IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 15, 2019, accepted October 13, 2019, date of publication October 16, 2019, date of current version November 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947657

Synthetic IR Image Refinement Using Adversarial Learning With Bidirectional Mappings

RUIHENG ZHANG[1,2], CHENGPO MU[1], MIN XU[2], LIXIN XU[1], QIAOLIN SHI[2], AND JUNBO WANG[3]

[1]School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China
[2]GBDTC, Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia
[3]Beijing Institute of Electronic System Engineering, Beijing 100039, China

Corresponding author: Chengpo Mu (muchengpo@bit.edu.cn)

This work was supported in part by the China Scholarship Council (CSC) and in part by the Beijing Institute of Technology Graduate School.

ABSTRACT Collecting a large dataset of real infrared (IR) images is expensive, time-consuming, and even unavailable in some specific scenarios. With recent progress in machine learning, it has become more feasible to replace real IR images with qualified synthetic IR images in learning-based IR systems. However, this alternative may fail to achieve the desired performance, due to the gap between real and synthetic IR images. Inspired by adversarial learning for image-to-image translation, we propose the Synthetic IR Refinement Generative Adversarial Network (SIR-GAN) to narrow this gap. By learning the bidirectional mappings between two unpaired domains, the realism of the simulated IR images generated from the IR Simulator are significantly improved, where the source domain contains a large number of simulated IR images, where the target domain contains a limited quantity of real IR images. Specifically, driven by the idea of transferring infrared characteristic and protect target semantic information simultaneously, we propose a SIR refinement loss to consider an infrared loss and a structure loss further to the adversarial loss and the consistency loss. To further reduce the gap, stabilize training, and avoid artefacts, we modify the proposed algorithm by developing a training strategy, adding the U-net in the generators, using the dilated convolution in the discriminators and invoking the N-Adam acts as the optimizer. Qualitative, quantitative, and ablation study experiments demonstrate the superiority of the proposed approach compared with the state-of-the-art techniques in terms of realism and fidelity. In addition, our refined IR images are evaluated in the context of a feasibility study, where the accuracy of the trained classifier is significantly improved by adding our refined data into a small real-data training set.

INDEX TERMS Infrared simulation, synthetic refinement, convolutional neural networks, adversarial learning.

I. INTRODUCTION

Different kinds of target detectors based on infrared (IR) thermal images are widely used in the field of remote sensing, such as unmanned vehicles [1], intelligent monitoring [2], [3], and automated target detection systems [4]–[6]. These technologies and systems require a large annotated infrared dataset. However, collecting a large amount of real IR data is expensive and time-consuming, even unavailable in some specific scenarios. Thus, the idea of utilizing synthetic instead of real IR images has become appealing, because the synthetic IR images can be mass generated and easily annotated. Several graphics rendering engines are well-suited to infrared simulation [18], [21], [28], [29]. Many efforts have explored using these generated synthetic data for various prediction tasks, including radar data classification [17], IR target detection [21], and semantic segmentation [20]. These studies illustrate that the model trained on a large quantity of synthetic images outperforms the model only trained on a small number of real images. Thus, it is distinct that synthetic data can be used as substitutes when it comes to lack of real data.

However, synthetic data are unable to replace real data completely. Learning from synthetic IR data can be

The associate editor coordinating the review of this manuscript and approving it for publication was Saeid Nahavandi.

problematic owing to a gap between real and simulated IR images. This gap will lead to bad generalization on a model trained with synthetic IR data, because synthetic IR images are not realistic enough and mislead the model to learn some information which is only present in synthetic data. Therefore, how to close the gap is an important topic. The gap mainly comes from the infrared texture of targets, which is caused by incomplete consideration of influences from various aspects on the target simulation results, such as solar radiation, sky background radiation, and ground radiation. Most of the traditional algorithms focus on improve infrared simulation systems, which may fail to bridge the gap effectively, and require a lot of work [9]–[14]. These methods are restricted by three main challenges, in the improvement of reality on simulated IR images: (1) Traditional infrared rendering methods fail to render correct infrared information of real images. The temperature simulation of the infrared target simulation is not real enough to accurately reflect the target temperature distribution. (2) The target modelling and the finite element analysis requires a massive amount of manual calculation. (3) Current methods fail to learn an effective mapping between real and simulated data automatically. It lacks a deep-learning-based framework to refine the synthetic IR data.

Recently, in the deep learning field, generative adversarial networks (GANs) have addressed the lack of training data via learning from real data probability distribution and generating synthetic data, proposed by [16]. GANs have achieved impressive results in image generation [37], image colorizing [42], [44], and image-to-image translation [41], [43]. In particular, the last few years have witnessed a variety of GANs with convolutional neural networks (CNNs) developed for image-to-image translation tasks. The feed-forward CNNs can be easily trained by using the standard back-propagation approach [45], and the transformed images are generated by forwardly passing the input image through the well-trained CNNs when testing. Pix2pix framework [41] is a strong pipeline to transform image style from a specific image to another image. However, it can not transform between two domains, such as our task, from simulated IR domains to real IR domains. Cycle-gan [8] designs a cycle architecture with a combination of adversarial losses and cycle consistency, to tackle the unpaired datasets, but the method only focuses on style or season transfer application. Shrivastava *et al.* [35] propose the Simulated+Unsupervised (S+U) learning, which is the first methodology that guides how one can use synthetic eye images to improve the performance of learning algorithms, by adopting generative adversarial networks, but it fails to learn mappings between two domains. Nevertheless, none of these GANs models cannot tackle the simulated IR refinement task, owing to limitations of these algorithms and their lack of guidance in the infrared field.

Different from traditional method, our formulation does focus on simulated IR images refinement with GANs, which is the first attempt to present an end-to-end deep adversarial learning method for this task. In this paper, we propose a Synthetic Infrared Refinement Generative Adversarial Network (SIR-GAN) to refine simulated IR object. Different from Dual-gan [7] and Cycle-gan [8], the proposed SIR-GAN model can effectively learn bidirectional mappings between real and simulated IR domain, by taking infrared radiation and edge features into consideration. Besides concerns about the adversarial loss and the cycle consistency loss, we construct a new loss function that extra adds a SIR refinement loss including the infrared loss and the structure loss. The SIR refinement loss can guide two Generators and Discriminators to keep the target structure and only transfer thermal features. The whole pipeline including training and testing is shown in Figure 1. Firstly, we prepare two training sets: one is a small set of the real IR images captured by a specific thermal imager, the other one is a large set of the simulated IR set generated by our IR Simulator. Then, the proposed SIR-GAN model trains on the two unpaired domains, where domain S contains a large number of simulated IR images, where domain R contains a small number of real IR images. For our task, the Generator G_{S2R} with the mapping from source domain S to target domain R is utilized to refine the test simulated IR data when testing. In this method, only a small amount of unpaired data is needed, greatly reducing the difficulty of data collection.

In summary, our paper makes the following contributions:

(1) To our knowledge, this is the first demonstration of GANs successfully refining simulated IR target images. The proposed SIR-GAN algorithm can easily learn the bidirectional mappings between two unpaired and imbalanced domains. The target domain is a limited number of real IR images, while the source domain is a large quantity of simulated IR images that can be automatically produced by the IR simulator. This study provides new insights into using limited real IR data to generate a large amount of simulated IR data.

(2) To further bridge the gap between simulated and real IR data, the infrared radiation and structure features of the IR target is taken into consideration. We study an Infrared loss to take account of the local infrared details, and a Structure loss to protect the semantic information. The aforementioned two losses are referred as the SIR refinement loss. To this end, we construct a cumulative loss function, consisting of a SIR refinement loss, an adversarial loss and a cycle consistency loss.

(3) To stabilize training and prevent artefacts, we develop a training strategy with the Discriminators initialization, add the U-net and Dilated convolution in the architecture of the Generators and the Discriminators, and invoke the Nesterov-accelerated Adam as the optimizer, so that the refined IR images can keep the target structure and only be transformed in thermal features.

(4) Qualitative, quantitative and ablation study experiments illustrate that the proposed method brings a significant improvement to the realism of the IR simulator output, compared with the state-of-the-art techniques. Furthermore,

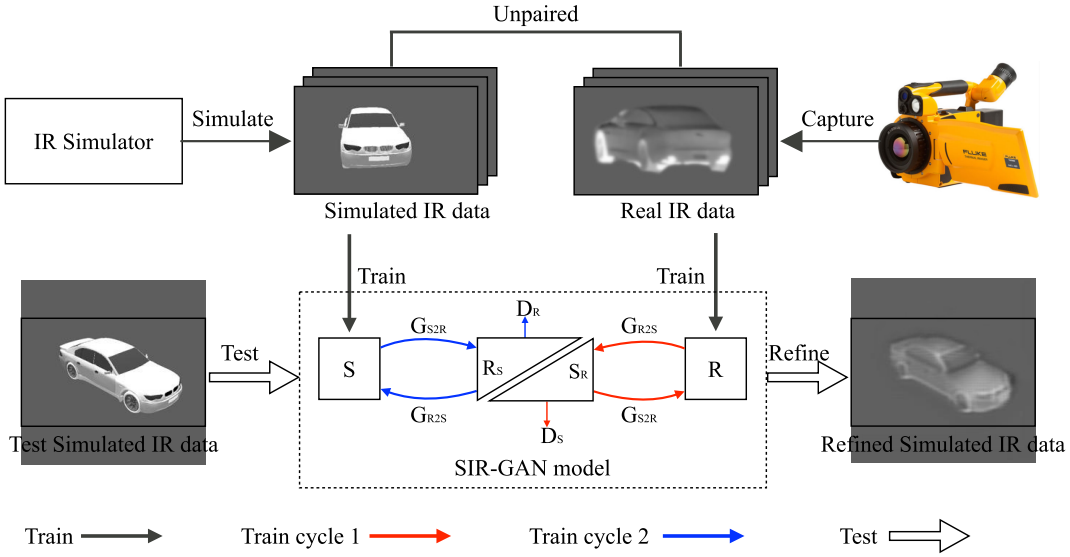


FIGURE 1. The whole pipeline. In the training step, the SIR-GAN model includes two learning cycles:(1) $S \xrightarrow{G_{S2R}} R_S \xrightarrow{G_{R2S}} S$; (2) $R \xrightarrow{G_{R2S}} S_R \xrightarrow{G_{S2R}} R$. The training data consists of two unpaired domains: (1) a large quantity of simulated IR images are generated by IR simulator, and (2) a limited number of real IR images are captured by a thermal sensor. For testing, the well-trained G_{S2R} is used to refine the test samples.

we implement a utility evaluation study to investigate whether the refined IR images are practically usable or not. It is interesting to note that adding our refined data to a training set of a limited amount of real data, the accuracy of the trained classifier will be significantly improved. The findings should make an important contribution to the field of IR image recognition with limited data.

