DE-CycleGAN: An Object Enhancement Network for Weak Vehicle Detection in Satellite Images

Peng Gao ⓘ, Tian Tian ⓘ, Linfeng Li, Jiayi Ma ⓘ, and Jinwen Tian

Abstract—**Vehicle detection is a very important application of remote sensing. However, suffering from the low acutance and insufficient color information, the detection of weak vehicles in satellite imagery still remains a challenge. Image enhancement can improve the visual effects of remote sensing images. Nevertheless, most existing image enhancement methods aim to improve the quality of the entire image without target guidance, which have ambiguous contributions to the detection performance. Methods based on generative adversarial networks (GANs) have realized image enhancement with target guidance by the addition of target-guided branches, but paired training data is not available in some scenarios. In this article, a novel model of detection-guided CycleGAN (DE-CycleGAN) is proposed to enhance the weak targets for the purpose of accurate vehicle detection, where a backbone GAN with a target-guided branch is learned in the absence of paired images. Specifically, enhancements of two levels are mutually executed. At the image level, the color information of the entire satellite image is enriched by refined CycleGAN, and its sharpness is enhanced by the gradient enhancement model. At the object level, the target-guided branch for detection is added to enhance features of the target. The experimental results validate that the detection performance has been significantly improved on the images enhanced by the proposed DE-CycleGAN model, which shows a positive effect on weak target detection.**

Index Terms—**CycleGAN, generative adversarial networks (GANs), object enhancement, target-guided branch, weak vehicle detection.**

I. INTRODUCTION

VEHICLE detection is one of the topics that are being widely studied for object detection in remote sensing imagery. Most public vehicle datasets are aerial images of high quality, such as DLR 3K [1], UCAS-AOD [2], and DOTA [3]. Many vehicle detection works [4] have yielded high detection scores on these open high-resolution aerial images. However, in some scenes, some targets in satellite imagery are weak, poorly defined, and lacking of color information. Self-limited by imaging conditions of the satellite instrument, filters of

Peng Gao, Tian Tian, Linfeng Li, and Jinwen Tian are with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: gaopengde@126.com; tiantian_hust@foxmail.com; d201780635@hust.edu.cn; jwtian@mail.hust.edu.cn).

Jiayi Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jyma2010@gmail.com).

the imaging instrument inevitably generate blurry images. And during the denoising process of image prepossessing, details will be probably lost. For adversarial purposes, targets in some scenes tend to hide their colors and textures in the background, making them difficult to be detected. Moreover, the targets in these scenes are usually small and categories are unevenly distributed. All of the above have posed a great challenge for weak vehicle detection in satellite imagery.

A. Image Enhancement Methods

Due to the imaging differences, the same detection model applied to the satellite imagery usually cannot perform as well as which on the aerial imagery. How to elevate the detection performance in satellite images has become an urgent study in remote sensing applications. Image enhancement aims to improve image quality of high-level vision tasks, and two kinds of image enhancement approaches are considered to improve the performance of weak object detection. One is to improve image quality through image prepossessing, such as denoising [5], image sharpening [6], and histogram equalization. The other is to provide external information through supervised methods, such as super resolution (SR) [7], high dynamic range (HDR) [8], and salience enhancement [9]. However, both of them are implemented on the image level, which treats the image as a whole without target prior knowledge, and therefore results in few improvement on the detection tasks.

B. GAN for Image Generation

Generative adversarial network (GAN) [10] is a common generation model which can generate different images for particular tasks. It uses an adversarial architecture including the generator and the discriminator to generate images. Many target-guided branches [11], [12] are proposed with GAN to improve performance in detection tasks. Compared with other image enhancement methods, GAN can reconstruct images for specific tasks, which means that targets can be reconstructed with more salient features. In the task of weak target detection, many features of weak targets can be learned during the generation process. Furthermore, features of weak targets can be enhanced by GAN with a target-guided branch. Many supervised manners can be added to the generation, and the generation can be adjusted for special tasks, e.g., the detection task. With target-guided branches, many GANs have achieved improvements in computing vision tasks such as image translation and super-resolution. Targets will be enhanced in the generating process of GAN with the help of

target-guided branches, and the output images will present more details, color information, and salience than enhancement results generated by other methods.

However, the generation process of GAN-based models is difficult to converge and the training processes are not stable either. Without the additional constraint imposed by extra supervised manners, the detection improvement gained from image enhancement is not remarkable. Strong regulation beneficial for target detection is required in the generation process, but this requires additional paired training samples, which are always difficult to collect. CycleGAN [13] employs unpaired images for image-to-image translation, and this valuable character for image generation has aroused our interests. CycleGAN uses adversarial loss in the generator and discriminator as the global style loss, and employs the identity mapping loss and the cycle consistency loss as the content loss functions. It is more practical to use CycleGAN for image generation than many other algorithms which require paired training data.

C. Motivation

From the aforementioned methods, we are motivated to realize the weak vehicle enhancement for target detection using a framework of CycleGAN with the addition of target-guided branch. Specifically, we collect two datasets, one is the aerial imagery of clear land surfaces and the other is the satellite imagery containing blurry vehicles. Vehicles in the aerial imagery are characterized by low noise, rich color information, and fine textures, whereas targets in the satellite imagery lack color information and sharp contours. CycleGAN can transfer blurry targets to clear targets and enrich color information. An auxiliary target guidance branch is added to the enhancement model to enrich the target features for detection.

However, some problems need to be solved when we directly utilize CycleGAN for target enhancement. First, training process is hard to converge because CycleGAN only employs a discriminator to supervise the generator, which makes it hard to capture the information distribution of the target domain. Second, the enhanced vehicles may appear distorted and unrealistic and the color distribution may be not uniform since CycleGAN is originally designed for style transfer. Third, since the model adopts single classification in the discriminator as the regulation, local features are ignored which lead to the loss of some tiny objects. Finally, false targets will be randomly generated in the enhanced images, resulting in more false alarms in the detection process. The above problems reveal that it is necessary to develop stronger supervised model to control the image generation process and enrich the spatial information if CycleGAN is employed as the image enhancement model for vehicle detection tasks.

D. Proposed Model

In order to address these problems and build an object enhancement network for weak vehicle detection, we propose DE-CycleGAN. It is designed based on CycleGAN with a detection guided branch. Fig. 1 shows some samples of generated images

Fig. 1. Samples of the image enhancement using DE-CycleGAN. From the left column to the right column are DLR 3 K images, generated images form DLR 3 K images, weak satellite images, generated images form weak satellite images. The rightmost column is used for weak vehicle detection.

using DE-CycleGAN. Two levels of enhancements are done as following.

- At the image level, vehicles are enhanced by image-to-image translation. Some improvements for target detection tasks are made on CycleGAN. Aerial images are chosen as the reference imagery for their better qualities. Gradient loss function is replaced by content loss function, and identity loss function is also introduced to enforce the global style learning process.
- At the object level, vehicle generation is controlled by the directed branch. In the object detection task, target branch for detection is added to guide the generation.

The major implements and improvements we have employed include the following.

