End to End Infrared and Visible Image Fusion
With Texture Details and Contrast Information

JINGYU JI\textsuperscript{1}, YUHUA ZHANG\textsuperscript{1}, ZHILONG LIN\textsuperscript{1}, YONGKE LI\textsuperscript{1}, CHANGLONG WANG\textsuperscript{1},
YONGJIANG HU\textsuperscript{2}, FUYU HUANG\textsuperscript{2}, AND JIANGYI YAO\textsuperscript{3}

\textsuperscript{1}Department of UAV, Army Engineering University, Shijiazhuang, Hebei 050000, China
\textsuperscript{2}Department of Electronic and Optical Engineering, Army Engineering University, Shijiazhuang, Hebei 050000, China
\textsuperscript{3}Equipment Simulation Training Center, Army Engineering University, Shijiazhuang, Hebei 050000, China

Corresponding authors: Yongjiang Hu (hyj70001@163.com) and Changlong Wang (wcl_lgd@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62171467, and in part by the Natural Science Foundation of Hebei Province under Grant F2021506004.

\textbf{ABSTRACT} Infrared and visible image fusion combine data information from different sensors to achieve a richer description of the same scene. In order to highlight the salient features of the infrared image and the visible image in the fusion image and obtain a fusion image with good performance, an end-to-end infrared and visible image fusion algorithm is proposed in this paper. The contrast attention module and visible image cascade part are introduced in the generator, so that the fusion image can focus on the detail information in the visible image and the contrast information in the infrared image. And in order to retain more structural contour information in the original image, the contour loss is added to the content loss function. In addition, the contrast and detail information in infrared and visible images are balanced by two discriminators. And a goal-guided reward function is introduced into the discriminator, which further facilitates the generator to produce effective fused images. Finally, extensive fusion experiments on public datasets verify the advantages of the proposed algorithm compared with other classical algorithms, and ablation experiments demonstrate the effectiveness of the improved part of the algorithm.

\textbf{INDEX TERMS} End-to-end infrared and visible image fusion, contrast attention module, contour loss, target-guided reward function.

\section{I. INTRODUCTION}
The purpose of image fusion is to extract useful data information from images obtained from different sensors, and then fuse them to generate an image that contains more information. Among the existing multi-sensor fusion, infrared and visible image fusion is the most widely utilized [1]. Infrared images are thermal radiation images obtained from infrared sensors that are minimally affected by changes in lighting and obstructions, and they can be easily captured at night even with obstructions. But infrared images usually lack texture, and their ability to express details is not enough. Visible images are light reflection images obtained from visible sensors, which have rich structural texture information. But at night, with obstructions, or with a cluttered background, objects in visible images may not be easily observed. Due to the complementary characteristics of infrared and visible images, the fusion of the two can obtain rich detailed information and important thermal target information, and obtain ideal fusion results. Infrared and visible image fusion has been widely utilized in target detection [2], tracking [3], agricultural activities [4], military operations [5], and many other fields [6], [7].

For the fusion of infrared and visible images, the key is to maintain the significant contrast information in the infrared images and the rich detailed texture information in the visible images. To achieve this goal, many traditional infrared and visible image fusion algorithms have been widely proposed in the past, and these traditional algorithms can be divided into six categories: feature-based methods [8], decision-based methods [9], sparse representation-based methods [10], transformation-based methods [11], subspace-based methods [12], and hybrid methods [13], which have achieved good results in many application scenarios. In order to better
consider the inherent properties of infrared and visible images, and to design more convenient activity level measurement and fusion rules, recently, with the in-depth development of deep learning, many researchers have found that algorithms using deep learning can achieve this purpose. Therefore, deep learning-based methods are now widely studied. Generally, methods based on deep learning can be divided into two categories, one is fusion methods based on convolutional neural networks, and the other is fusion methods based on generative adversarial networks (GAN). Fusion methods based on convolutional neural networks usually apply convolutional neural networks in part of the fusion process, while the overall fusion process is still in the traditional framework [14], [15]. The fusion method based on GAN is an end-to-end model, which does not require manual design of fusion rules or reconstruction strategies. This feature enables GAN-based fusion algorithms to have good performance in applications, and various GAN-based image fusion methods have emerged in an endless stream, such as FusionGAN [16], DDcGAN [17] and ResNetFusion [18].