The rest of the paper has been organised in the following way. Section II briefly reviews related work on IR simulator and synthetic-based system, synthetic IR refinement, and generative adversarial networks. After introducing the real and simulated IR datasets in Section III, we illustrate the proposed SIR-GAN framework in Section IV. Then we exhibit the experimental validation of the whole method in Section V. Finally, we conclude the paper with future directions in Section VI.

II. RELATED WORK

A. IR SIMULATOR AND SYNTHETIC-BASED SYSTEM

Several graphics rendering engines are well-suited to infrared simulation [18], [21], [28], [29]. For instance, Lahoud *et al.* [18] use Unity3D and Oculus Rift to build an IR augmented reality system, which can simulate a thermal camera. In addition, through calculating the radiation model, object-oriented graphics rendering engine (OGRE) is also utilized to simulate a real three-dimensional infrared complex scene, which is developed by [29]. On the basis of Mu's work, Gao and Zhang [21]–[23] design an infrared scene simulator to generate thousands of simulated IR images, which can be used to train classifiers and detectors [28]. Meanwhile, for the scene simulation, Guo *et al.* [24] produce a semi-automatic system to simulate large-scale IR urban scenes in the form of levels of detail. Xiong *et al.* [25] propose a simplified watertight module through piece wise planar 3D reconstruction from raw meshes simulated by multi-view stereo. Zhang *et al.* [26] build an infrared scene system to fulfil the real-time and accuracy requirements. Yang and Lv [27] carry out the dynamic IR simulation based on Vega and its IR module.

Many efforts have explored using these generated synthetic data for various prediction tasks, including radar data classification, IR target detection, and semantic segmentation. Karabacak *et al.* [17] simulate micro-Doppler signatures and use the simulated signatures as a source of a priori knowledge to improve the classification performance of real radar data, particularly in the case when the total amount of data is small. Zhang *et al.* [21] train a vehicle detector on a mixed dataset, containing real and simulated IR images with a specific ratio. The experiments show that simulated IR images play an important role in improving the precision of target detection. Richter *et al.* [19] propose that a semantic segmentation model trained only with synthetic data can even outperform the model trained with real data if the amount of synthetic data is large enough. Johnson *et al.* [20] develop a method to incorporate photo-realistic computer images from a simulation engine to rapidly generate annotated data that can be used for the training of machine learning algorithms. As a result, they show that the model trained with a large number of synthetic images outperforms the model trained with a small number of real images. Thus, it is distinct that synthetic data can be used as substitutes when it comes to lack of real data.

B. SYNTHETIC IR REFINEMENT

Traditional methods are based on the infrared calculation, which requires a huge amount of manual work. Liu *et al.* [9] study an infrared radiation calculation strategy based on the principle of the simulation of the infrared thermal image, for more accurate IR texture calculation. Wang *et al.* [10] introduce a portable infrared/visible composite target simulator, which adopts the reflective optical design, through the combination of blackbody and visible light source system design and a variety of target plate, to provide $0.4\mu m \sim 12\mu m$ band simulation target. Liu *et al.* [11] develop a fast numerical simulator for infrared thermography testing (IRT) by using the database of unflawed IRT information, to accelerate the calculations. Ren *et al.* [12] propose a method for infrared 3D scene building based on pseudo color and infrared particle effects includes decoy projectile and smoke. Chengpo *et al.* [29] refine a 3D infrared scene, through calculating thermal radiation model, object-oriented graphics rendering engine. These methods are only effective for specific scenarios. However, they fail to automatically learn a mapping between synthetic and real IR images, and require a huge amount of work in each scene. Our work is essentially different from these approaches, where we improve the realism of the simulator using deep learning.

C. GENERATIVE ADVERSARIAL NETWORKS

GANs [16] introduce the concept of adversarial learning between the generator and the discriminator. The generator and the discriminator act as adversaries with respect to each other to produce real-like samples. The generative adversarial networks learn two models (a generator and a discriminator) with competing losses. The goal of the generator network is to map a random vector to a realistic image, whereas the goal of the discriminator is to distinguish the generated from the real images. GANs have achieved impressive results in image generation [37], image colorizing [42], [44], and image-to-image translation [41], [43]. This inspires us to design a generative model based on simulated IR images to improve the reality of simulated IR images. Meanwhile, we adopt an adversarial loss to learn the mapping, so that the refined IR images cannot be distinguished from images in the target domain.

The last few years have witnessed a variety of GANs with convolutional neural networks (CNNs) developed for image-to-image translation tasks. The feed-forward CNNs are able to be easily trained by using the standard back-propagation approach [45], and the transformed images are generated by forwardly passing the input image through the well-trained CNNs when testing. Pix2pix framework [41] is a successful pipeline to transform image style from a specific image to another specific image. Similar ideas have been applied to various tasks such as generating photographs from attributes and semantic layouts [47] or from sketches [46]. However, it can not transform between two domains, such as our task, from simulated IR domains to real IR domains, because pix2pix model needs a pair of training samples, one as input, another as ground truth. Unlike the above prior work, we learn the mapping without paired training examples.

More recently, several other approaches tackle the unpaired training. CoGAN [48] employs a weight-sharing method to learn a common representation across different domains. Liu *et al.* [49] extend the above baseline by combining GANs and variational autoencoders [50]. Another line of concurrent study is Cycle-gan [8], which designs a cycle architecture with a combination of adversarial losses and cycle consistency. Shrivastava *et al.* [35] propose Simulated+Unsupervised (S+U) learning, which is the first method that uses synthetic eye images to improve the performance of learning algorithms, by using generative adversarial networks, but it fails to learn mappings between two domains. However, none of these GANs models cannot tackle the simulated IR refinement task, owing to limitations of these algorithms and their lack of guidance in the infrared field. In contrast, we propose an end-to-end solution that does refine the simulated IR images.

III. DATA PREPARATION

Before introducing the SIR-GAN model, we briefly present the dataset. The proposed method learning bidirectional mappings between real IR domain and simulated IR domain. The real IR images are captured by a specific thermal imager, while the simulated IR images are rendered by the IR target simulation system.

A. REAL IR IMAGES

The experimental site is a typical plain area, where longitude and latitude are $118°43'22''E$ and $44°54'82''N$, where the altitude is 1190 meters. The thermal sensor to capture real IR images is FlexCam Expert IR thermal imager TiX660, where IFoV is 0.8 mRad, and image resolution of output is 320*240. The collected real IR vehicle dataset includes 800 samples. For real IR dataset establishment, we manually segment vehicles on each captured thermal images and save them separately, in order to exclude background interference and gain the undisturbed IR target images.

B. SIMULATED IR IMAGES

All the simulated IR images are rendered by the IR target simulation system, which consists of three parts: (1) we build a three-dimensional geometric model of a specific target; (2) through calculation of target thermal radiation model, infrared texture of the target can be inferred and mapped on the 3D target model; (3) after OGRE rendering with atmospheric effect model, a simulated IR target image is generated with the target label. Since we only focus on IR target modeling, background modeling is not taken into consideration. In this paper, we simulate an IR vehicle as a maneuvering target, as shown in Figure 2.

All objects emit infrared energy (heat) as a function of their temperature. The infrared energy emitted by an object is known as its heat signature. In general, the hotter an object

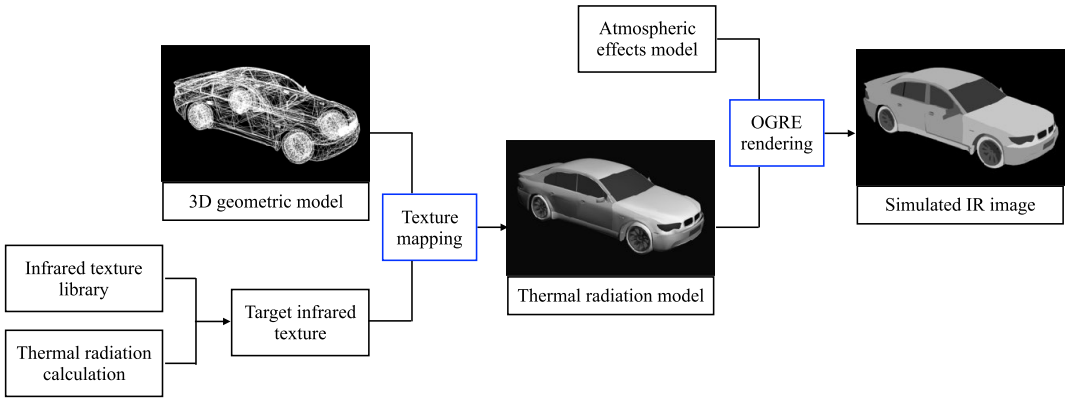


FIGURE 2. The pipeline of IR target simulation system. Firstly, a three-dimensional geometric model is built. After calculating thermal radiation of the target, infrared textures are mapped on the 3D target model. To the end, through OGRE rendering, the simulated IR target image is generated.

is, the more radiation it emits. A thermal imager is essentially a heat sensor that is capable of detecting tiny differences in temperature. Thus, an infrared image can reflect the target's temperature distribution, which depends on its infrared radiation characteristic. Actually, the infrared radiation characteristics are closely related to the heat transfer between the target and environment. When calculating the thermal radiation of the target, besides the target's own radiation, we should take the impact of environmental radiation into consideration, like solar radiation, sky background radiation and ground radiation.

We set up the IR target simulation system in the same environment as the captured real IR images, and adopt the same geographic data. The parameters used to calculate the IR radiation model are as follows, weather: cloudy with stratocumulus; environment temperature: $15.2°C$; humidity: 50.0%; wind speed: 2.7 m/s; target surface temperature: $22.8°C$; wave band: $8 \sim 14\mu m$. We employ Visual Studio 2017 as the development environment, through the OGRE texture mapping and rendering, to generate the simulated IR images. Please see the Appendix A for more details about the infrared radiation calculation.

IV. SIMULATED INFRARED REFINEMENT GENERATIVE ADVERSARIAL NETWORK

As mentioned before, the IR target simulation system can generate an arbitrary amount of synthetic images with annotations. However, there is still a gap between these simulated and real IR images. In this section, we propose a synthetic IR refiner named SIR-GAN model to refine these synthetic data, so as to bridge the gap. From Figure 1, we introduce a whole pipeline of the synthetic data refiner in both training and testing. We use a large number of simulated IR images and a small amount of real IR images to train the SIR-GAN model effectively, to make the model learn a mapping between simulated data and real data. What's more, the cycle design of our model can also train on the unpaired training data, which consists of a source set $S = \{s_i\}_{i=1}^{N}$ and a target set $R = \{r_j\}_{j=1}^{M}$, with no information provided as to which s_i matches which r_i [8].