- High spatial resolution aerial images are chosen as the reference images for learning because of their vivid color information and plenty details. Aerial images are scaled so that vehicles in them are similar in size to those in the satellite imagery, therefore the satellite images with weak targets can learn the global style beneficial for target detection. Vehicles in the aerial imagery are similar in size to those in the satellite imagery, therefore the satellite images with weak targets can learn the global style beneficial for target detection. Color information and details of the vehicles in the generated images are enriched by the process of image transformation.
- More adversarial losses are introduced into the whole network. In the identity mapping section, L1 loss function is replaced by the adversarial loss function. Since identity mapping loss function plays an important role in the content generation of CycleGAN, this modification affects

the contents learned by the model, and thus will improve color distributions of weak objects. In the reconstruction section, auxiliary adversarial loss is added to learn the whole distribution of the images. All the above works are beneficial to train the generator to converge.

- Gradient loss function is added as a content loss function in the generation and reconstruction sections, besides L1 loss function. Gradient feature preserves the details of tiny objects and improves the sharpness of targets.
- A target-guided branch for detection is added to Cycle-GAN. The classification loss function is utilized to define global style loss in the generator and discriminator. However, it is a weakly supervised approach compared to the detection and segmentation approaches, and generated images are sometimes unpredictable. In our model, we propose an auxiliary targeted detection branch for CycleGAN to get better performance on weak object detection tasks.

E. Structure of Article

The first section of this article introduces the background of our research and highlights of our work. The second section provides a brief description of the work related to image enhancement, image-to-image translation, and vehicle detection. The third section describes the structure of the model in detail, lists the loss functions, and explains the architecture. The fourth section is dedicated to experiments and analysis, listing, and discussing the results of comparisons. The final section presents our conclusions.

II. RELATED WORKS

A. Weak Vehicle Detection

Vehicle detection is one of the widely studied topics in the field of remote sensing. Some models, such as RetinaNet [14], Faster RCNN [15], and YOYO v3 [16], have turned out to be effective models in vehicle detection. RetinaNet is one of the one-stage detection models with a focal loss function. It is proposed to balance the extreme foreground–background class imbalance encountered during training of dense detectors. RetinaNet extracts features with VGGNet or ResNet as the backbone and the feature pyramid network (FPN) as the neck, and fuses different levels of the third, fourth, and fifth feature layers to obtain more local information to obtain accurate proposed anchor results.

With numerous works on object detection, many methods have been proposed on weak object detection, such as FPNs [17], attention mechanisms, data argumentation [18], and image enhancement [19]. The above methods have improved the performance on the weak target detection tasks, but shown limitations in many scenes due to the poor imaging environments.

B. Image Enhancement Method

Image enhancement is a fundamental part of image processing and it plays a pivotal role in weak object detection with different implementations. Image sharpening is a powerful tool to emphasize image texture using unsharp masks. Image denoising [5] removes noise and preserves objects. Image smoothing [20] is often used to reduce noise within an image or to produce an image with lower pixels. L0 gradient minimization [20] can smooth the image while sharpening major edges of an image and eliminating low-level structures. Histogram equalization is a method of image processing that uses histogram of an image for contrast adjustment, which provides better image quality without big loss of information. However, traditional image enhancement methods are transposed on empirically adjusted parameters and are limited to a single scene.

C. GAN for Image Generation

Image-generating methods are a very hot topic in the field of computer vision and image processing. With the generative models, targets in the generated images can be reconstructed and enhanced for advanced vision tasks. GAN is a common generative model in order to generate undistinguished outputs from inputs based on convolution neural networks. GAN contains two networks: a generator and a discriminator. The generative network (also known as generator) is used to generate samples. Discriminator network (also known as discriminator) is used to estimate the probability that a sample comes from the training data rather than the generator. Some supervised methods have been proposed recently based on GANs, such as image deblurring [21] or image denoising [5], super-resolution [22], high dynamic range [8], image-to-image translation [13], image classification [23], image fusion [24], [25], image generation [26], and domain adaption [27]. Structure-preserving super-resolution with gradient guidance [22] provides a structure-preserving super-resolution method to reduce distortions introduced by GAN. HDRNet employs bilateral mesh processing and local affine color transformations to provide real-time image enhancement of full-resolution images while still capturing high-frequency effects. Image-to-image translation maps images in one domain to other domains so that the synthetic images will display the similar style to the reference image.

To cope with the problem of insufficient training samples, some works have been done on vehicle enhancement using GAN. MC-GAN [28] is a multicondition constrained GAN, which can efficiently generate samples and improve the performance of trained detection for synthetic samples. Li *et al.* [29] proposed a vehicle detail enhancement method using GAN with foreground prior in order to evaluate the performance of aerial video for small and medium-sized vehicle detection. Zheng *et al.* [30] presented a learning method called vehicle synthesis GANs (VS-GANs) to generate annotated vehicles from very high-resolution remote sensing images. The above works that are carried out on very high resolution remote sensing images provide some motivations on weak vehicle enhancement.

It has been a hot topic that how to incorporate supervised methods in special vision tasks for better performance. Multitask networks [31], [32] introduce another parallel task to improve the main task. In image-to-image translation, in addition to classification, other supervised methods are introduced for strong supervision. Feature-guided GAN uses segmentation prior, instance prior, or attention prior to guide the generation for individual task-specific performance. Furthermore, detection-based unsupervised models [12] have been proposed to maintain the characteristics of object targets during the image generating.
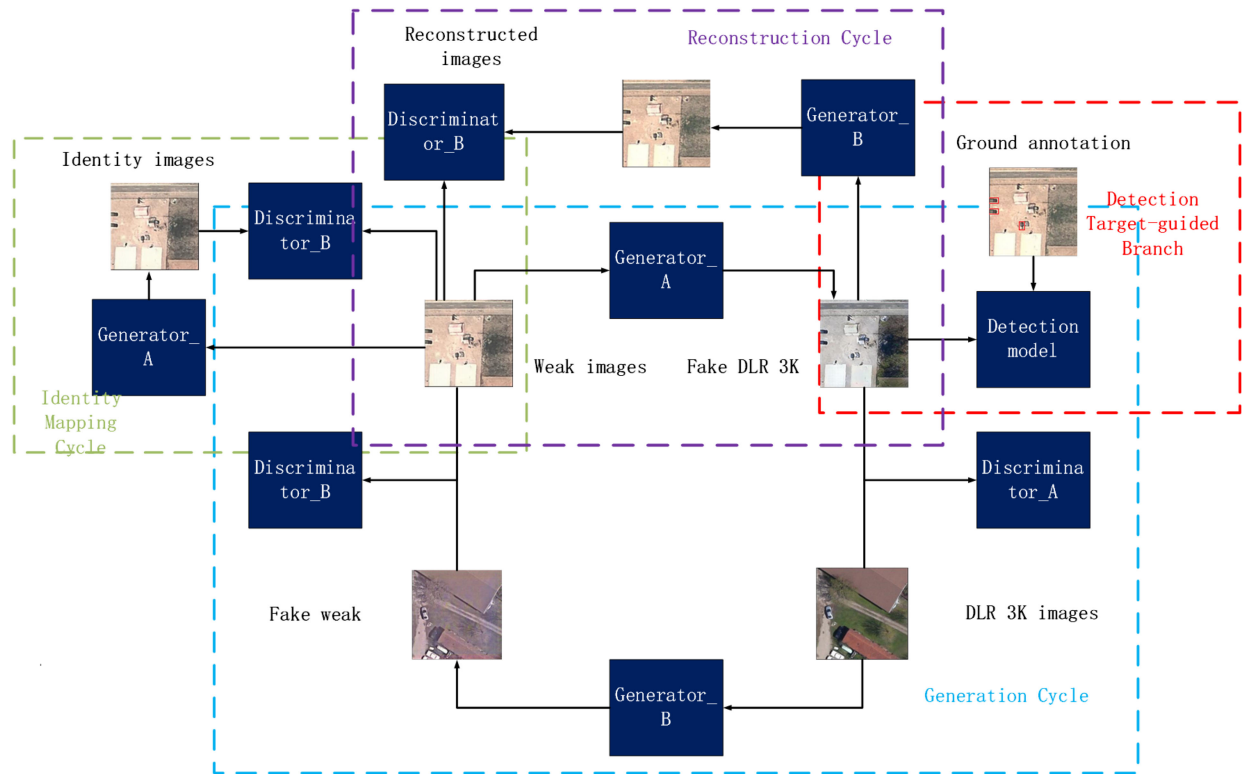
Fig. 2. Structure of DE-CycleGAN. The blue part is the generation cycle, the purple one is the reconstruction cycle, the light green denotes the identity mapping cycle, and the red one denotes the detection targeted-guided branch.