Although these methods can well solve some problems of the previous methods, their fusion results still have shortcomings such as loss of details, lack of contrast, and poor visual perception. On the one hand, these methods are based on improving a type of feature in the fused image, such as enhancing the detail features in visible images. However, such improvements may affect the features of the infrared images, resulting in an unsatisfactory final fusion performance. On the other hand, these methods do not fully address the balance of the respective features of infrared and visible images in the fusion image, resulting in poor fusion results.

In order to solve the above problems and improve the fusion quality of end-to-end models, a new GAN network framework was proposed in this paper. First, in order to increase the contrast information in infrared images, a contrast attention module is introduced in the generator, and this module is cascaded with each layer of the network to retain more contrast information. Simultaneously, in order to increase the detail information in visible images, visible images are connected at each layer, so that the fused image retains more visible information. Second, a contour loss function is introduced, that is, a contour loss is added to the content loss to control the similarity of the contour detail features of the generated image and the source image. In addition, a dual discriminator is designed to identify and enhance its corresponding modal features under the constraints of pixel intensity and contour variation and is supervised by a specific loss function to fully extract information from infrared and visible images. Simultaneously, in the discriminator, in order to improve the discriminator’s discriminative ability and promote the generator to produce better fusion images, a target-guided reward function module is utilized in the discriminator, which enables it to more accurately discriminate images from the generator. The entire network is an end-to-end model, which can not only overcome the shortcomings of the manual design of activity level measurement and fusion rules in traditional algorithms, but also solve the shortcomings of previous GAN-based fusion algorithms. A large number of experiments have proved that the proposed algorithm can achieve significant fusion results with visual effects, and because the trained model is directly utilized in the test stage, the real-time performance of the algorithm is also very high. The main contributions are as follows:

1. By introducing the contrast attention module and cascading the visible image with the network, not only the contrast information of the fused image is enhanced, but also the texture details of the fused image are effectively enhanced.
2. Based on the original loss function, the contour loss function is introduced to make the fusion image and the source images have more similar edge and contour details.
3. Two discriminators are employed to fully extract information from infrared and visible images. A target-guided reward function module is designed in the discriminator to improve the discriminator’s discriminative ability and thus the quality of the generator to generate fused images.
4. Comparative experiments and ablation experiments verify the good performance of the proposed algorithm from different perspectives.

The rest of this paper is organized as follows. Section 2 presents related work on image fusion, and Section 3 introduces the algorithm proposed in this paper. Section 4 is the experimental part, which validates the proposed algorithm on public data and compares it with other methods. The effectiveness of the key parts of the proposed algorithm is verified by ablation experiments. Conclusions are in Section 5.

II. RELATED WORKS

A. IMAGE FUSION BASED ON DEEP LEARNING

Deep learning is a research subfield of machine learning, with artificial neural network (ANN) as the main architecture of the model, aiming to make the model fit the end-to-end data mapping well in a data-driven manner. The deep learning fusion methods can not only mine the deep features of the image, but also have good model learning ability. In recent years, deep learning fusion methods have become a new research direction in the field of infrared and visible image fusion [19]. Compared with the traditional fusion algorithm, the deep learning network can optimize the error caused by the traditional algorithm from manually extracting features, constructing fusion rules and reconstruction, further to improve the performance of image fusion, and it is more robust.

Liu et al. [20] first used convolutional neural networks for learning and training to obtain feature maps, and then reconstructed feature maps and source images to obtain fused images. Li et al. decomposed the infrared and visible images separately to obtain the base layer and the detail layer [15]. They used a pre-trained VGG network [21] to extract the
detail layer features and proved that its computational efficiency is higher than the traditional methods. Different from the direct use of deep features by the above methods, for the deep features of infrared and visible images extracted by the pre-trained ResNet [22], Li et al. [23] used zero-phase component analysis (ZCA) and L1 regularization for post-processing, and the subjective and objective evaluations of the obtained fused images were better. In addition, Li et al. also proposed an autoencoder-based fusion method, named DenseFuse [24]. The similarity between images was used for training constraints during training, and two given fusion strategies of addition and L1-norm were utilized during testing, which can achieve adaptive fusion of infrared and visible images. Liu et al. applied CSR to image fusion [14] to perform image decomposition based on sparse coefficients obtained by CSR. Compared with other traditional methods based on sparse representation, this method preserved details more comprehensively, and had better robustness when multi-source image registration is poor.

