Z.; Sun, X.; Sun, H.; Yu, H. Vehicle Detection in Remote Sensing Images of Dense Areas Based on Deformable Convolution Neural Network. *J. Electron. Inf. Technol.* **2018**, *40*, 2812–2819.
16. Zhao, W.; Ma, W.; Jiao, L.; Chen, P.; Yang, S.; Hou, B. Multi-Scale Image Block-Level F-CNN for Remote Sensing Images Object Detection. *IEEE Access* **2019**, *7*, 43607–43621. [CrossRef]



17. Li, Y.; Zhang, Y.; Huang, X.; Yuille, A.L. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 182–196. [CrossRef]
18. Montserrat, D.M.; Lin, Q.; Allebach, J.; Delp, E. Training object detection and recognition cnn models using data augmentation. *Electron. Imaging* **2017**, 27–36. [CrossRef]
19. Oliveira, Í.; Medeiros, J.; de Sousa, V. A Data Augmentation Methodology to Improve Age Estimation using Convolutional Neural Networks. In Proceedings of the SIBGRAPI—Conference on Graphics Patterns and Images, Sao Paulo, Brazil, 4–7 October 2016. [CrossRef]
20. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Bing, X.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the International Conference on Neural Information Processing Systems, Kuching, Malaysia, 3–6 November 2014.
21. Toan, T.; Pham, T.; Carneiro, G.; Palmer, L.; Reid, L. A Bayesian Data Augmentation Approach for Learning Deep Model. Available online: <https://arxiv.org/pdf/1710.10564> (accessed on 20 August 2019).
22. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.K.; Wang, Z. Least squares generative adversarial networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
23. Denton, E.; Chintala, S.; Szlam, A.; Fergus, R. Deep Generative Image Models Using a Laplacian Pyramid of Adversarial Networks. Available online: <https://arxiv.org/abs/1506.05751> (accessed on 20 August 2019).
24. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. Available online: <https://arxiv.org/abs/1701.07875> (accessed on 20 August 2019).
25. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Available online: <https://arxiv.org/abs/1511.06434> (accessed on 20 August 2019).
26. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. Available online: <https://arxiv.org/abs/1703.10593> (accessed on 20 August 2019).
27. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
28. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
29. Taigman, Y.; Polyak, A.; Wolf, L. Unsupervised Cross-Domain Image Generation. Available online: <https://arxiv.org/abs/1611.02200> (accessed on 20 August 2019).
30. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
31. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
32. Yeh, R.A.; Chen, C.; Lim, T.Y.; Schwing, A.G.; Hasegawa-Johnson, M.; Do, M.N. Semantic image inpainting with deep generative models. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
33. Denton, E.; Gross, S.; Fergus, R. Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks. Available online: <https://arxiv.org/pdf/1611.06430> (accessed on 20 August 2019).
34. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
35. Liu, W.; Liu, X.; Ma, H.; Cheng, P. Beyond Human-level License Plate Super-resolution with Progressive Vehicle Search and Domain Prior GAN. In Proceedings of the 25th ACM international conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1618–1626.
36. Inoue, N.; Furuta, R.; Yamasaki, T.; Aizawa, K. Cross-domain weakly-supervised object detection through progressive domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]

37. Li, J.; Liang, X.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Perceptual generative adversarial networks for small object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. Available online: <https://arxiv.org/abs/1512.03385> (accessed on 20 August 2019).
39. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein Gans. Available online: <https://arxiv.org/abs/1704.00028> (accessed on 20 August 2019).
40. Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015. [CrossRef]
41. Redmon, J.; Farhadi, A. Yolov3: An Incremental Improvement. Available online: <https://arxiv.org/pdf/1804.02767> (accessed on 20 August 2019).
42. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Piotr, D. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, 2999–3007. [CrossRef]
43. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. Dota: A large-scale dataset for object detection in aerial images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).