A segmentation task branch is added into the domain adaption with GAN by Shi *et al.* [27].

However, there are many drawbacks for exploiting those proposed GANs into the high-vision tasks directly. First, generation process is difficult to force and paired training data is needed for those supervised manners. However, paired training data is not available in some scenarios and more extra labeling works need to be done. CycleGAN [13] employs unpaired images as inputs and translates the image from source domain to reference domain effectively. It designs a cyclic consistency loss function to maintain consistency between input and output images. Thus, generation task with CycleGAN instead of GAN is a more efficient approach since there is no more work to collect paired images with annotations. Since CycleGAN uses unpaired images, more supervised methods need to be introduced into the training process in order to converge. More classification loss functions and gradient loss functions are considered and the training process is adapted. XGAN [33] divides the generation into two parts: the encoder part and the decoder part. By reducing the input and output variance through the encoder–decoder part, XGAN obtains a stable generation model. DRIT [34] is an example of CycleGAN untangling structure. DRIT trains two branches of the encoder model: content adversarial part and style adversarial part. Content information domain is shared across unpaired domains, while the style information domain is preserved within each domain. Moreover, those GANs with task-oriented branches are designed to generate images with more efficient target features, while no more work is done

for advanced vision tasks such as detection, salience detection, and instance segmentation. Targets in the synthetic images are salient, but they are evaluated by human perceptual psychophysical similarity, not by detection or segmentation metrics. Therefore, these enhancement models are not suitable for direct application to detection or segmentation tasks. There are more works to be done for GAN with task-guided branches to generate more salient objects for high-vision tasks.

III. METHOD

The structure of the DE-CycleGAN is illustrated in Fig. 2, and the entire model is divided into two parts: one part is the generation model, where a GAN used for generating images for weak vehicle enhancement is designed based on CycleGAN. It is designed to enhance targets on the image level. And the other part is the detection model which introduces a target-guided branch to the generation part for the detection task. It is designed on the object level. The former adds refinements to CycleGAN to enforce the image generation process and enrich the target details. CycleGAN is divided into three parts: the generation cycle, the reconstruction cycle, and the identity mapping cycle. The generator in the generation cycle is responsible for generating satellite images from the aerial images that are difficult to distinguish from real images. The reconstruction cycle makes synthetic satellite images similar to the inputs. The identity mapping cycle is used to regulate the style of the generated targets. All of three parts are redesigned and more loss functions

are introduced into our model. The detection part introduces a detection-guided branch, which is used to control the image generation process of GAN so that targets in the synthetic image are suitable for detection tasks. The detection part adopts the RetinaNet model with annotations as the ground truth for target detection.

The generation cycle includes two generators and two discriminators: the first generator generates fake DLR 3 K images from weak images and the second generator generates fake images from DLR 3 K images. Both fake DLR 3 K images and DLR 3 K images are distinguished in the first discriminator and both of weak images and fake images are brought to the second discriminator. The generation cycle and the identity cycle share the same generator, and these cycles are designed to keep the image generated by the generator consistent with the input. The reconstruction cycle contains two generators and a discriminator: the generator shared with the generation cycle generates fake DLR 3 K images and the second generator translates it into the reconstructed images. Both weak images and reconstructed images are fed into the discriminator to be distinguished from each other. The detection targeted-guided branch includes a detection model, and detection annotations are employed as the ground truth. Two kinds of generators and discriminators are designed in our model. One generator is used to generate fake DLR 3 K images from weak images and the other is used to generator fake images from DLR 3 K images. The former is shared in the identity cycle and the latter is shared in the reconstruction cycle. One discriminator is used to distinguish the fake DLR 3 K images and DLR 3 K images, and the other is used to distinguish the fake images and DLR 3 K images. The latter is shared in identity cycle and reconstruction cycle. The difference between the two generators, as well as the two discriminators, is marked in Fig. 2.

A. Generation Model

CycleGAN chooses the adversarial loss in the generator and discriminator as the style loss, and utilizes the identity mapping loss and cycle loss as the content loss. We perform image translation from two unpaired datasets using refined CycleGAN. The weak vehicle imagery is translated to the fake DLR 3 K imagery in order to enhance the targets. We introduce gradient loss to the content loss to enrich target details and add an adversarial loss into the reconstruction cycle to force reconstruction. Furthermore, we apply the gradient loss and adversarial loss instead of the L1 loss as the identity mapping loss to adjust the style of synthetic images for detection tasks.

1) Gradient Loss: We add Structural SIMilarity (SSIM) loss as the gradient loss to L1 loss as our content loss. We compute the content loss function between two pairs: input and synthetic images, and input and reconstructed images. The gradient loss function based on SSIM loss contributes to learning more details compared to the L1 loss. In addition, sliding windows used in the SSIM loss function learn more local target features.

L1 loss is denoted as follows:

$$\mathcal{L}_{\text{pixel}}(X, G(X)) = \mathbb{E}\left[\|\, M(X) - M(G(X))\|_1\right] \quad (1)$$

where M is the gradient calculation model, and G is the generator. SSIM loss function is formulated as follows:

$$SSIM(X, G(X)) = \frac{2\mu_x\mu_{G(x)} + C_1}{\mu_x^2 + \mu_{G(x)}^2 + C_1} \times \frac{2\sigma_{xG(x)} + C_2}{\sigma_x^2 + \sigma_{G(x)}^2 + C_2} \quad (2)$$

where C_1, C_2 are constants and μ_x, $\mu_{G(x)}$, σ, $\sigma_{G(x)}$, $\sigma_{xG(x)}$ denote the mean, standard deviation, and cross-covariance of the image pair $(X, G(X))$.

Two gradient loss functions are computed as follows.

For the generation cycle:

$$\mathcal{L}_{gg}(X, G) = \mathcal{L}_{\text{pixel}}(M(X), M(G(X))) \\ + SSIM(M(X), M(G(X))) \quad (3)$$

where $\mathcal{L}_{gg}(X, G)$ is the gradient loss function in the generation cycle. X is the input of the weak vehicle images, G is the generator, and $G(X)$ is the synthetic fake DLR 3 K image.