The above methods either still require manual design of partial fusion rules, or a manual design of ground truth to train the network. On the one hand, there will be some errors in artificially designed fusion rules, which are usually designed with human subjective consciousness. On the other hand, there is no ground truth in image fusion, and the artificially constructed ground truth will limit the learning and training of the network. With the expansion of the application scope, the scene information represented by multimodal images becomes more and more complex and diverse. Ideally, image fusion is an end-to-end image generation task with multimodal images as the input and fused images as the output. Since 2019, researchers have proposed a data-driven end-to-end deep learning fusion method, which overcomes the complex and difficult design of fusion strategies in the previous methods, and significantly improves the computational efficiency and universality of image fusion.

Researchers in the field of deep learning continue to optimize the combination of deep networks. Goodfellow et al. proposed the Generative Adversarial Network (GAN) [25] architecture based on zero-sum game theory. GAN can generate new samples that conform to given constraints, which promotes the application of deep learning from data discrimination to data generation. In 2019, Ma et al. proposed the first image fusion method based on generative adversarial networks, named FusionGAN [16], which transformed the fusion task into an adversarial learning process for infrared and visible image information retention, opening up new ideas for deep learning fusion methods. Xu et al. proposed a conditional generative adversarial network-based fusion method (DDcGAN) [17] at the IJCAI 2019 conference, and designed two discriminators that constrain the information retention of infrared and visible images, respectively. Based on their research basis in the literature [26] and literature [16], Ma et al. proposed an end-to-end fusion method based on detail preservation [27]. The designed detail loss helped preserve visible texture information and can significantly improve the clarity of fused images. With the development of deep learning technology, starting from the modal characteristics of infrared and visible images and the target task of image fusion, the semantic information of images is mined and utilized to improve the fusion efficiency and quality in different scenarios, which is a new direction for deep learning fusion method research worthy of expansion.

B. GENERATIVE ADVERSARIAL NETWORKS

GAN is a probabilistic generative model, which makes the generated sample distribution obey the real data distribution through a game. The traditional GAN model consists of a generator $G$ and a discriminator $D$. In order for $G$ to learn the distribution $P_g$ of the real data $x$, a noise variable $P_z(z)$ is first defined. The $G$ is mapped to the data space $G(z, \theta_G)$; the $D(z, \theta_D)$ is a binary classification network, which is utilized to judge whether the data generated by the $G$ comes from the $P_g$. During the model training, $G$ tries to “deceive” $D$ by generating real samples as much as possible, and the discriminator tries to identify fake samples as much as possible, and the two play against each other. In other words, the generative adversarial network is a minimax optimization problem, and the optimal value is a saddle point at which the generator reaches the minimum value and the discriminator reaches the maximum value. The objective function $V(D, G)$ [25] of the traditional generative adversarial network is shown in Eq. (1):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} \left[ \log D(x) \right] + E_{z \sim P_z(z)} \left[ \log (1 - D(G(z))) \right]$$ (1)

where $P_{data}$ represents the distribution of real training data, $P_z(z)$ represents the input noise variable. $D(x)$ represents the probability that the discriminator distinguishes the sample from the real sample, $\log D(x)$ represents the cross entropy (CE) of $[1, 0]^2$ and $[D(x), 1 - D(x)]^2$, $\log (1 - D(G(z)))$ represents the cross entropy of $[1, 0]^2$ and $[D(G(z)), 1 - D(G(z))]^2$.

However, in the early stage of traditional GAN model training, the learning performance of the generator is very poor, and the generated data distribution is quite different from the real data distribution. So that the discriminator has a high degree of confidence in identifying the generated data, and it causes the problem that the gradient of the objective function in (1) back-propagated to the generator is very small, resulting in the problem of gradient disappearance. The Least Squares Generative Adversarial Nets (LSGAN) [28] introduced to this paper proposed a least squares loss function based on the traditional GAN model. The new decision boundary generated by the discriminator $D$ will penalize those who are far away from the decision boundary. The samples are generated to provide a larger gradient for the generator $G$ during the training process, which overcomes the problem that the gradient of the original GAN model disappears. The LSGAN objective function is shown
in Eq. (2) and Eq. (3):
\[
\min_D V_{\text{LSGAN}}(D) = E_{x \sim P_{\text{data}}(x)}[(D(x) - b)^2] + E_{z \sim P_z(\mathcal{Z})}[(D(G(z)) - a)^2] \quad (2)
\]
\[
\min_G V_{\text{LSGAN}}(G) = E_{z \sim P_z(\mathcal{Z})}[(D(G(z)) - c)^2] \quad (3)
\]
where \(a\) and \(b\) represent the labels of the generated samples and real samples, respectively, and \(c\) represents the boundary value that the discriminator identifies as the generated samples.