A. OVERALL OF SIR-GAN

The SIR-GAN consists of two Generators and two Discriminators. The Generator model is responsible for generating the image, and constantly make the generated 'fake' image closer to the target dataset, in order to achieve to cheat Discriminator model. The Discriminator model should reinforce identification ability, and distinguish between real samples and 'fake' samples to guide the generation process of the Generator model.

Inspired by the cycle structure of [8] for image-to-image translation, we use the similar design in our SIR-GAN model. The goal is to learn a bidirectional mapping function between two domains S (simulated IR data) and R (real IR data) provided training samples $\{s_i\}_{i=1}^{N}$ where $s_i \in S$, and $\{r_j\}_{j=1}^{M}$ where $r_j \in R$. The data distribution is donated as $s \sim p_{data}(s)$ and $r \sim p_{data}(r)$. As illustrated in Figure 1, our SIR-GAN model consists two mappings $G_{S2R} : S \rightarrow R$ and $G_{R2S} : R \rightarrow S$. Then, we introduce two adversarial discriminators D_R and D_S, where D_R aims to discriminate between images $\{r\}$ and refined images $\{G_{S2R}(s)\}$; in the same way, D_S aims to distinguish between $\{s\}$ and $\{G_{R2S}(r)\}$.

Then, we introduce the detailed process and architecture of our SIR-GAN model step by step, as shown in Figure 3.

(1) Firstly, we put a sample s_i of the simulated IR dataset S into the Generator G_{S2R}, to output a generated image $r_{si} = G_{S2R}(s_i)$. Then, the Discriminator D_R distinguishes whether r_{si} belongs to the simulated IR dataset S or the real IR dataset R. The judgment will be fed back to G_{S2R}, so as to reinforce the generator to generate a more realistic image. Thus, under the supervision of D_R, r_{si} will continuously narrow the gap with the image in R. Finally, we put r_{si} into the Generator $\{G_{R2S}\}$ to generate a new image $s_i' = G_{R2S}(G_{S2R}(s_i))$, which is expected to be similar to the samples in R.

(2) On the other hand, an image r_j of the real IR dataset R is put into the Generator G_{R2S}, to output a generated image

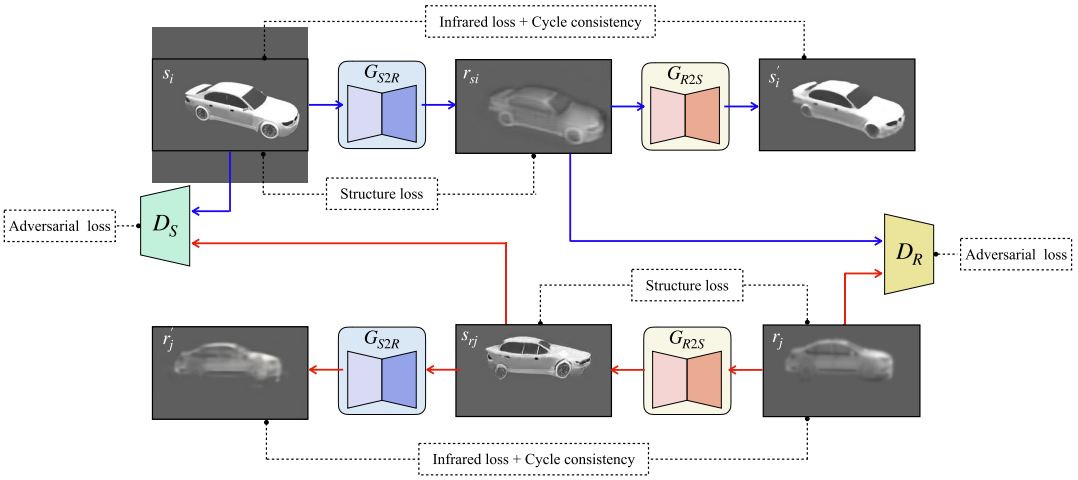


FIGURE 3. Network architecture and data flow chart of SIR-GAN. The blue line represents a cycle from domain *S* to domain *R*, and back to domain *S*. The red line represents a cycle from domain *R* to domain *S*, and back to domain *R*. The dotted line means the losses.

$s_{rj} = G_{R2S}(r_j)$. After the Discriminator D_R distinguishing whether s_{rj} belongs to S or R, the result will be fed back to G_{R2S} to update the generator. Under the supervision of D_S, s_{rj} should be close to the image in S. After s_{rj} putting into the Generator $\{G_{S2R}\}$, the generated image r_j' should be expected to be close to the samples in S.

That is to say, Generators not only can generate real IR images from simulated IR images, but also have the ability to reduce from real IR images to simulated IR images. This specific cycle design make Generators and Discriminators combat each other. D_R guides G_{S2R}, D_S trains G_{R2S}, consequently two Generators can generate more similar images to the target dataset. Meanwhile, Generators in turn train Discriminator to continuously enhance the discriminating ability. For s_i' and r_j', they can be treated as verification for whether G_{S2R} and G_{R2S} can generate enough quality or not.

B. LOSS FUNCTION

Our objective includes three types of terms: the SIR adversarial losses, the SIR cycle consistency losses, and the SIR refinement losses. (1) SIR adversarial losses for G_{S2R} aim to match the distribution of refined images to that of simulated images, and it is the same reason for G_{R2S}. (2) SIR cycle consistency losses aim to prevent the learned mapping G_{S2R} and G_{S2R} from contradicting each other. (3) SIR refinement losses which include structure losses and infrared losses, aim to not only reserve object structure information, but also transfer the infrared texture, in order to strengthen the generation ability of G_{S2R}.

1) SIR ADVERSARIAL LOSS

We apply adversarial losses to both mapping $G_{S2R} : S \rightarrow R$ and $G_{R2S} : R \rightarrow S$. We define the G_{S2R} objective as Equation (1), and the excepted result is $s \rightarrow G_{S2R}(s) \approx r$.

$$\begin{aligned}&\mathcal{L}_{S2R}(G_{S2R}, D_R, S, R)\\&\quad = E_{r\sim p_{\text{data}}(r)}[\log D_R(r)]\\&\qquad + E_{s\sim p_{\text{data}}(s)}[\log(1 - D_R(G_{S2R}(s))]\end{aligned} \tag{1}$$

where G_{S2R} tries to generate image $G_{S2R}(s)$ which looks similar to real IR images of domain R, and D_R aims to discriminate between $G_{S2R}(s)$ and r. G_{S2R} and D_R compete with each other, to minimize the generation error of G_{S2R} and maximize the discrimination ability of D_R. In the same way, we introduce the $G_{R2S} : R \rightarrow S$ objective as:

$$\begin{aligned}&\mathcal{L}_{R2S}(G_{R2S}, D_S, R, S)\\&\quad = E_{s\sim p_{\text{data}}(s)}[\log D_S(s)]\\&\qquad + E_{r\sim p_{\text{data}}(r)}[\log(1 - D_S(G_{R2S}(r))]\end{aligned} \tag{2}$$

2) SIR CYCLE CONSISTENCY LOSS

The aim is that the image translation should be able to bring s to r_s and back to domain S, for instance, $s \rightarrow G_{S2R}(s) \rightarrow G_{R2S}(G_{S2R}(s)) \approx s$. If the input s and the refined r_s do not share semantic information, it is impossible to regenerate the input by using the refined image. Thus, by forcing the learned transformation to have an effective inverse refinement, the generated image can be further forced to share semantics with the input. If the generated $G_{S2R}(s)$ can be recovered to a simulated IR image, we can believe that the main structure information of input image s is well-preserved, like object's outline, shape, and orientation. In other words, only the infrared information is transferred in the whole process. As illustrated in Figure 3, for each sample r from domain R, G_{R2S} and G_{S2R} should satisfy: $r \rightarrow G_{R2S}(r) \rightarrow G_{S2R}(G_{R2S}(r)) \approx r$. We express this cycle consistency loss to ensure that the transformed image shares semantics with the input image, in Equation (3), where L1-normalize is employed in this loss.

$$\begin{aligned}&\mathcal{L}_{cyc}(G_{S2R}, G_{R2S})\\&\quad = E_{s\sim p_{\text{data}}(s)}[\|G_{R2S}(G_{S2R}(s)) - s\|_1]\\&\qquad + E_{r\sim p_{\text{data}}(r)}[\|G_{S2R}(G_{R2S}(r) - r\|_1]\end{aligned} \tag{3}$$

3) SIR REFINEMENT LOSS

To further strengthen the infrared refinement ability of Generator G_{S2R} and reduce the loss of semantic information,

we propose an Infrared loss $\mathcal{L}_{ir}$ and a Structure loss $\mathcal{L}_{strc}$ to form the SIR refinement loss:

$$\mathcal{L}_{ref} = \mathcal{L}_{ir} + \mathcal{L}_{strc} \tag{4}$$

a: INFRARED LOSS

As the consistency loss with global L1 focuses on the entire image space, it ignores many local infrared details, which are critical in infrared images. Since the grey value of an IR image depends on the infrared radiation of the target, we try to make the input and the transformed images consistent in infrared radiation, to ensure the successful transmission of infrared information. In order to further improve the quality of the refined images regarding infrared details, we propose an infrared loss to restore the infrared radiation information in the refined IR image.

$$\begin{aligned}&\mathcal{L}_{ir}(G_{S2R}, G_{R2S})\\ &= E_{s\sim p_{\text{data}}(s)}\left[\|\Phi\left(G_{S2R}(s)\right) - \Phi(s)\|_1\right]\\ &\quad + E_{r\sim p_{\text{data}}(r)}\left[\|\Phi\left(G_{R2S}(r)\right) - \Phi(r)\|_1\right],\end{aligned} \tag{5}$$

where the loss is raised from the infrared radiation value metric:

$$\Phi = \Phi_{min} + \frac{(g/255 - r) \times (\Phi_{max} - \Phi_{min})}{1 - r}, \tag{6}$$

where g is the grey level of each pixel of an image. The maximum and minimum radiation intensity of the target is Φ_{max} and Φ_{min}. r is constant and follows $0 \leqslant r \leqslant 1$, which depends on the specific scene and the type of the IR thermal imager.