For the reconstruction cycle:

$$\mathcal{L}_{\text{gr}}(X, G, F) = \mathcal{L}_{\text{pixel}}(M(X), M(F(G(X)))) \\ + SSIM(M(X), M(F(G(X)))) \quad (4)$$

where $\mathcal{L}_{\text{gr}}(X, G, F)$ is the gradient loss function in the reconstruction cycle, F is the opposite generator, and $F(G(X))$ is the reconstructed images similar to input X.

2) Adversarial Loss: CycleGAN employs adversarial loss in the generation cycle, and we introduce another adversarial loss to the reconstruction cycle in order to force the whole image generation. The composite of the gradient loss function and the adversarial loss function replaces the L1 loss function to capture more features for our detection model.

For generator $G: X \to Y$ and its discriminator D_Y, adversarial loss formulation is as follows:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))] \quad (5)$$

where $x \sim p_{\text{data}}(x)$ and $y \sim p_{\text{data}}(y)$ are the data distribution of the X and Y. Generator G tries to generate images $G(x)$ that are similar to the images from the domain Y, while discriminator D_Y aims to distinguish between translated samples $G(x)$ and real samples y. A similar adversarial loss for the generator $F: Y \to X$ and its discriminator D_X as well.

Especially, for the generation cycle

$$\mathcal{L}_g(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))] \quad (6)$$

$$\mathcal{L}_d(F, D_x, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_X(x)] \\ + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log(1 - D_X(F(y)))] \quad (7)$$

where $\mathcal{L}_g(G, D_Y, X, Y)$ is the adversarial loss for generator, $\mathcal{L}_d(F, D_x, X, Y)$ is the adversarial loss function for discriminator, X is the weak vehicle datasets, and Y is the DLR 3 K

datasets. The whole loss function in the generation cycle is

$$\mathcal{L}_{ag}\left(G, F, D_x, D_Y, X, Y\right) = \mathcal{L}_{gg}(X, G) + \mathcal{L}_g\left(G, D_Y, X, Y\right)$$
$$+ \mathcal{L}_d\left(F, D_x, X, Y\right) \quad (8)$$

where $\mathcal{L}_{ag}(G, F, D_x, D_Y, X, Y)$ is the adversarial loss in the generation cycle.

Especially, for reconstruction cycle

$$\mathcal{L}_g\left(G, D_X, X, X\right) = \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log D_X(x)\right]$$
$$+ \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log(1 - D_X\left(F(G(x))\right))\right] \quad (9)$$

$$\mathcal{L}_d\left(F, D_y, Y, Y\right) = \mathbb{E}_{y \sim p_{\text{data}}(y)}\left[\log D_Y(y)\right]$$
$$+ \mathbb{E}_{y \sim p_{\text{data}}(y)}\left[\log(1 - D_Y\left(G(F(y))\right))\right] \quad (10)$$

where $\mathcal{L}_g(G, D_Y, X, Y)$ is the adversarial loss for generator, and $\mathcal{L}_d(F, D_x, X, Y)$ is the adversarial loss function for discriminator.

The whole loss function in reconstruction cycle is

$$\mathcal{L}_{ar}\left(G, F, D_x, D_Y, X, Y\right) = \mathcal{L}_{gr}(X, G, F)$$
$$+ \mathcal{L}_g\left(G, D_X, X, X\right)$$
$$+ \mathcal{L}_d\left(F, D_y, Y, Y\right) \quad (11)$$

where $\mathcal{L}_{ar}(G, F, D_x, D_Y, X, Y)$ is the adversarial loss in the generation cycle.

3) Identity Mapping Loss: Identity mapping loss in Cycle-GAN is used to maintain consistency between the input image and the generated image. Given a DLR 3 K image as the input, CycleGAN uses the generator of L1 loss function (weak image to DLR 3 K image) to keep outputs the same as inputs. If there is no identity mapping cycle, the learned style will be changed. In other words, L1 loss function affects the style of generated images. Therefore, we can change the style of outputs by designing the identity loss function.

With a gradient loss function and an adversarial loss function added in the generation cycle and reconstruction cycle, the style learned by CycleGAN is changed. In order to refine the style learned by the generator, we apply an adversarial loss function instead of L1 loss function as the identity mapping loss. In this way, the style of vehicles in the synthetic images is changed, which will be suitable for detection tasks.

For the identity mapping cycle

$$\mathcal{L}_g\left(G, D_x, X, X\right) = \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log D_X(x)\right]$$
$$+ \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log(1 - D_X\left(G(x)\right))\right] \quad (12)$$

$$\mathcal{L}_d\left(F, D_x, X, X\right) = \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log D_X(x)\right]$$
$$+ \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log(1 - D_X\left(F(x)\right))\right] \quad (13)$$

where $\mathcal{L}_g(G, D_x, X, X)$ is the loss in the generator, and $\mathcal{L}_d(F, D_x, X, X)$ is the loss in the discriminator.

The whole loss function in the identity cycle will be

$$\mathcal{L}_i\left(G, F, D_x, D_Y, X, Y\right) = \mathcal{L}_g\left(G, D_X, X, X\right)$$
$$+ \mathcal{L}_d\left(F, D_y, Y, Y\right). \quad (14)$$

B. Detection Model

With the learned global style, it is also necessary to add a supervisory branch to train the generator to get result guidance by the detection task. With a targeted detection branch, the synthetic images by GAN will become more suitable for detection tasks. We add a detection model pretrained with weak images to the discriminant section. The generation model refreshes parameters with its own parameters fixed in order to optimize the image generation process. We also design a model that refreshes the detection model during the training step to obtain better performance of the detection task. We use RetinaNet as the detection model because it has been proven to be a suitable detection model in vehicle detection. RetinaNet exploits focal loss function to solve the class imbalance problem for positive and negative samples.

The loss function of the refreshed detection targeted branch is

$$\mathcal{L}_{\text{detection}}(g, p) = \mathcal{L}_{\text{regression}}(g, p) + FL(p_t) \quad (15)$$

where $\mathcal{L}_{\text{regression}}(g, p)$ is L1 loss function computing the annotations and the predicted boxes, where g is the annotation and p is the detection predicted boxes. $FL(p_t)$ is the focal loss function, and p_t is the refined classification prediction. The inputs are the outputs of the former generations.

In conclusion, the whole loss function in the DE-CycleGAN becomes

$$\min_{G, F, g, p} \max_{D_x, D_Y} \mathcal{L} = \lambda_1 \mathcal{L}_{gg}(X, G)$$
$$+ \lambda_2 \mathcal{L}_{ag}\left(G, F, D_x, D_Y, X, Y\right)$$
$$+ \lambda_3 \mathcal{L}_{gr}(X, G, F)$$
$$+ \lambda_4 \mathcal{L}_{ar}\left(G, F, D_x, D_Y, X, Y\right)$$
$$+ \lambda_5 \mathcal{L}_i\left(G, F, D_x, D_Y, X, Y\right)$$
$$+ \lambda_6 \mathcal{L}_{\text{detection}}(g, p) \quad (16)$$

where λ_1, λ_2, λ_3, λ_4, λ_5, and λ_6 are the coefficients of different components.