III. THE PROPOSED ALGORITHM

In a generative adversarial network model, the purpose of training is to let the generator fool the discriminator to achieve the final goal. Corresponding to image fusion, that is, the purpose of training is to make the generator generate a fusion image with good visual performance, rich information, and prominent contrast. Since infrared and visible images have different characteristics, infrared images are thermal radiation imaging, which has the contrast characteristics of the target; visible images rely on light reflection imaging, which has rich texture details. Therefore, in order to better combine the respective feature attributes of these two types of images and achieve good image fusion results, a generative adversarial network model based on contrast enhancement and detail preservation is designed. This section introduces the proposed model in detail, including the generator and discriminator structures in the network, and the loss function of the network.

A. NETWORK STRUCTURE

The key of the infrared and visible fusion algorithms is that they can effectively extract the features of the source images, and complete the reconstruction of the fused image through a reasonable fusion strategy, so as to improve the fused image expression ability. In this paper, an infrared and visible image fusion algorithm based on GAN is proposed, which transforms the fusion problem into a process of confrontation between source images and fused images. The network structure of the algorithm is shown in Figure 1, and the entire network consists of generators and discriminators.

In the training phase, the infrared and visible images are input into the generator as shown in Figure 1. The infrared image goes through the convolutional blocks and the contrast attention module. Simultaneously, the visible image goes through the convolutional blocks, and each convolutional block is cascaded with the visible image. The two parts are converged before the fourth convolution block, and then the fused image is obtained through the remaining convolution blocks. Then the fused image, infrared image, and visible image are fed into the discriminator for adversarial training. The visible image and the fusion image go through a designed discriminator, and the discriminator judges whether the details of the generated fusion image are rich enough and whether it is natural compared with the original visible image.

Similarly, the infrared image and the fused image also go through a discriminator, and the discriminator is used to judge whether the generated fused image has sufficient contrast and the structural information compared with the infrared image. Both discriminators output a compared value. The generator and discriminator in the training phase alternate until the desired effect is achieved.

In the testing phase, the discriminator is not utilized and only the generator is kept. The infrared image and the visible image are directly input into the trained generator to obtain fused images.

1) GENERATOR NETWORK STRUCTURE MODEL

The generator model \(G\) is an improved network based on VGG-16 [18], which is divided into visible path and infrared path. The basic structure of each path is the same, each path contains 3 convolutional blocks, each block has 2-3 convolutional layers and a max pooling layer at the end of each block. First, the infrared image and the visible image pass through the three convolutional blocks of their respective paths simultaneously. Then the two paths together feed into the fourth convolutional block. Finally, the fused image is output after the fifth convolution block. Before the two-way convergence, in order to better extract the contrast features of the infrared image, a contrast attention module is added after the third convolution block of the infrared path, which can better refine the feature mapping in the network. Inspired by SE-Net [29], the network structure of this module includes convolutional layers, activation layers and pooling layers. As shown in Figure 2.

Contrast attention generating features can be computed as:

\[
A = F \otimes S(\text{Conv}_{1 \times 1}((\text{ReLU} (\text{Conv}_{1 \times 1} (\text{Avg} (F)))))) \quad (4)
\]

where \(F\) is the output of the three convolutional blocks and the input of the contrast attention module. RELU is the Leaky RELU activation function, \(\text{Conv}_{1 \times 1}\) is the convolution operation with a \(1 \times 1 \times 64\) size, and the size of the convolution kernel is set to 3, and the stride is set to 1. \(S\) represent Sigmoid activation function. The contrast attention module finds out the contrast salient features in the infrared images by learning the context information of the samples, and enhances the network’s ability to express the contrast saliency regions. In order to obtain the contrast attention feature, the average value of the contrast feature map is obtained from the input \(F\) by using average pooling, and then the non-linear interaction of the average value of the contrast feature is learned with two convolutions and the Leaky RELU activation function, resulting in a contrast attention feature map. Finally, the Sigmoid activation function is used to map the contrast attention feature to the [0, 1] interval to obtain the output contrast attention descriptor.