b: STRUCTURE LOSS

In order to retain the semantic information, the target structure of input images should be maintained in the refinement process. The gradient correlation defined by the normalized cross correlation between two images, is used to predict the structures of both input images and the generated images. The structures can be regard as the prior knowledge guiding better IR image generation. Given gradients in horizontal and vertical directions of two images, s_i and r_{si}, GC is formulated as:

$$GC(s_i, r_{si}) = \frac{1}{2}[NC\left(\nabla_x s_i, \nabla_x r_{si}\right) + NC\left(\nabla_y s_i, \nabla_y r_{si}\right)], \tag{7}$$

where,

$$NC(s_i, r_{si}) = \frac{\sum_{(i,j)}(s_i - \overline{s_i})(r_{si} - \overline{r_{si}})}{\sqrt{\sum_{(i,j)}(s_i - \overline{s_i})^2}\sqrt{\sum_{(i,j)}(r_{si} - \overline{r_{si}})^2}}, \tag{8}$$

and ∇_x and ∇_y are the gradient operator of each direction, $\overline{s_i}$ is the mean value of s_i. We formulate the structure loss $\mathcal{L}_{strc}$ with L1 distance as:

$$\begin{aligned}&\mathcal{L}_{strc}(G_{S2R}, G_{R2S})\\ &= \frac{1}{2}E_{s\sim p_{\text{data}}(s)}[1 - GC\left(s, G_{S2R}(s)\right)]\\ &\quad + \frac{1}{2}E_{r\sim p_{\text{data}}(r)}[1 - GC\left(r, G_{R2S}(r)\right)].\end{aligned} \tag{9}$$

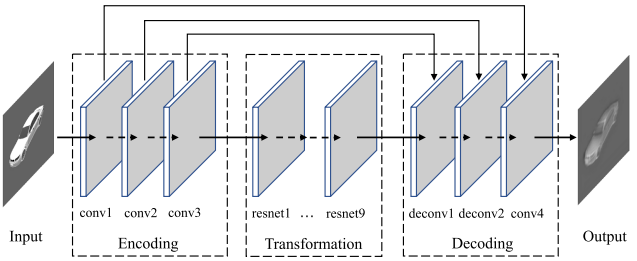


FIGURE 4. The architecture of the Generator with U-net.

4) CUMULATIVE LOSS

Overall, we define the full loss function as follows:

$$\begin{aligned}\mathcal{L}(G_{S2R}, G_{R2S}, D_R, D_S) = \lambda_1[&\mathcal{L}_{S2R}(G_{S2R}, D_R, S, R)\\ &+ \mathcal{L}_{R2S}(G_{R2S}, D_S, R, S)]\\ &+ \lambda_2\mathcal{L}_{cyc}(G_{S2R}, G_{R2S})\\ &+ \lambda_3\mathcal{L}_{ref}(G_{S2R}, G_{R2S})\end{aligned} \tag{10}$$

The hyper-parameter $\lambda_1, \lambda_2, \lambda_3$ control the relative importance of the three losses. Each of the three terms has a loss weight indicated to adjust the importance of each loss part. In the Cycle-gan [8], it fixes the weights of the adversarial loss and the cycle consistency equal. In addition, we find that the susceptibility to mode collapse is different when facing different tasks, so we introduce $\lambda_1, \lambda_2, \lambda_3$ as the hyper-parameter. All experiments use $\lambda_1 = 1$, $\lambda_2 = 5$, $\lambda_3 = 2$, to make our network focus on the simulated IR images refinement.

C. NETWORK ARCHITECTURE

The SIR-GAN model consists of 2 Generators (G_{S2R} and G_{R2S}) and 2 Discriminators (D_R and D_S). Besides the new loss function, several vital modifications are adopted to the architecture and the training step to stabilize training and prevent the network from producing artefacts, compared to other GANs. Thus, the refined IR images can retain the target structure and transfer thermal information.

1) GENERATOR ARCHITECTURE

G_{S2R} and G_{R2S} have the same architecture. We adopt the networks from [8], and modify it by using U-type cross-connection (U-net) method [30]. The motivation is that the Generator is expected only to transform infrared feature but not change the target structure information. U-net is a usual contracting network by successive, which contains a large number of feature channels in the upsampling part. These channels allow the network to propagate context information to higher resolution layers. The same idea is employed in our model. The architecture of Generator is illustrated in Figure 4. The input and output image is 256 × 256. The Generator includes 4 convolutional layers, 9 resnet convolutional layers [31], and 2 deconvolutional layers. Convolution and deconvolution are the opposite operations. The structure of resnet block includes two convolutional layers, and the feature map plus input equals output, as shown in Figure 5. The first

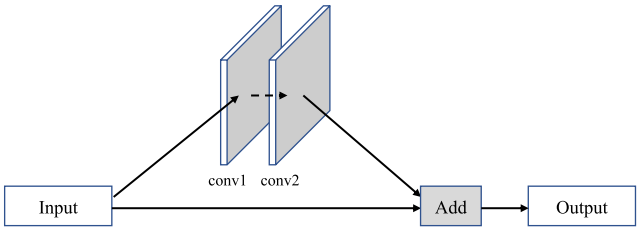


FIGURE 5. The structure of the resnet block.

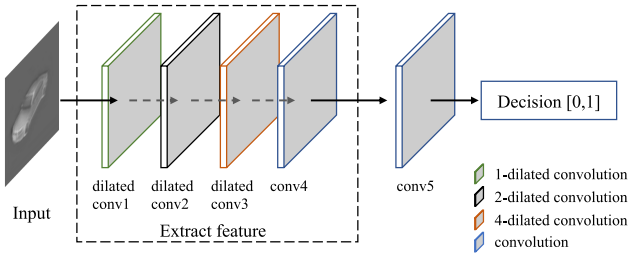


FIGURE 6. The architecture of the Discriminator with dilated convolution.

three convolutional layers are regarded as the encoding step, while two deconvolutional layers and one convolutional layer can be treated as the decoding step. The Generator G_{S2R} is only used to refine simulated IR images to real IR images, rather than changing other features of the vehicle, such as the outline and orientation of the vehicle. Thus, besides sequential convolution operation, we add three joint feature map fusion in $conv1 \rightarrow conv4$, $conv2 \rightarrow deconv1$ and $conv3 \rightarrow deconv2$. This kind of joint feature map can protect the shape, outline, orientation and other fundamental features, and only transform the infrared features of the target. After each convolutional layer, Instance Normalization [32] is used instead of Batch Normalization to achieve normalization for a single image, thereby improving the quality of the generated images based on accelerating model convergence and preventing gradient explosion.

2) DISCRIMINATOR ARCHITECTURE

The input is a 256 × 256 image, and the output is the result of binary classification, where 0 means the class of input is not the target class, where 1 means the class of input is the target class. The input is put into three dilated convolutional layers with different kernels, followed by two standard convolutional layers. Different from [8], our discriminators adopt three varying dilated convolutional layers instead of standard convolutional layers. The motivation is that we find that the object may be deformed during training, because spatially hierarchical information is easily lost in the standard convolution. In order to solve the deformation problem, an effective module named dilated convolution is used to increase the receptive field corresponding to the pixel points in the last layer. Unlike traditional convolution kernels, dilated convolution has a 3 × 3 convolution kernel with a certain interval. In the left figure of Figure 7, there is a 1-dilated convolution of 3 × 3, and the red dot position

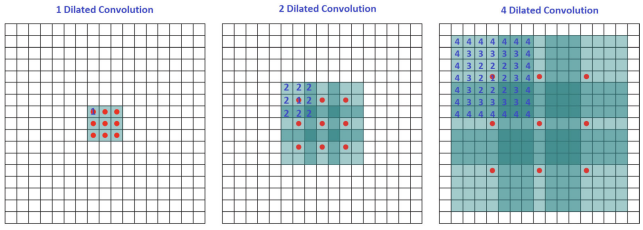


FIGURE 7. Dilated convolution with a kernel size of 3 × 3 in different dilation rates.

represents the sampling point, which is the same as the standard convolution operation. In the middle of Figure 7 illustrates a 2-dilated convolution of 3 × 3. We can see that each sample point in the convolution kernel has a unit interval, so that a combination of 1-dilated convolution and 2-dilated convolution can reach a 7 × 7 receptive field. The right of Figure 7 shows the 16 × 16 receptive field achieved by the combination of 1-dilated convolution, 2-dilated convolution, and 4-dilated convolution. The reason we replace the standard convolutional layers with dilated convolutional networks is listed below. (1) Deep convolutional neural networks consist of convolutional layers and pooling/up-sampling layers. The weights of convolutional layers are learnable, while pooling/up-sampling layers are deterministic. The dilated convolutional layer can be understood as a learnable mixed layer that combines a convolutional layer with a pooling layer. (2) Compared to three dilated convolutional layers of 16 × 16 receptive field, the three standard convolutional layer combination of the same 3 × 3 receptive field convolution kernel only has a receptive field of 7 × 7. The wider receptive field will have a positive influence on discrimination and reduction of the input image. (3) The standard CNNs may cause loss of internal data structure and spatial hierarchy information, which is especially important in the simulated IR refinement task. Because the shape and outline information of the target is expected to be retained. Similarly, we also use Instance Normalization instead of Batch Normalization after each convolutional layer. Note that, the PatchGAN [41] is employed in this architecture as the classifier, which tries to distinguish whether the input image is natural or generated through such $N \times N$ patch in conv5.

D. TRAINING PROCEDURE WITH UNPAIRED AND IMBALANCED DATA

The sum of training data for SIR-GAN consists of two unpaired and imbalanced datasets. One is the real IR vehicle dataset including 800 samples captured by a specific thermal sensor. The other one is the simulated IR vehicle dataset which includes 5000 samples. These samples are built easily through varying view of the target in the IR simulator. The SIR-GAN model learns the bidirectional mappings between simulated and real IR domains. Consequently, the training process does not need one-to-one correspondence between input and output, to tackle the unpaired learning. On the other hand, unbalanced data will cause insufficient training. To deal

Algorithm 1 The SIR-GAN Training Procedure, Using the Nesterov-Accelerated Adam Optimizer

Require:
The set of simulated IR images for current batch, $s_i \in S$;
The set of real IR images for current batch, $r_j \in R$
The number of training batch iterations $\mathcal{N}$ with a batch size of $\mathcal{B}$;

Ensure:
Update Generator and Discrimitor weights $\theta_{G(S\to R)}$, $\theta_{G(R\to S)}$, $\theta_{D(S)}$, $\theta_{D(R)}$;

initialize network parameters $\theta_{G(S\to R)}$, $\theta_{G(R\to S)}$, $\theta_{D(S)}$, $\theta_{D(R)}$;
for $n = 1$ to 20 **do**
  pretrain D_S as a binary classifier with S and R, to discriminate simulated IR images;
  pretrain D_R as a binary classifier with R and S, to discriminate real IR images;
end for
for $n = 21$ to $\mathcal{N}$ **do**
  pick a sample s_i from S, and put it into G_{S2R}, then the model outputs a generated image $r_{si} = G_{S2R}(s_i)$;
  the D_R discriminates the generated image r_{si};
  put r_{si} into G_{R2S} to generate a new image $s_i' = G_{R2S}(G_{S2R}(s_i))$;
  calculate the error between r_{si} and R;
  calculate the error between s_i' and S;
  for $b = 1$ to $\mathcal{B}$ **do**
    pick a sample r_j from R, and put it into G_{R2S}, to generate $s_{rj} = G_{R2S}(r_j)$;
    put s_{rj} into D_S to distinguish between the simulated and the real;
    send s_{rj} into G_{S2R} to generate a new one $r_j' = G_{S2R}(G_{R2S}(r_j))$;
    calculate the error between s_{rj} and S;
    calculate the error between r_j' and R;
  end for
  average the batch of the error between s_{rj} and S;
  average the batch of the error between r_j' and R;
  update $\theta_{G(S\to R)}$, $\theta_{G(R\to S)}$, $\theta_{D(S)}$, $\theta_{D(R)}$, according to Equation (10).
end for

with it, we use the mean error of a batch size with 1 real sample and 6 simulated samples, so that all the simulated IR samples can be evenly used. Different from other GANs, we initialize the Discriminator as a binomial classifier to distinguish real and simulated IR images. This operation can enlarge the gradients of the Generators when starting bidirectional training, so as to accelerate convergence.