IV. Experiments

Two datasets are utilized in our experiment. One is the open Munich DLR 3 K vehicle imagery in 2015 with 20 aerial images. It contains 14 235 vehicles with a ground sampling distance of 0.13 m. The other is a weak vehicle target imagery that we collect from a wilderness area including wild and urban area via Google Earth. Weak scenes contain a lot of valuable targets. Train datasets contains 150 satellite images with a spatial resolution of less than 0.5 m and 12000 vehicle objects divided into nine classes: cars, pickups, trucks, armored vehicles, tanks, trailers, artillery, pickups, and tent. Trucks, armored vehicles, tanks, trailers, and tents are large-scale targets with a size of about 40×60 pixels. Pickup trucks are medium-scale targets with a size of about 20×30 pixels. Cars and cannons are small-scale targets, measuring approximately 10×20 pixels. Test datasets contain 2100 vehicles, of which 1204 are trucks, 500 are pickups, 150 are artillery, 210 are pickups, and 38 are tents. There are large differences in the number of vehicles in different categories,

TABLE I
SPEED AND RESULTS OF DIFFERENT COEFFICIENTS FOR λ_6

λ_6	1	3	5	7	10
Epoch of convergence	145	45	**15**	55	120
Average score	0.640	0.682	**0.712**	0.645	0.512

The bold entities indicate the best results compared to other numbers in each row.

which lead to a large impact on the detection results. Trucks, armored vehicles, tanks, trailers, pickups, and tents are large targets, while cars, pickups, and cannons are small targets. Truck, small truck, pickup, artillery, and tent are selected as evaluation indicators for the amount of rest of vehicle categories in both train and test datasets are small. Vehicle categories are judged by both visual interpretation and prior knowledge. Prior knowledge is used to select weak scenes containing both military vehicles and civilian vehicles. And visual interpretation is used to judge the vehicle categories. Two scenarios are included in the experiments such as the parking lot and road. Satellite images which should contain those types of vehicles are collected. Urban and wild areas containing wild scenes, parking lot, road, and buildings are included in our datasets.

Targets in two datasets are similar in size and shape. However, weak field scenario datasets differ from the aerial datasets greatly. Vehicles in the Munich DLR 3 K can be divided into multiple components, whereas vehicles in the weak vehicle datasets lack detail and cannot be divided into local components. Targets in the Munich DLR 3 K are rich in color information and have high image acutance compared to targets in the weak vehicle datasets. Vehicles also have different color distributions and spatial resolutions.

We used RetinaNet based on pretrained ResNet-101 weights as a baseline for comparison with the image-enhanced detection model. Three comparisons were performed to evaluate the effectiveness of our model. We set λ_1, λ_2, λ_3, λ_4, and λ_5 as 1 and λ_6 as 5 in formula (16) to strengthen the detection-guided performance in the synthetic images. The coefficients affect the speed of convergence and the result. λ_1, λ_2, λ_3, λ_4, and λ_5 are set to the same value, and λ_6 is chosen from 1, 3, 5, 7, and 10 empirically. A too small or too large λ_6 will result in difficulties on training convergence. The speed and results were tested in the experiments of RetinaNet with DE-CycleGAN and listed in Table I. The weights with the best target detection results in training are saved.

A server with GPU of 1080Ti is used in our experiment. The learning rate of GAN is 0.001 and it reduces to 0.1 times every 10 epochs. The learning rate of detection model is 0.0001 and it reduces to 0.1 times every 10 epochs. The optimizers of GAN and detection model are Adam optimizer. The batch size of GAN without detection branch is 8 and the batch size of DE-CycleGAN is 4. The image size in the GAN is 256 and the image size in detection branch is 416. For other target detection tasks, we choose the mean average accuracy (mAP) as the metric for target detection results, which is the mean value of the average accuracy with IoU from 50 to 95. It is the most popular evaluation metric used in the detection tasks.

A. Effects on Different Detection Models

Three quantitative comparisons are executed with three common detection models (RetinaNet, Faster RCNN, YOLO V3) to evaluate our target enhancement model. Three models are trained and tested in source and enhanced images, respectively. Detection scores are listed in Table II.

RetinaNet achieves the highest scores compared with other two detection models in the raw images while Faster RCNN suffers the lowest scores in four kinds of vehicles. Our weak vehicle datasets contain a class imbalance distribution of different kinds of vehicles. Trucks have the most samples and the other categories are much less. RetinaNet exploiting focal loss can solve this class imbalance problem for positive and negative samples. All the above demonstrate that RetinaNet is suitable to be used in our weak vehicle datasets. Since targets have been enhanced by the proposed DE-CycleGAN, all of three detection models achieve improvements on detection performance.

Table II shows that target enhancement algorithm achieves performance improvement of all three detection models. RetinaNet based on focal loss achieves the highest detection results in four categories of targets and obtains the highest improvement in the detection of two categories of targets, while Faster RCNN achieves the most categories of three. Faster RCNN achieves the greatest performance on the artillery, and the artillery has smaller and less targets compared to other kinds of vehicles in the test dataset. The RPN used in Faster RCNN suffers from small target miss detection and has a low raw detection rate for artillery. Once artillery is enhanced by GAN, Faster RCNN with DE-CycleGAN can distinguish the wrong anchor points and keep more accurate proposals to improve the detection performance.

B. Comparison Against Different Image Enhancement Models

Five image enhancement algorithms are compared with our target enhancement model to evaluate the effectiveness of our model. Raw input is used as baseline. Three of them are unsupervised methods (histograms equalization, L0 smooth, and sharpening) and the other two are supervised methods (SPSP, MirNet). Parameters of unsupervised methods are manually configured empirically. SPSR is pretrained using DLR 3K dataset and MirNet is not retrained due to lack of paired high-light and low-light images. Target-guided GAN-based image generation algorithm lacks publicly available source code and weights, so it is not selected as a comparison model. Results are listed in Table III, and samples of enhanced images are illustrated in Fig. 3.

Fig. 3 visualizes examples of different image enhancement methods for weak vehicle images, from left to right: original image, histogram equalization, L0 smoothing, sharpening, SPSR, MirNet, DRIT, and DE-CycleGAN. Supervisory-based method achieves a higher improvement on the detection task compared to unsupervised image enhancement methods, while DE-CycleGAN obtains the best results on the detection task

TABLE II
PERFORMANCE OF DIFFERENT DETECTION METHODS FOR WEAK OBJECT DETECTION

Categories	RetinaNet	RetinaNet + DE-CycleGAN	Faster RCNN	Faster RCNN + DE-CycleGAN	YOLO v3	YOLO v3 + DE-CycleGAN
Truck	0.813	**0.919(+0.096)**	0.750	0.802(+0.052)	0.803	0.812(+0.011)
Small truck	0.787	**0.847(+0.06)**	0.396	0.662(+0.266)	0.619	0.638(+0.019)
Pickup	0.295	**0.510(+0.186)**	0.128	0.233(+0.105)	0.075	0.116(+0.041)
Artillery	0.531	0.771(+0.240)	0	**0.853(+0.853)**	0.100	0.244(+0.144)
Tent	0.348	**0.474(+0.126)**	0.225	0.406(+0.181)	0.169	0.193(+0.024)

The bold entities indicate the best results compared to other numbers in each row.