In addition, in order to obtain more significant detailed texture features of the visible image, the visible image is cascaded with each convolutional block of the visible path to correlate to the deeper detailed features of the visible image.
2) DISCRIMINATOR NETWORK STRUCTURE MODEL

In order to make the feature information of both infrared and visible image reflected in the fusion image, two discriminators with the same network structure are used to judge the quality of the generated fusion image. As shown in Figure 1. $D_{\text{visible}}$ is designed to discriminate the similarity between the generated fusion image and the original visible image, and $D_{\text{infrared}}$ is designed to discriminate the similarity between the generated fusion image and the original infrared image. The two discriminators contain 4 convolutional blocks, a linear layer, and a reward function module. Among them, every convolutional block designs a convolutional layer, a switchable normalization (SN) layer and a Leaky RELU. The size of samples of each convolutional layer is designed to 32, 64, 128 and 256, respectively. The size of all convolutional kernels is set to 3, and the stride of all convolutional layer is set to 2. Especially, SN is adopted in the discriminator, which can not only improve the speed of network training, but also obtain well-normalized data.

The setting of the reward function can improve the discriminator’s discriminative ability, thereby improving the quality of the fused image. In this paper, a goal-guided reward function module is designed, and its specific expression...
is as follows.

\[ r = \begin{cases} 
0, & s_t < s_{t-1} \\
\theta(s_t - S), & s_t \geq s_{t-1}
\end{cases} \quad (5) \]

where \( r \) represents the reward function, \( \theta \in (0, \infty) \) is the reward coefficient, \( s_t \) is the output value obtained by the discriminator to determine whether the fused image is similar to the source image. The value is between 0 and 1, and the closer to 1 it is, the more similar the images are. When the value reaches 1, the two images are judged to be exactly the same by the machine. \( s_t \) and \( s_{t-1} \) are the network output values in the current state and the previous state, respectively. \( S \) is a constant, the value is greater than 1. It can be seen from the equation that the reward \( r \) is closely related to the state before and after the network update. Whenever the fusion result is more similar to the original image, it will receive a negative reward. Its essence is to improve the judgment ability of the discriminator to motivate the generator to generate better result. It can be seen that the reward function module designed can generate dynamic rewards in real time in combination with state update information, so that the discriminator has good control performance and can be optimized according to the continuously updated similarity results. Since the reward situation is related in real time to the task goal, it is called a goal-directed reward function.

**B. LOSS FUNCTION**

The loss function includes the loss function of the generator and the loss function of the discriminator. The loss function of the generator guides the generator to produce a good fused image, and the loss function of the discriminator is used to facilitate the ability of the discriminator to judge the fused image and the source images.

1) THE LOSS FUNCTION OF THE GENERATOR

Similar to the previous algorithm based on the generative adversarial model, the loss function \( L_{\text{Generator}} \) of the generator in this paper is also composed of the adversarial loss \( L_{\text{adv}}^{\text{Generator}} \) and the content loss \( L_{\text{con}}^{\text{Generator}} \), the difference is that the content loss includes the intensity loss \( L_{\text{Intensity}}^{\text{Generator}} \) and the contour loss \( L_{\text{Contour}}^{\text{Generator}} \) proposed in this paper. The expression of the loss function is as follows:

\[ L_{\text{Generator}} = L_{\text{adv}}^{\text{Generator}} + \lambda L_{\text{con}}^{\text{Generator}} \quad (6) \]

where \( \lambda \) is the weight coefficient, the purpose is to balance the two-part loss function, and this paper sets it to 100. \( L_{\text{Generator}} \) is the loss function of generator, \( L_{\text{adv}}^{\text{Generator}} \) is the adversarial loss, and \( L_{\text{con}}^{\text{Generator}} \) is the content loss. The adversarial loss can be expressed as:

\[ L_{\text{adv}}^{\text{Generator}} = \frac{1}{N} \sum_{n=1}^{N} (D(F_n) - t)^2 \quad (7) \]

where \( D(\cdot) \) represents the discriminator model function, \( F_n \) represents the fusion image, \( t \) is the threshold for the discriminator to judge that the input image is a fake image, \( n \) represents the \( n \)-th image input into the discriminator, and \( N \) represents the number of images input to the discriminator. The intensity loss is expressed as follows:

\[ L_{\text{Intensity}}^{\text{Generator}} = \|F_n - I_{ir}\|^2 \quad (8) \]

where \( F_n \) represents the fusion image, and \( I_{ir} \) represents the infrared image.