For the parameter setting of training, we set $\lambda_1 = 1$, $\lambda_2 = 5$, $\lambda_3 = 2$ in Equation (10). We adopt the Nesterov-accelerated Adam solver [51] with a batch size of 6, $\mu = 0.975$, $\upsilon = 0.999$, $\eta = 1e^{-8}$. We do not use any pre-trained networks. The networks of SIR-GAN are trained from scratch. We keep the learning rate of 0.0002 in the first 100 epochs, and then linearly decay the learning rate to 0 for the next 100 epochs. We introduce the training procedure step by step as shown in Algorithm 1. After many times of the error back propagation, the Generators can generate the images whose distribution approximates to the target domain, while the Discriminators are unable to distinguish the generated images.

E. INFERENCE PROCEDURE

During training, we train the SIR-GAN model effectively, which includes 2 Generators (G_{S2R} and G_{R2S}) and 2 Discriminators (D_R and D_S). For testing, we only select the trained G_{S2R} to refine the test samples. All the settings are the same as the training step. The test image is put into the well-trained G_{S2R}, to generate a new image which is expected to be a real IR image. And we call this process as the simulated IR image refinement.

V. EXPERIMENTS AND DISCUSSIONS

In this section, we firstly introduce the experimental setup, and then present the results of our SIR-GAN model in refining the simulated IR images.

A. EXPERIMENTAL SETUP

For the SIR-GAN model, we implement the Tensorflow [33] framework for all inference, training and testing, in a powerful server (2.7-GHz 4-core CPU, 16G RAM, $4 * 8$GB GPU and Ubuntu 16.04). The whole training costs 10 hours on four NVIDIA 1080Ti Pascals.

The sum of training data for SIR-GAN consists of two datasets. One is the real IR vehicle dataset including 800 samples. For real IR dataset establishment, we manually segment vehicles on each captured thermal images and save them separately, in order to exclude background interference and gain the pure IR target images. The other one is the simulated IR vehicle dataset which includes 5000 samples. These samples are build easily through varying view of the target in our IR simulator.

B. EVALUATION METRICS

Evaluating the quality of synthesized images is an open and challenging problem [34]. Traditional metrics such as per-pixel mean-squared error do not assess joint statistics of the result, and therefore do not measure the very structure that structured losses aim to capture. To better illustrate the performance of the proposed method, we evaluate the results with three tactics.

1) Visual and quantitative study. The results are displayed to give an intuitive evaluation. Meanwhile, we calculate the mean grey value, grey variance, contrast, and infrared cross entropy of each domain on the different methods.
2) AMT perceptual study. We run "real v.s. fake" perceptual studies on Amazon Mechanical Turk (AMT).

TABLE 1. The visual indicators of all baselines, where value is on the left and relative error is on the right.

	Grey value		Grey variance		Contrast		IEC
Method	**value**	**relative error**	**value**	**relative error**	**value**	**relative error**	
Real	107.3	-	530.1	-	0.158	-	-
Simulated	120.8	12.58%	785.6	48.19%	0.212	34.17%	9.74
SIR-GAN	105.4	**1.77%**	420.5	**20.67%**	0.160	**1.26%**	**5.91**
DCGAN	117.5	9.50%	650.7	22.75%	0.198	25.31%	9.11
Cycle-gan	111.7	4.10%	640.5	20.82%	0.188	18.98%	7.25
S+U	102.1	4.84%	350.9	33.80%	0.168	6.32%	7.94
NSU	103.5	3.54%	362.4	31.63%	0.166	5.06%	6.33

For graphics colorization and image generation, plausibility to an observer is often the ultimate goal.

3) "FCN-score" study. We measure whether the refined images are realistic enough that off-the-shelf classification system can classify the images or not.

C. COMPARISION WITH STATE-OF-THE-ART METHODS

To evaluate the proposed SIR-GAN model, we compare it with several state-of-the-art baselines: (1) Simulated+ Unsupervised (S+U) learning [35] try to use adversarial training to make simulated eye images more realistic; (2) The new S+U (NSU) Learning with adaptive data [36] generate eye images with pseudo labels for gazing estimation; (3) Cycle-gan [8] introduce a novel method for unpaired image-to-image translation; (4) DCGAN [37] is the first GAN model combined with CNN for image generation. For equality, all the above approaches are all trained on our training data with standard settings.

1) VISUAL AND QUANTITATIVE STUDY

In order to illustrate the result in visual intuition, we first display the result of each method in Figure 8. As mentioned before, the mappings are learning from simulated IR domain to real IR domain, other than from a particular simulated IR image to a corresponding real IR image, consequently there is no ground truth for each simulated IR image, but we present some real samples as a reference in the rightmost column of Figure 8. It is obvious, that our SIR-GAN performs best, and it can refine the simulated IR images effectively. The refined IR images much resemble the real IR images. DCGAN only learns a directional mapping, whose learning ability is too weak to generate realistic images. It is not only unable to learn the infrared mapping, but also changes the structure of the target. Both Cycle-gan and S+U learn some informative mappings including infrared and outline features, but they happen some irregular spots on their results. It may be caused by a specific pattern which is learned from the training set. Although NSU has no such concerns, this model sometimes leads to non-convergence, which means that the model parameters oscillate, destabilize and never converge. In addition, all the baselines easily occur diminished gradient, meaning that the discriminator gets too successful that the generator gradient vanishes and learns nothing. The reason is that our loss function adds a penalty term named SIR refinement loss, and this term can guarantee the generator G_{S2R} to converge to the global optimum.

$$\eta = \frac{|v - v_m|_1}{v} \times 100\% \tag{11}$$

To quantify the results, we count mean grey value, grey variance, and contrast with relative error of each approach as the qualitative results. The relative error is defined in Equation (11), where v is the reference value of real IR set and v_m measured value of synthetic IR set. In addition, cross entropy is a common measurement in information theory to assess the distance of two probability distributions. Inspired by this, we define an infrared cross entropy (ICE) to evaluate the results, as shown in Equation (12). $P_{real}(i)$ and $P_{generate}(i)$ are the probabilities of infrared intensity i in the real IR images and the generated IR images. The smaller value of ICE represents that the generated IR image is more consistent with the real one in infrared intensity distribution.

$$ICE = \sum_{i=1}^{255} P_{real}(i) \log\left[\frac{1}{P_{generate}(i)}\right] \tag{12}$$

In Table 1, we show the assessment value and the relative error, and see that the distance between real domain and the proposed SIR-GAN model is minimal on the 1.77%, 20.67%, 1.26% and 5.91 for grey value, grey variance, contrast, and ICE. These four results of our method are much better than those of IR simulation and other methods. The results of simulated data reasonably has the largest gap on 12.58%, 48.19%, 34.17%, and 9.74. Compared with DCGAN and S+U algorithm, our results look more complete, and their results is blurry and incomplete, in Figure 8. The reason why our model outperforms them is that they use directional mappings while our model utilizes bidirectional mappings. The bidirectional mappings can be trained more effectively and converge more easily than directional mappings. Though Cycle-gan and NSU also employ the cycle consistency design for bidirectional mappings, their results of grey value are too low while their results of grey variances are large, and the outlines of their target are unclear. This is caused by the lack of the U-type module and Dilated convolutional layer, which are benefits for keeping the original information, compared

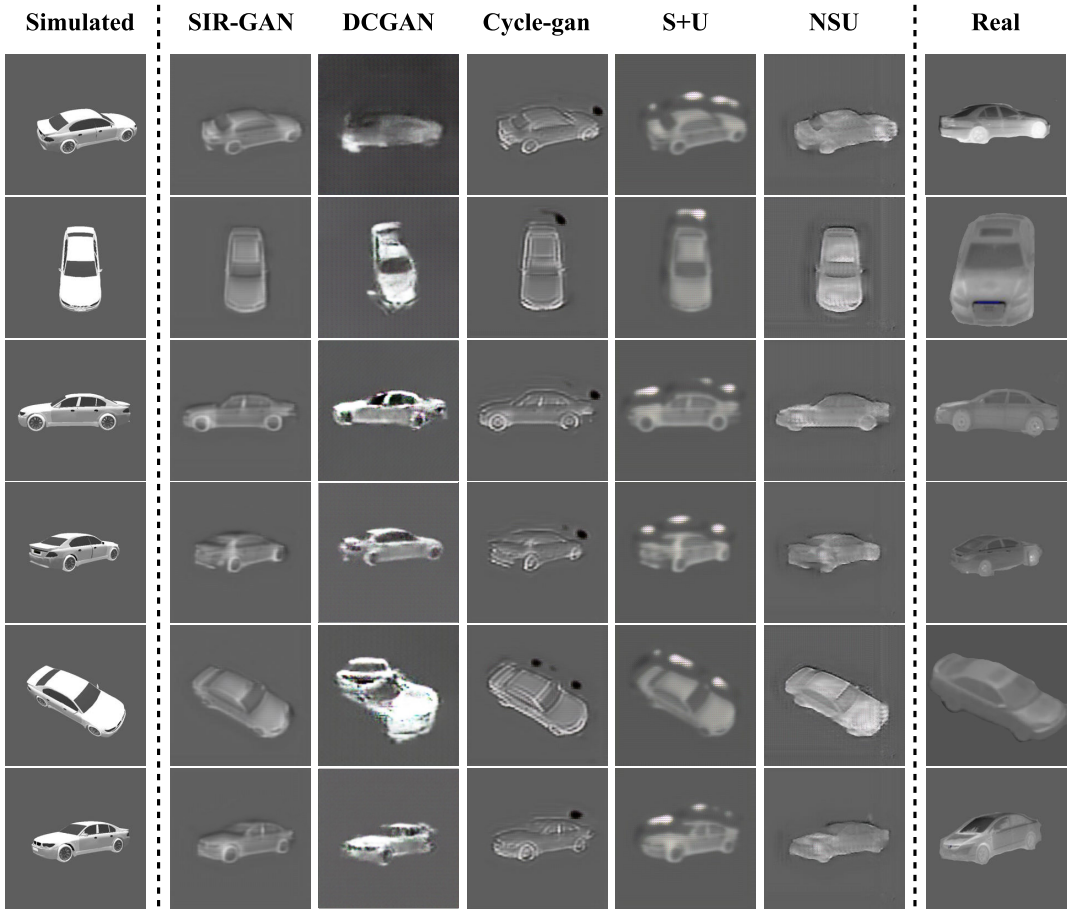


FIGURE 8. Visual presentation of the results in different methods. We select 6 poses of the vehicle. The first column is the simulated IR images, while the last column is the real IR images.