TABLE III
SCORES OF DIFFERENT IMAGE ENHANCEMENT METHODS IN WEAK OBJECT DETECTION

Categories	Raw	Histograms Equalization	L0 smoothing	Sharpening	SPSR	MirNet	DRIT	DE-CycleGAN
Truck	0.813	0.829(+0.013)	0.844(+0.031)	0.815(+0.002)	0.869(+0.056)	0.806(+0.044)	0.570(-0.243)	**0.919(+0.096)**
Small truck	0.787	0.763(-0.024)	0.766(-0.021)	0.748(-0.039)	0.844(+0.057)	0.744(-0.043)	0.411(-0.376)	**0.847(+0.06)**
Pickup	0.295	0.103(-0.192)	0.225(-0.070)	0.303(+0.008)	0.411(+0.106)	0.005 (-0.290)	0.164(-0.131)	**0.510(+0.186)**
Artillery	0.531	0.6609(+0.129)	0.7619(+0.230)	0.610(+0.079)	0.5179(-0.014)	0.016(-0.525)	0.015(-0.416)	**0.771(+0.240)**
Tent	0.348	0.4659(+0.127)	0.340(-0.008)	0.290(-0.058)	**0.520(+0.172)**	0.381(+0.033)	0.112(-0.236)	0.474(+0.126)

Green numbers are results below the baseline and italic numbers are the highest scores in the Table.
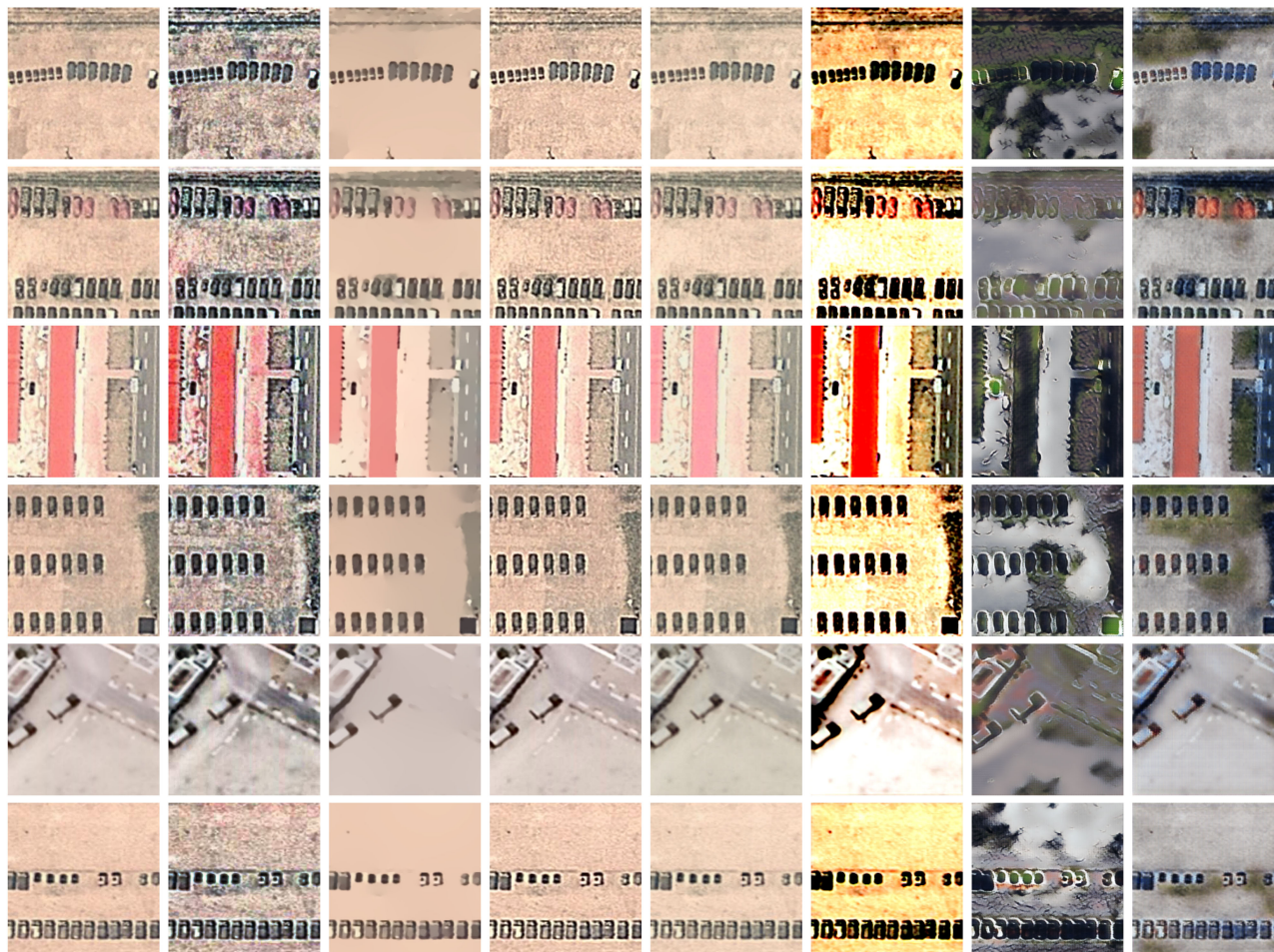

Fig. 3. Examples of different image enhancement methods for weak vehicle images; from left to right: original image, histogram equalization, L0 smoothing, sharpening, SPSR, MirNet, DRIT, and DE-CycleGAN.

TABLE IV
DETECTION RESULTS IN THE DIFFERENT MODELS: RETINANET, CYCLEGAN, DE-CYCLEGAN WIHOUT DETECTION PART, DE-CYCLEGAN, AND DE-CYCLEGAN
WITH UPDATED RETINANET WEIGHTS.

Categories	RetinaNet	CycleGAN	DE-CycleGAN without detection part	DE-CycleGAN	DE-CycleGAN with updated RetinaNet weights
Truck	0.813	0.885(+0.072)	0.875(+0.062)	0.901(+0.068)	**0.919(+0.096)**
Small truck	0.787	0.741(-0.046)	0.777(-0.010)	0.794(+0.007)	**0.847(+0.06)**
Pickup	0.295	0.469(+0.174)	0.409(+0.105)	0.419(+0.095)	**0.510(+0.186)**
Artillery	0.531	0.460(-0.061)	0.464(-0.067)	0.613(+0.082)	**0.771(+0.240)**
Tent	0.348	**0.514(+0.166)**	0.454(+0.106)	0.465(+0.117)	0.474(+0.126)

The bold entities indicate the best results compared to other numbers in each row.
Green numbers are results lower than the baseline.

for four of all target categories. MirNet performs the best on the tent detection. Without any adaptions for detection task, DRIT suffers the lowest detection scores in the comparison experiments.

Six image enhancement samples are illustrated in Fig. 3. Histogram equalization introduces noise while making image sharper. L0 smooths and flattens the image, and blurs details of the object. Sharpening and SPSR have little effect on the whole image. MirNet exposes the image and enhances sharpness. DE-CycleGAN colors the vehicle, and enhances sharpness while reducing noise.

C. Ablation Study for DE-CycleGAN

The third comparison includes RetinaNet, CycleGAN, modified CycleGAN, modified CycleGAN with target detection-guided branch, and modified CycleGAN with updated RetinaNet weights. RetinaNet with ResNet-101 is baseline. Modified CycleGAN refers to CycleGAN with refinements. Modified CycleGAN with detection-guided branch refers to our DE-CycleGAN without detection model weights refreshed in the training. Target-enhanced model with updated RetinaNet weights refers to our DE-CycleGAN with detection model weights refreshed in the training. RetinaNet is trained in the source images and is used as detector in enhanced images. The weights of RetinaNet are updated in the model of the target-enhanced model with updated RetinaNet weights.