\( L_{\text{Contour}}^{\text{Generator}} \) is the contour loss function proposed in this paper. It uses the feature description operator to map the infrared image, visible image and fusion image from the pixel space to the shallow gradient space. By calculating the shallow contour feature distance between the fused image and the source images, the contour feature distribution of the infrared and visible images is learned, and the fusion image is realized. At the same time, the contour information of the infrared target and the detailed texture information of the visible image are preserved. In this paper, the Laplacian operator is used to extract the contour information of the source images, and the Manhattan distance is used to calculate the distance between the gradient of the fusion image and the gradient of the infrared and visible images. The edge loss function \( L_{\text{Contour}}^{\text{Generator}} \) is expressed as follows:

\[ L_{\text{Contour}}^{\text{Generator}} = \frac{1}{HW} \| \nabla F_n - \nabla I_{ir} - \nabla I_{vis} \|_1 \quad (9) \]

where \( H \) and \( W \) are the height and width of the input image, respectively, \( \nabla \) represents the Laplacian operator, and \( \|\cdot\|_1 \) represents the \( L_1 \) norm.

2) DISCRIMINATOR LOSS FUNCTION

Two separate discriminator loss functions work simultaneously for dual adversarial training, namely \( L_{\text{Discriminator}}^{\text{ir}} \) and \( L_{\text{Discriminator}}^{\text{vis}} \). \( L_{\text{Discriminator}}^{\text{ir}} \) is used to discriminate the difference between the generated fused image and the infrared image, and \( L_{\text{Discriminator}}^{\text{vis}} \) is used to discriminate the difference between the generated fused image and the visible image. Through these two loss functions, the discriminator parameters can be well updated, and a fusion image that can balance infrared features and visible features can be generated. The loss functions of the two discriminators are defined as follows:

\[ L_{\text{Discriminator}}^{\text{ir}} = \frac{1}{N} \sum_{n=1}^{N} (D(I_{ir}) - t_{ir})^2 + \frac{1}{N} \sum_{n=1}^{N} (D(F_n) - t_F)^2 \quad (10) \]

\[ L_{\text{Discriminator}}^{\text{vis}} = \frac{1}{N} \sum_{n=1}^{N} (D(I_{vis}) - t_{vis})^2 + \frac{1}{N} \sum_{n=1}^{N} (D(F_n) - t_F)^2 \quad (11) \]

where \( N \) is the total number of input images, \( D(I_{ir}) \) is the discrimination result of infrared images, \( D(I_{vis}) \) is the discrimination result of visible images, and \( D(F_n) \) is the discrimination result of the fused image. \( t_{ir}, t_{vis}, t_F \) represent the threshold of infrared image, visible image, and fusion image, respectively. In our paper, \( t_F \) is a constant number...
between [0, 0.3], and \( t_{ir} \) and \( t_{vis} \) are set as constant random numbers between [0.7, 1.2].

**IV. EXPERIMENTS**

In this section, first, the detailed experimental configuration is presented. Second, a competent visual evaluation and objective quantitative analysis of the proposed method and other commonly used methods are performed. Finally, additional ablation experiments are provided.

**A. EXPERIMENTAL SETTING**

During training, 40 pairs of images from the TNO [21] dataset and 20 pairs of images from the RoadScene [30] dataset were selected for training. Since only these image pairs cannot be used for good learning and training, we use a sliding step size of 14 for the source images, and split these images into images of size 128 x 128, which effectively expands the training dataset. In addition, the Adam optimizer is used to update the network parameters, the Batch size is set to 32, the epoch is set to 16, the learning rate is set to 0.0001, and the training step of the discriminator is set to 2. The experimental platform is adopted PyCharm and MATLAB 2018a on a PC with twelve Intel (R) Core (TM) i7-8700 CPU @ 3.2 GHz and NVIDIA GeForce RTX 2060.

During the testing process, TNO dataset, RoadScene dataset and other infrared and visible image datasets were selected for testing. And we select five classic methods such as FusionGAN [16], CNN [31], ResNet [23], GTF [26], GANMcC [32] to compare with the proposed method. In addition, the fusion performance is objectively evaluated by six indicators, including entropy (EN) [33], spatial frequency (SF) [34], correlation coefficient (CC) [35], mutual information (MI) [35], structural similarity (SSIM) [36], and \( Q_{ABF} \) [37]. These metrics are defined as follows:

Entropy can indicate the richness of the information of the image. The larger the entropy value of the fusion image is, the richer the feature information contains, and the better the fusion performance is. The formula for calculating EN is as follows:

\[
EN = -\sum_{n=1}^{R-1} p(n) \log_2 p(n)
\]

where \( R \) is the dynamic range of the analyzed image, and \( p(n) \) is the distribution probability of the \( n^{th} \) gray level.