TABLE 2. Training time of different errors on G_{S2R} loss (unit: hour).

Error	SIR-GAN	DCGAN	Cyclegan	S+U	NSU
0.25	**8.2**	14.9	17.4	14.4	16.7
0.20	**9.3**	-	32.3	-	20.1
0.15	**10.6**	-	-	-	-

to the proposed approach. These histogram statistics and visual intuition intuitively illustrate that our algorithm for simulated IR image refinement outperforms other state-of-the-art methods. Although our visual statistics results do not widen the gap with the results of Cycle-gan, S+U, and NSU algorithms, we can see the distinct gap according to Figure 8, and also the next two experiments.

In addition, we carry out the experiments to present the training time of these methods. In Table 2, we show the training time of different errors on G_{S2R} loss, which is the Generator for IR refinement. For the error of 0.25, each algorithm can reach the threshold. The SIR-GAN converges faster than others, at only 8.2 hours. For the error of 0.20, DCGAN and S+U fail to reach the threshold. For threshold 0.15, only the SIR-GAN can achieve the best result, at 10.6 hours.

2) AMT PERCEPTUAL STUDIES

To more holistically evaluate the visual quality of our results, we employ AMT perceptual studies for our experiments. We follow the protocol from [38]: Participants are presented with a sequence of pairs of images, one a real IR image, one simulated IR image(generated by our IR simulator), and told which image is real. Every image appears for one second on each trial, after which the images disappear and participants respond as to which is real. If the subjects find the image very hard to tell the difference between the real images and the refined images, the Turker mark it ''hard to label''. The first 10 trials of each session are practice and participants are given feedback. The remaining 40 trials are utilized to measure the rate at which each algorithm fools participants. Each session tests just one approach at a time, and participants are only allowed to complete a single session.

We validate the perceptual realism of the refined IR images. Results of the AMT study are given in Table 3. For synthetic images hard to label, we find that only our method can confuse Turkers hard to tell real or fake on 42.7%, much larger than labeled fake on 28.9%. In contrast, all the other approach hard to confuse the Turkers. The simulated

TABLE 3. AMT "real vs fake" experiments.

	Simulated	SIR-GAN	DCGAN	Cycle-gan	S+U	NSU
Turkers labeled real (%)	0.3%	**28.4%**	1.4%	8.7%	5.1%	9.8%
Turkers hard to label (%)	1.1%	**42.7%**	4.2%	10.9%	11.2%	18.2%
Turkers labeled fake (%)	98.6%	**28.9%**	94.4%	80.4%	83.7%	72.0%

IR domain and DCGAN model almost never fool subjects. Cycle-gan and S+U algorithms just can confuse less than 20%. NSU model has a relatively good result on 9.8% and 18.2%, which also underperform our method. Simultaneously with Figure 8, it is obvious that our results are much more realistic than others. Especially the quality of the refined IR images have a huge improvement on the basis of the simulated IR images. Here, we see the proposed method can fool participants on around three quarters of trials at 256×256 resolution. Turkers can distinguish most of real from fake on the generated IR images by Cycle-gan, S+U, and NSU algorithm.

3) "FCN-SCORE" STUDY

Although ATM perceptual studies may be the gold standard for assessing graphical realism, recent works [34], [38] have started using pre-trained semantic classifiers to discriminate whether the refined IR image is real or fake. The intuition is that if the refined IR images are realistic, the pre-trained classifiers trained on real IR data can correctly classify the refined IR images as well. For this reason, we adopt the FCN [39] architecture for semantic classification and segmentation, and train it on our real IR training set. Finally, we score refined IR images by the classification accuracy.

In Table 4, we assesses the performance of the simulated IR image $\rightarrow$ refined IR image task on the real IR dataset. The methods employing the cycle-consistenct achieve higher scores, indicating that the refined IR images include more recognizable structure and thermal information. During training, the cycle-consistent-based approaches like SIR-GAN, Cycle-gan, and NSU are able to learn the bidirectional mappings, while DCGAN and S+U only learn the directional mapping. In the step of the bidirectional training, the Generators repeat correcting the network through error back-propagation, whose learning ability and efficiency is several times that of the directional models. With the same cycle-consistent design, our results are much better than those of Cycle-gan and NSU algorithm. This variant results in high performance; examining the results reveals that the proposed network architecture and loss function are more effective on simulated IR refinement task. Clearly, it is significant, in this case, that the loss measures the quality of the match between input and output, and indeed SIR-GAN performs much better than other GANs.

D. ABLATION STUDY

In order to better understand our method, we conduct a series of ablation evaluation experiments. All the results are shown in every subsections and discussed in detail.

TABLE 4. "FCN-score" experiments for different methods.

Method	Per-pixel acc.	Per-class acc.	Class IoU
Simulated	0.08	0.03	0.01
SIR-GAN	**0.59**	**0.23**	**0.16**
DCGAN	0.11	0.06	0.02
Cycle-gan	0.23	0.10	0.07
S+U	0.18	0.08	0.05
NSU	0.23	0.11	0.08

1) LOSS FUNCTION ANALYSIS

The cumulative loss consists of the SIR adversarial loss ($\mathcal{L}_{adv}$), the SIR cycle consistency loss ($\mathcal{L}_{cyc}$) and the SIR refinement loss ($\mathcal{L}_{ref}$), where the SIR refinement loss contains the Infrared loss ($\mathcal{L}_{ir}$) and the Structure loss ($\mathcal{L}_{strc}$). In Table 5, we show the comparison against ablations of all parts of the full loss. We can see that the cumulative loss is superior to any others on both AMT and FCN-scores studies. Removing the SIR refinement loss substantially degrades performance, as does removing the SIR cycle consistency loss. There is a same trend on removing the Infrared loss or the Structure loss. Thus, we conclude that each term is critical to our model. In the AMT experiments, $\mathcal{L}_{adv}+\mathcal{L}_{cyc}$ outperforms $\mathcal{L}_{adv}$ and $\mathcal{L}_{adv}+\mathcal{L}_{ref}$, on 23.2% labeled real, 22.7% hard to label and 54.1% labeled fake, which shows that $\mathcal{L}_{adv}$ and $\mathcal{L}_{cyc}$ is the fundamental terms. This illustrates that the bidirectional mappings can learn more informative features than only directional mappings. It often incurs training instability and causes mode collapse, especially for the direction of the mapping that is removed. Furthermore, $\mathcal{L}_{adv}+\mathcal{L}_{cyc}+\mathcal{L}_{ir}$ and $\mathcal{L}_{adv}+\mathcal{L}_{cyc}+\mathcal{L}_{strc}$ outperform $\mathcal{L}_{adv}+\mathcal{L}_{cyc}$, because the infrared radiation and structure information are protected.

Through comparison with the result of $\mathcal{L}_{adv}+\mathcal{L}_{cyc}+\mathcal{L}_{ref}$, it is obvious that $\mathcal{L}_{ref}$ is a significant term for optimizing the Generator G_{S2R}. This term makes participants hard to label on 42.7%, far above the result of $\mathcal{L}_{adv}+\mathcal{L}_{cyc}$. It shows similar trends in FCN-score experiments. Thus, $\mathcal{L}_{ref}$ can help the objective to improve the Generator G_{S2R}'s ability, because $\mathcal{L}_{ref}$ enlarge the updated weight of G_{S2R} in every epoch of training.

2) THE ROLE OF U-NET

The U-net architecture allows low-level features to shortcut across the neural network. Does this cause better results? To demonstrate the necessity of U-net in the architecture of the Generators, we compare the results of two groups of experiments, which are with U-net method and without U-net method. Figure 9 and Table 6 show the performance

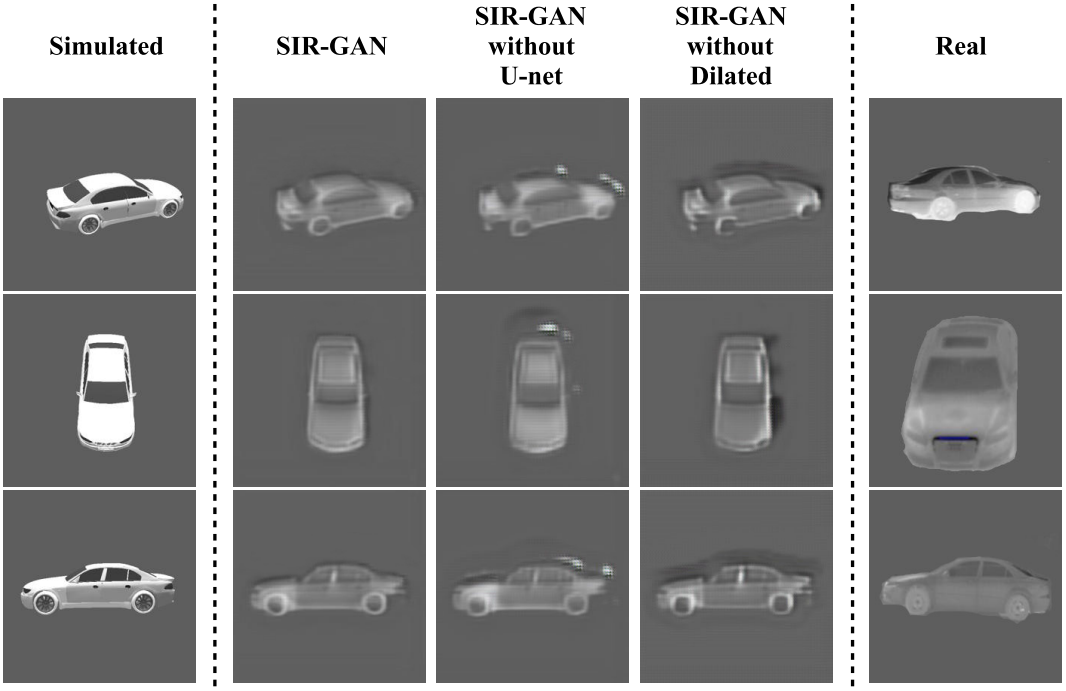


FIGURE 9. Visual presentation of the results of the proposed SIR-GAN with different components. We select 3 poses of the vehicle. The first column is the simulated IR images, while the last column is the real IR images.