The detection scores of five detection models are listed in Table IV. Samples generated by the five models are illustrated in Fig. 4. Table IV shows that refinements against CycleGAN in our target enhancement model have a positive effect on weak vehicle target detection. The whole target enhancement model obtains the highest detection performance among five models. CycleGAN shows a decrease in two classes of vehicles compared to the baseline model. Modified CycleGAN decreases degradation in detection performance compared with Cycle-GAN. Two models with detection-guided branches are improved in all five categories of weak vehicle target detection.

Global style learned of vehicles and detection results of vehicles in the synthetic images vary in different generation models as shown in Fig. 4. In the generated image, the color of vehicles is changed and the acutance is improved. Compared to original inputs, outputs of CycleGAN, modified CycleGAN, modified CycleGAN with detection-guided branch, and modified CycleGAN with updated RetinaNet weights own more color

information. Compared to CycleGAN, the latter three models are able to maintain a gradient distribution similar to the input image, with more vehicle detail and less noise. There is a degraded detection performance in both artillery detection and small car detection for most of five detection models. Both targets are small. Compared to the ground truth, CycleGAN suffers the most false alarms and omissions, while modified CycleGAN, modified CycleGAN with detection-guided branch, and modified CycleGAN with updated RetinaNet weights achieve fewer false alarms and omissions. Our DE-CycleGAN with updated detection weights has the fewest omissions.

D. Evaluations on Different Scenarios

Four scenarios are considered in our experiments. The first scenario is a parking lot, which contains many densely packed vehicles of different types. A variety of targets are densely parked in this scenario, making it easy to miss targets or mis-classify. The second column of images in Fig. 4 shows that the modified CycleGAN with updated RetinaNet weights achieves the best detection performance, which is improved over the modified CycleGAN without updated RetinaNet weights. The improvement in the detection of parking lots proves the effectiveness of our detection of branches. The second scenario is the road containing some small vehicles which will be missing detection easily. The third and fifth columns in Fig. 4 provide the detection improvement from the original image to the image enhanced by CycleGAN, which demonstrates that the model with GAN can improve the detection performance of weak vehicles on the road. The third scenario is a residential area which contains some buildings similar to vehicles that will cause false alarms. The first and fourth columns in Fig. 4 show the decrease of false alarms by modified CycleGAN. The fourth scenario is wild area which contains fewer and smaller vehicles, which are difficult to be detected. The sixth column in Fig. 4 shows the decrease of miss detection by CycleGAN and modified Cycle-GAN. From the experimental results, we can see that our method can improve the image quality and target detection for different scenes. Besides, CycleGAN and modified CycleGAN achieve a larger detection improvement on parking lot, road, and wild scenes, which demonstrates the effectiveness of the generating models by GAN in the image enhancement. An improvement on buildings introduced by modified CycleGAN with updated RetinaNet weights shows the value of the detection branch.
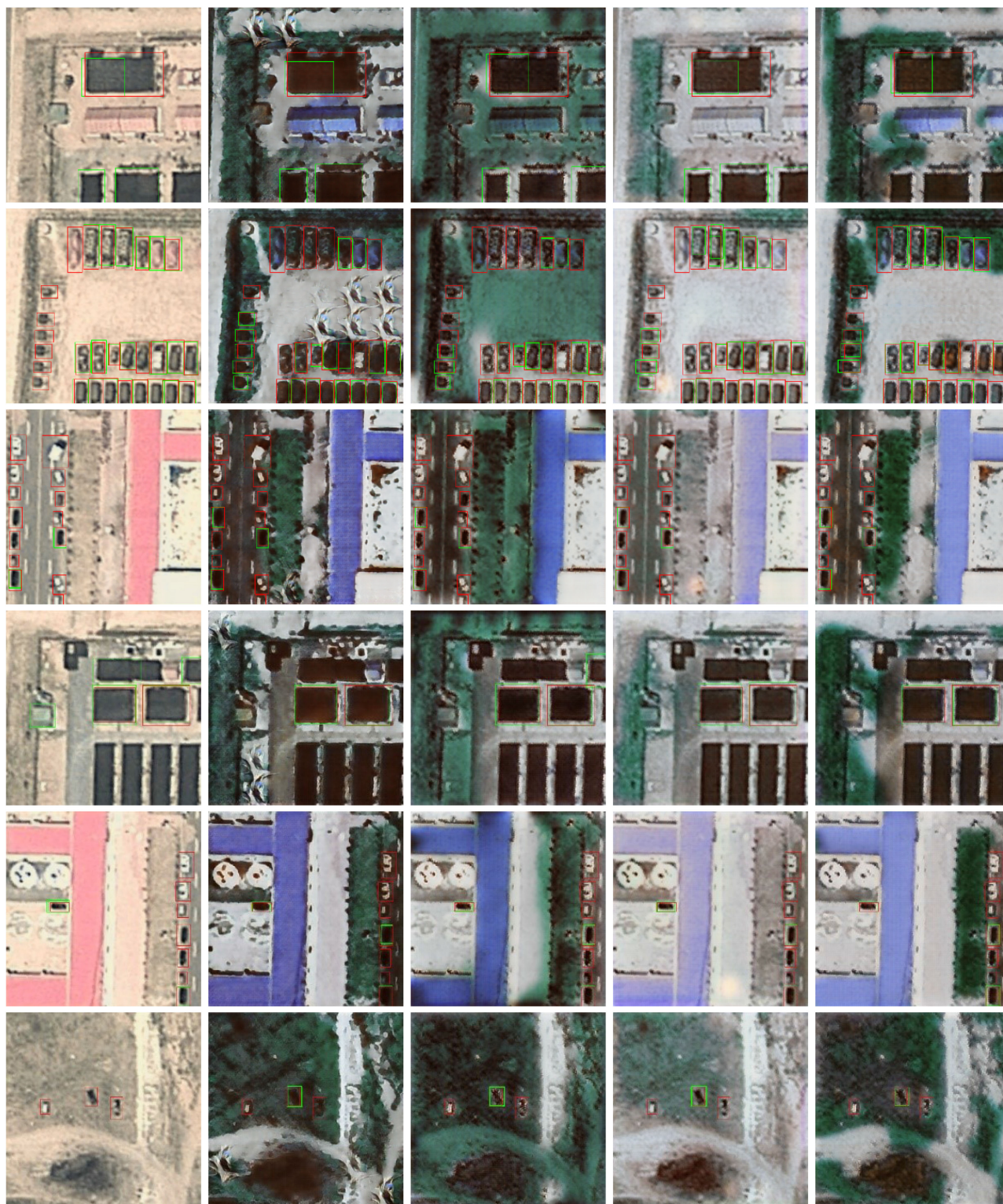
Fig. 4. Detection results of five models. From left to right: raw images, enhanced by CycleGAN, enhanced by modified CycleGAN, enhanced by modified CycleGAN with detection-guided branch, and enhanced by modified CycleGAN with updated RetinaNet weights. Green boxes are predicted boxes and red boxes are annotation boxes.

V. Discussion

The increase in detection score has demonstrated that our model has a positive effect on improving weak vehicle detection performance. Three reasons are elaborated as follows.

- Compared with enhanced methods on the images, image-generating methods, such as GAN, are able to enhance target features from end to end for detection tasks. More features are learned during the reconstruction of images in the generator of GAN. Global style learned from reference imagery contributes to detection tasks, and color of targets is enriched and sharpness is heightened. Image-to-image translation plays a positive role of target enhancement in the detection tasks. Furthermore, extra adversarial loss introduced in the reconstruction cycle and identity mapping cycle forces the learned style of generated images.

- More details of the objects are preserved by gradient loss, which helps to improve detection performance. Compared with CycleGAN, gradient loss function utilized in the DE-CycleGAN is beneficial to keep the details of targets, and is beneficial for the detection tasks.

- Detection target branch acts as a detection guidance for the training of image generation. Extra supervised information is introduced into the generation of GAN, and the detection prior contributes to the detection tasks for the target enhancement. In addition, detection refreshed model achieves the best results on the detection task, demonstrating the effectiveness of the target detection branch.

VI. Conclusion

In this article, we propose a novel weak vehicle enhancement model, named DE-CycleGAN, to improve the performance of weak vehicle detection in satellite images. Compared to other object enhancement algorithms, our model achieves a better enhancement performance for weak vehicles via GANs without paired training data and better training convergence compared to other GAN-based models. We select DLR 3 K imagery as the reference datasets and collect weak vehicle images as the source imagery. The enhancement of weak vehicles is implemented on two levels. On the image level, image sharpness and color information are enhanced by the refined CycleGAN. On the object level, vehicle generation is enforced by the target branch, which will facilitate the detection results. Experimental results have validated that our model achieves the best performance compared to other image enhancement methods and significantly improves the detection accuracy for all three common detection models. The performances of the proposed DE-CycleGAN have shown us the potentials of image enhancement by means of image transformation network for the purpose of weak vehicle detection. In the future work, object enhancement with adversarial networks in more scenarios will be further studied.

Acknowledgment

References

[1] M. Mandal, M. Shah, P. Meena, S. Devi, and S. K. Vipparthi, "AVDNet: A small-sized vehicle detection network for aerial visual data," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 494–498, Mar. 2020.

[2] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 3735–3739.

[3] G. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.

[4] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3652–3664, Aug. 2017.

[5] Q. Xiang and X. Pang, "Improved denoising auto-encoders for image denoising," in *Proc. 11th Int. Congr. Image Signal Process. BioMed. Eng. Inform.*, 2018, pp. 1–9.

[6] X. Tian, Y. Chen, C. Yang, and J. Ma, "Variational pansharpening by exploiting cartoon-texture similarities," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–16, 2021.

[7] J. Ma, X. Wang, and J. Jiang, "Image superresolution via dense discriminative network," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5687–5695, Jul. 2020.

[8] H. Xu, J. Ma, and X.-P. Zhang, "MEF-GAN: Multi-exposure image fusion via generative adversarial networks," *IEEE Trans. Image Process.*, vol. 29, pp. 7203–7216, 2020.

[9] Y. Kim and A. Varshney, "Saliency-guided enhancement for volume visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 925–932, Sep.–Oct. 2006.

[10] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[11] J. Ma *et al.*, "Infrared and visible image fusion via detail preserving adversarial learning," *Inf. Fusion*, vol. 54, pp. 85–98, 2020.

[12] D. Bhattacharjee, S. Kim, G. Vizier, and M. Salzmann, "DUNIT: Detection-based unsupervised image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4786–4795.

[13] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.

[14] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollr, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[16] J. Redmon and A. Farhadi, "YOLOV3: An incremental improvement," 2018, *arXiv:1804.02767.*

[17] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 936–944.

[18] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," 2019, *arXiv:1902.07296.*

[19] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[20] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," in *Proc. SIGGRAPH Asia Conf.*, 2011, pp. 1–12.

[21] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8174–8182.

[22] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7766–7775.

[23] Y. Xi *et al.*, "DRL-GAN: Dual-stream representation learning GAN for low-resolution image classification in UAV applications," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1705–1716, 2021.

[24] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, 2019.

[25] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, 2020.

[26] M. B. Bejiga, F. Melgani, and A. Vascotto, "Retro-remote sensing: Generating images from ancient texts," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 950–960, Mar. 2019.

[27] Q. Shi *et al.*, "Domain adaption for fine-grained urban village extraction from satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1430–1434, Aug. 2020.

[28] J. Shen, N. Liu, H. Sun, and H. Zhou, "Vehicle detection in aerial images based on lightweight deep convolutional network and generative adversarial network," *IEEE Access*, vol. 7, pp. 148 119–148130, 2019.

[29] J. Li *et al.*, "Visual detail augmented mapping for small aerial target detection," *Remote Sens.*, vol. 11, no. 1, p. 14, 2019.

[30] K. Zheng, M. Wei, G. Sun, B. Anas, and Y. Li, "Using vehicle synthesis generative adversarial networks to improve vehicle detection in remote sensing images," *ISPRS Int. J. of Geo-Inf.*, vol. 8, no. 9, 2019, Art. no. 390.

[31] H. Guo, Q. Shi, B. Du, L. Zhang, D. Wang, and H. Ding, "Scene-driven multitask parallel attention network for building extraction in high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–20, 2020.

[32] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–18, 2020.

[33] A. Royer *et al.*, "XGAN: Unsupervised image-to-image translation for many-to-many mappings," in *Domain Adaptation for Visual Understanding*. Berlin, Germany: Springer, 2020, pp. 33–49.

[34] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 35–51.
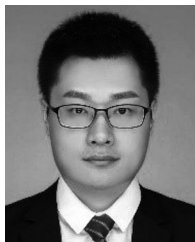
Peng Gao received the B.S. degree in optical information science and technology from the Huazhong University of Science and Technology, Wuhan, China, in 2013, and the M.S. degree in optical engineering from the Huazhong Institute of Optoelectronic Technology, Wuhan, China, in 2016. He is currently working toward the Doctor's degree from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include computer vision, machine learning, and remote sensing object detection.

Tian Tian received the B.S. degree in electronic information engineering and Ph.D degrees in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively.

She joined the School of Computer Sciences, China University of Geosciences, Wuhan, China, in 2015, as a Postdoctoral Lecturer. Currently, she is an Associate Professor with School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. Her research interests include remote sensing image processing, computer vision, and applications.

Linfeng Li received the B.S. degree in automation from the School of Automation, Chongqing University, Chongqing, China, in 2017. He is currently working toward the Ph.D. degree at the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include image processing, pattern recognition, and remote sensing.

Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

He is currently a Professor with the Electronic Information School, Wuhan University, China. He has authored or co-authored more than 150 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, ECCV, etc. His research interests include computer vision, machine learning, and robotics.

Dr. Ma has been identified in the 2020 and 2019 Highly Cited Researcher lists from the Web of Science Group. He is an Area Editor of *Information Fusion*, an Associate Editor of *Neurocomputing* and *Entropy*, and a Guest Editor of *Remote Sensing*.

Jinwen Tian received the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology (HUST), China, in 1998.

He is currently a Professor with HUST. His research interests include remote sensing image analysis, image compression, computer vision, and fractal geometry.