TABLE 5. Ablation study for different losses of our method.

Method	AMT			FCN-score		
	labeled real	hard to label	labeled fake	Per-pixel	Per-class	Class IoU
$\mathcal{L}_{adv}$	8.3%	11.5%	80.2%	0.39	0.14	0.10
$\mathcal{L}_{adv} + \mathcal{L}_{cyc}$	23.2%	22.7%	54.1%	0.52	0.20	0.14
$\mathcal{L}_{adv} + \mathcal{L}_{ref}$	12.6%	20.4%	67.0%	0.44	0.15	0.11
$\mathcal{L}_{adv} + \mathcal{L}_{cyc} + \mathcal{L}_{ir}$	25.8%	35.4%	38.8%	0.54	0.22	0.15
$\mathcal{L}_{adv} + \mathcal{L}_{cyc} + \mathcal{L}_{strc}$	24.4%	29.5%	46.1%	0.53	0.21	0.14
$\mathcal{L}_{adv} + \mathcal{L}_{cyc} + \mathcal{L}_{ref}$	**28.4%**	**42.7%**	**28.9%**	**0.59**	**0.23**	**0.16**

TABLE 6. Ablation study for different architectures of our method.

Method	AMT			FCN-score		
	labeled real	hard to label	labeled fake	Per-pixel	Per-class	Class IoU
Ours	28.4%	42.7%	28.9%	0.59	0.23	0.16
Ours without U-net	22.8%	37.1%	40.1%	0.50	0.20	0.14
Ours without Dilated	25.8%	39.1%	33.1%	0.53	0.21	0.15

of our method with and without the U-net. For the AMT experiment, it is distinct that ours without the U-net on 22.8% labeled real, 37.1% hard to label and 40.1% labeled fake underperforms the standard SIR-GAN, but outperforms other state-of-the-art models in Table 3. Similarly, the FCN-score experiment illustrates the same trend. Without the U-net, our generator model created simply by severing the skip connections in the U-Net, is unable to learn how to refine the simulated IR images to the realistic IR images. As we can see in Figure 9, the results of SIR-GAN model without U-net have blurred details, compared to standard SIR-GAN model.

TABLE 7. Ablation study for different architectures of our method.

Data from	Trained on	Tested with	Accuracy
IR Simulator	100% S	R	0.22
SIR-GAN	100% RS	R	0.68
SIR-GAN+Real	75% RS + 25% R	R	0.82
SIR-GAN+Real	50% RS + 50% R	R	0.89
SIR-GAN+Real	25% RS + 75% R	R	0.90
Real	100% R	R	0.92

This is caused by the loss of the target outline information transfer.

TABLE 8. Radiation data of each component of the vehicle.

Component	Material	Temperature/(°C)	Infrared emissivity	Radiant exitance/($W \cdot m^{-2}$)	Grey value
hood	paint	70	0.95	292.70	212
shell	paint	35	0.95	184.68	131
tire	rubber	70	0.90	277.29	200
wheel	metal	50	0.35	83.82	55
window	glass	35	0.85	165.24	116

3) THE ROLE OF DILATED CONVOLUTION

To investigate the behavior of dilated convolution, we conducted several ablation studies. We compared our SIR-GAN model with SIR-GAN without dilated convolution. Figure 9 shows some examples of comparison, including the results of Simulated domain, SIR-GAN, SIR-GAN without dilated convolution and Real domain. It is distinct that, without dilated convolution, a dark blur and fuzz appears at the edges of the target, caused by the loss of internal data structure and spatial hierarchy information. Table 6 presents AMT and FCN-score experiments. Though ours without dilated convolution gets a quite good result compared with ours without U-net, it still has a gap with the standard SIR-GAN model.

The dilated convolution tackles three problems caused by the standard CNNs, as mentioned before. (1)Unlike pooling layer of CNNs, the dilated convolution is learnable, through combining a convolutional layer with a pooling layer. (2)The Dilated convolutional layer can expand the receptive field and maintain the original resolution. Consequently, under the same computational complexity, it will reduce the information loss and increase precision. (3)The dilatied convolution retains the shape and outline information of the target, while the standard CNNs may cause loss of internal data structure and spatial hierarchy information, which is especially important in the simulated IR refinement task. There is no doubt that, the dilated convolutional layer is an indispensable component.

E. FEASIBILITY STUDY

We have illustrated that our approach can generate a large number of qualified IR images, which is realistic enough according to several scientific experiments. Furthermore, whether the refined IR images is practically usable is still a question. Then, we try to use our refined IR images to train the off-the-shelf classifier, and test on the real IR images. In addition, to further study on how to use the refined IR images, we train the Inception V4 model [40] with several groups of training data, which mix refined and real IR images by different ratios. The experimental results are shown in Table 7. The result of the classifier trained on simulated IR images is 0.22. The reason is simulated images possess the outline and shape features. Through comparison results between IR Simulator and SIR-GAN, it can be seen that the classifier trained on the refined data outperforms the classifier trained on the simulated data. This means that the proposed method improves the quality of simulated IR image in content. If we change the ratio of training samples between Refined Simulated data and Real data, the classification accuracy is positively related to the proportion of Real data.

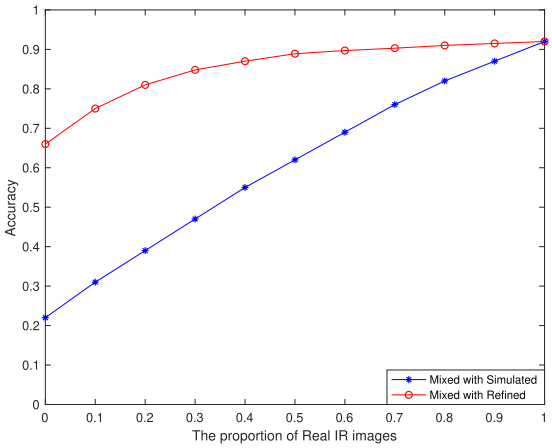


FIGURE 10. The line chart of the test accuracy in training sets with the varying proportion of real IR images. For the red line, the training set consists of real and simulated IR samples. For the blue line, the training set consists of real and refined IR samples.

Furthermore, in order to investigate the effect of mixing ratio changes in the training set on classification accuracy, we design an experiment that as the proportion of real IR data increases linearly at a linear ratio of 0.1, we train the Inception V4 with these different training sets and test on the same real IR set, as presented in Figure 10. It shows the relationship between accuracy and the proportion of real IR images in training set. For the blue line, the training set mixes simulated IR images with real IR images, and the line is close to linear. In contrast, the red line reveals that as the proportion of Real data increases linearly, the increase in accuracy continues to decrease. In other words, when our refined IR images as training set are used to train a learning-based classifier model, the accuracy will be significantly increased if the training set adds a limited quantity of real IR training samples. This characteristic makes our refined IR images extraordinarily practical and meaningful in engineering applications.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose an end-to-end SIR-GAN algorithm for simulated IR images refinement, by taking into account the infrared radiation and semantic information. As a refiner framework of learning bidirectional mappings relationship between two unpaired and imbalanced domains, the proposed

SIR-GAN model combines the SIR adversarial loss, the SIR cycle consistency loss and the SIR refinement loss as a novel training objective function. The training set consists of two domains, i.e., the source domain where the simulated IR images are automatically generated by the proposed IR target simulation system, and the target domain which is a limited quantity of real IR images. We tackle the unpaired and imbalance training by using the proposed training strategy. To further reduce the gap between synthetic and real data, we make several key modifications on the architecture and the training stage. To our knowledge, this work is the first attempt to propose a GAN-based end-to-end framework for synthetic IR refinement task. Finally, experimental results reveal that the proposed SIR-GAN achieves significant improvement, compared with other state-of-the-art methods. To the end, the refined IR images are practically utilized to train the Inception V4 classifier and test on the real data. Consequently, we find out that if our refined IR images add a limited quantity of real IR samples, as the training set, the accuracy of the classifier can be greatly increased.

In future, we intend to study the complex environment generation by adversarial learning. In addition, we will explore modeling the infrared radiation distribution to generate more than one refined IR images for each simulated IR images according to different environment and motion states, and even investigate refining IR videos rather than single images.

APPENDIX A
INFRARED RADIATION CALCULATION

The radiation flux of the vehicle is defined by the Planck formula:

$$\hat{\Phi}_{\lambda_1-\lambda_2} = \int_{\lambda_1}^{\lambda_2} \Gamma(\lambda, T) \cdot \frac{c_1}{\lambda^5[exp(\frac{c_2}{\lambda T}) - 1]} d\lambda \qquad (13)$$

where $[\lambda_1, \lambda_2]$ is a specified infrared band range, $\Gamma(\lambda, T)$ represents emissivity of the material under a specific band and temperature. c_1 and c_2 are the first and second radiation constant respectively. T is the surface temperature of the vehicle.

Then, we calculate the radiation flux of environment. The reflected radiation flux from environment can be expressed as:

$$\Phi_{f(\lambda_1-\lambda_2)} = \rho_e * (Q_{solar} + Q_{sky} + Q_{ground}) \qquad (14)$$

where Q_{solar}, Q_{sky} and Q_{ground} are the reflected radiation of solar, sky and ground, respectively.

The sum of radiation consists of radiation of the vehicle $\hat{\Phi}_{\lambda_1-\lambda_2}$ and reflected radiation of environment $\Phi_{f(\lambda_1-\lambda_2)}$, as defined in Equation (15).

$$\Phi = \hat{\Phi}_{\lambda_1-\lambda_2} + \Phi_{f(\lambda_1-\lambda_2)} \qquad (15)$$

After calculating the sum radiation flux of the vehicle, each component of the vehicle can be inferred by numerical calculation according to their temperature distribution, in order to build the vehicle radiation model. According to the experimental data from Table 8, we can calculate the sum radiation.

Finally, we convert the invisible infrared radiation into the visible-infrared image. Since human eyes can not sense infrared radiation, a mapping relationship must be established between the infrared radiation and the image, so that the infrared radiation distribution of the target can be visually displayed in different brightness of a grey image. Our quantitative criterion is the linear grey level mapping method, which is defined as follows.

$$g = \frac{\Phi - \Phi_{min}}{\Phi_{max} - \Phi_{min}} * (g_{max} - g_{min}) + g_{min} \qquad (16)$$

where g is the quantized grey level. g_{max} and g_{min} are the upper and lower value of grey level. We choose $g_{max} = 255$ and $g_{min} = 0$. Φ_{max} and $\Phi_{min}$are the upper and lower of the infrared radiation value of the vehicle. After infrared texture mapping, the simulated IR images are generated through OGRE rendering based on GPU acceleration. In addition, every simulated IR target image can be easily auto-annotated with labels, including target class, band range, and environment temperature.

REFERENCES

[1] T. Yang, Z. Li, F. Zhang, B. Xie, J. Li, and L. Liu, "Panoramic uav surveillance and recycling system based on structure-free camera array," *IEEE Access*, vol. 7, pp. 25763–25778, 2019.
[2] Q. Zhang, Y. Zhou, S. Song, G. Liang, and H. Ni, "Heart rate extraction based on near-infrared camera: Towards driver state monitoring," *IEEE Access*, vol. 6, pp. 33076–33087, 2018.
[3] E. Resendiz-Ochoa, R. A. Osornio-Rios, J. P. Benitez-Rangel, R. De J. Romero-Troncoso, and L. A. Morales-Hernandez, "Induction motor failure analysis: An automatic methodology based on infrared imaging," *IEEE Access*, vol. 6, pp. 76993–77003, 2018.
[4] Y. Li, A.-B. Ming, H. Mao, G.-F. Jin, Z.-W. Yang, W. Zhang, and S.-Q. Wu, "Detection and characterization of mechanical impact damage within multi-layer carbon fiber reinforced polymer (CFRP) laminate using passive thermography," *IEEE Access*, vol. 7, pp. 27689–27698, 2019.
[5] M. A. Zulkifley, "Two streams multiple-model object tracker for thermal infrared video," *IEEE Access*, vol. 7, pp. 32383–32392, 2019.
[6] J. Gao, Z. Lin, and W. An, "Infrared small target detection using a temporal variance and spatial patch contrast filter," *IEEE Access*, vol. 7, pp. 32217–32226, 2019.
[7] Z. Yi, H. Zhang, and P. Tan, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2849–2857.
[8] R. R. Atallah, A. Kamsin, M. A. Ismail, S. A. Abdelrahman, and S. Zerdoumi, "Face recognition and age estimation implications of changes in facial features: A critical review study," *IEEE Access*, vol. 6, pp. 28290–28304, 2018.
[9] L. Liu, X.-F. Gu, T. Yu, X.-Y. Li, J.-G. Li, H.-L. Gao, and Y. Sun, "Target identification based on the simulation of infrared thermal image," in *Proc. Int. Conf. Electr. Inf. Control Eng.*, Apr. 2011, pp. 3653–3656.
[10] S. Wang, L. Hu, D. Zhou, B. Du, and Y. Mao, "Design of a portable infrared/visible composite target simulator," *Proc. SPIE*, vol. 11023, Mar. 2019, Art. no. 110230K.
[11] H. Liu, S. Xie, C. Pei, J. Qiu, Y. Li, and Z. Chen, "Development of a fast numerical simulator for infrared thermography testing signals of delamination defect in a multilayered plate," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5544–5552, Dec. 2019.
[12] L. Ren, L. Ni, and G. Guanghong, "Research of Key technology about scene rendering based on ogre engine," *J. Syst. Simul.*, vol. 29, no. Suppl. 1, pp. 161–172, Dec. 2017.
[13] W. Lin, X. Li, Z. Yang, L. Lin, S. Xiong, Z. Wang, X. Wang, and Q. Xiao, "A new improved threshold segmentation method for scanning images of reservoir rocks considering pore fractal characteristics," *Fractals*, vol. 26, no. 2, 2018, Art. no. 1840003.

[14] W. Lin, X. Li, Z. Yang, J. Wang, S. Xiong, Y. Luo, and G. Wu, "Construction of dual pore 3-D digital cores with a hybrid method combined with physical experiment method and numerical reconstruction method," *Transp. Porous Media*, vol. 120, pp. 227–238, Oct. 2017.

[15] W. Lin, Z. Yang, X. Li, J. Wang, Y. He, G. Wu, S. Xiong, and Y. Wei, "A method to select representative rock samples for digital core modeling," *Fractals*, vol. 25, no. 4, 2017, Art. no. 1740013.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[17] C. Karabacak, S. Z. Gurbuz, A. C. Gurbuz, M. B. Guldogan, G. Hendeby, and F. Gustafsson, "Knowledge exploitation for human micro-Doppler classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2125–2129, Oct. 2015.

[18] F. Lahoud and S. Susstrunk, "Ar in vr: Simulating infrared augmented vision," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3893–3897.

[19] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 102–118.

[20] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks?" in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May/Jun. 2017, pp. 746–753.

[21] R. Zhang, C. Mu, Y. Yang, and L. Xu, "Research on simulated infrared image utility evaluation using deep representation," *J. Electron. Imag.*, vol. 27, no. 1, 2018, Art. no. 013012.

[22] R. Zhang, C. Mu, X. Gao, K. Liu, and Y. Ma, "A fusion algorithm of template matching based on infrared simulation image," *Proc. SPIE*, vol. 10033, Aug. 2016, Art. no. 1003307.

[23] X. Gao, C.-P. Mu, M.-S. Peng, Q.-X. Dong, and R.-H. Zhang, "The infrared image simulation of the tank under different movement states," *Proc. SPIE*, vol. 10420, Jul. 2017, Art. no. 1042024.

[24] S. Guo, X. Xiong, Z. Liu, X. Bai, and F. Zhou, "Infrared simulation of large-scale urban scene through LOD," *Opt. Express*, vol. 26, no. 18, pp. 23980–24002, 2018.

[25] X. Xiong, F. Zhou, X. Bai, B. Xue, and C. Sun, "Semi-automated infrared simulation on real urban scenes based on multi-view images," *Opt. Express*, vol. 24, no. 11, pp. 11345–11375, 2016.

[26] Z. Fan, W. Tong, H. Kemeng, M. Jiaming, L. Meiling, and W. Zhangye, "One improved real-time infrared simulation system based on unity3D," *J. Comput.-Aided Des. Comput. Graph.*, vol. 30, p. 1177, Jul. 2018.

[27] B. Yang and M. Lv, "The research on dynamic infrared scene simulation for infrared seeker," in *Proc. 2nd Int. Conf. Mechatronics Eng. Inf. Technol. (ICMEIT)*. Atlantis Press, May 2017.

[28] R. Zhang, C. Mu, M. Xu, L. Xu, and X. Xu, "Facial component-landmark detection with weakly-supervised LR-CNN," *IEEE Access*, vol. 7, pp. 10263–10277, 2019.

[29] C. Mu, M. Peng, X. Gao, R. Zhang, and Q. Dong, "Infrared image simulation of ground maneuver target and scene," *J. Beijing Inst. Technol.*, vol. 2, p. 13, 2016.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[32] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," Jul. 2016, *arXiv:1607.08022*. [Online]. Available: https://arxiv.org/abs/1607.08022

[33] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 1–21.

[34] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.

[35] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2107–2116.

[36] K. Lee, H. Kim, and C. Suh, "Simulated+ unsupervised learning with adaptive data generation and bidirectional mappings," in *Proc. Int. Conf. Learn. Represent.*, 2018.

[37] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," Nov. 2015, *arXiv:1511.06434*. [Online]. Available: https://arxiv.org/abs/1511.06434

[38] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 649–666.

[39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[40] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 21st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.

[41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[42] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 172–189.

[43] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4066–4079, Aug. 2018.

[44] Y. Cao, Z. Zhou, W. Zhang, and Y. Yu, "Unsupervised diverse colorization via generative adversarial networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2017, pp. 151–166.

[45] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Cognit. Model.*, vol. 5, no. 3, p. 1, 1988.

[46] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5400–5409.

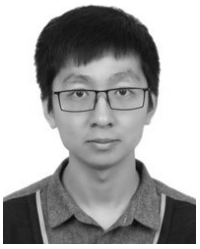
[47] L. Karacan, Z. Akata, A. Erdem, and E. Erdem, "Learning to generate images of outdoor scenes from attributes and semantic layouts," Dec. 2016, *arXiv:1612.00215*. [Online]. Available: https://arxiv.org/abs/1612.00215

[48] M.-Y. Liu, O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 469–477.

[49] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 700–708.

[50] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," Dec. 2013, *arXiv:1312.6114*. [Online]. Available: https://arxiv.org/abs/1312.6114

[51] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016, *arXiv:1609.04747*. [Online]. Available: https://arxiv.org/abs/1609.04747



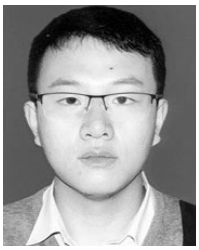
RUIHENG ZHANG received the B.S. degree in information engineering from the Beijing Institute of Technology, China, in 2014. He is currently pursuing the joint Ph.D. degree with the Beijing Institute of Technology and the University of Technology Sydney. He is the author of more than ten research articles and one book. His current research interests include deep learning, computer vision, and object detection. He has served as the Reviewer for several international journals and conferences.

CHENGPO MU received the B.S. and M.S. degrees from the Beijing Institute of Technology and the Ph.D. degree from Beijing Jiaotong University. He is currently an Associate Professor with the Beijing Institute of Technology. He has published more than 60 journal articles and two books. His current research interests include deep learning, cyber security, and 3D simulation.

MIN XU received the B.E. degree from the University of Science and Technology of China, in 2000, the M.S. degree from the National University of Singapore, in 2004, and the Ph.D. degree from The University of Newcastle, Australia, in 2010. She is currently an Associate Professor with the University of Technology Sydney. She has published over 100 research articles in high quality international journals and conferences. Her research interests include multimedia data analytics, pattern recognition, and computer vision. Over 1500 citations of her research articles show her reputation in her research field.

LIXIN XU received the Ph.D. degree in information engineering from the Harbin Institute of Technology. He is currently a Professor with the Beijing Institute of Technology. He has published 100 journal and conference papers. His current research interests include deep learning, MEMS, and target detection. He has served as an Editor and a Reviewer for several international journals and conferences. He is the Editorial Board of *Journal of Detection and Control*.

QIAOLIN SHI received the B.S. degree from the Beijing Institute of Technology, Beijing, China, in 2014. She is currently pursuing the Ph.D. degree with the School of Information and Electronics, Beijing Institute of Technology. Her research interests include statistical learning on graphical models and its application to wireless communications. She has served as a Reviewer for several international journals and conferences.



JUNBO WANG received the Ph.D. degree in navigation, guidance, and control from Beihang University, China. He is currently a Senior Engineer with the Beijing Institute of Electronic System Engineering. He is the author of five research articles and two patents. His current research interests include machine learning, fuzzy, navigation, and control.