Spatial frequency is a measure of the overall activity level of the image, which can reflect the ability of the fusion image to describe the contrast of details, expressed as row frequency and column frequency. The higher the spatial frequency value of the fused image is, the clearer the fused image is. The formula for calculating SF is as follows:

\[
SF = \sqrt{RF^2 + CF^2}
\]

where \( RF \) and \( CF \) are given by:

\[
RF = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=2}^{N} [F(m, n) - F(m, n - 1)]^2
\]

\[
CF = \frac{1}{M \times N} \sum_{i=2}^{M} \sum_{j=1}^{N} [F(m, n) - F(m, n - 1)]^2
\]

where \( F(m, n) \) is the pixel value of the fusion image \( F \) at position \((m, n)\), and \( M \times N \) represents the image size.

Correlation coefficient represents the correlation between the fusion image and the source images. The larger the value of the correlation coefficient is, the better the quality of the fusion image is. The formula for calculating CC is as follows:

\[
CC(s, f) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (s_{i, j} - \bar{s}) (f_{i, j} - \bar{f})}{\sqrt{\left(\sum_{i=1}^{M} \sum_{j=1}^{N} (s_{i, j} - \bar{s})^2 \right) \left(\sum_{i=1}^{M} \sum_{j=1}^{N} (f_{i, j} - \bar{f})^2 \right)}}
\]

where \( s \) and \( f \) are the source image and fusion images of size \( M \times N \), respectively. \( \bar{s} \) and \( \bar{f} \) are the mean values of the source image and fusion images, respectively.

Mutual information means that the fusion image retains the amount of information in the source image. The larger the mutual information value is, the better the image fusion performance is. The formula for calculating MI is as follows:

\[
MI_{IF}(i, f) = \sum_{i, f} P_{IF}(i, f) \log \frac{P_{IF}(i, f)}{P_I(i) P_F(f)}
\]

Structural similarity can express the similarity of the structural texture between the fusion images and the source images. It is used to measure the structural similarity between the source images and the fusion images. The larger the structural similarity value is, the better the quality of the fusion image is. The formula for calculating SSIM is as follows:

\[
SSIM = \frac{(2\mu_s \mu_f + C_1)(2\sigma_{sf} + C_2)}{(\mu_s^2 + \mu_f^2 + C_1)(\sigma_s^2 + \sigma_f^2 + C_2)}
\]

where \( s \) and \( f \) represent the source image and the fusion image, respectively, \( \mu_s \) and \( \mu_f \) are their average values, \( \sigma_s^2 \) and \( \sigma_f^2 \) are their variances, \( \sigma_{sf} \) are the covariances between them; \( C_1 \) and \( C_2 \) are variable used to stabilize the denominator.

\( Q_{ABF} \) can express the degree to which the salient information of the source images is contained in the fusion images through local analysis. The higher the value is, the better the quality of the fused image is. The formula for calculating \( Q_{ABF} \) is as follows:

\[
Q(a, b, f) = \frac{1}{|W|} \sum_{\omega \in W} (\lambda(\omega) Q_0(a, f | \omega) + (1 - \lambda(\omega)) Q_0(b, f | \omega))
\]

where \( W \) is the family of all windows and \( |W| \) is the cardinality of \( W \). \( \lambda(\omega) \) is a local weight, \( a \) and \( b \) are the pixels, \( Q_0(\cdot) \) is the local quality index.

**B. FUSION RESULTS AND EVALUATION**

Figure 3 and Figure 4 show the fusion results of the proposed algorithm and the other five comparison algorithms. Due to space limitations, this paper only shows the results of six pairs
FIGURE 3. Image fusion and comparison results. ((a1, b1, c1) are infrared images, (a2, b2, c2) are visible images, (a3-a8, b3-b8, c3-c8) are the fusion results acquired by different methods).
FIGURE 4. Image fusion and comparison results. ((a1, b1, c1) are infrared images, (a2, b2, c2) are visible images, (a3-a8, b3-b8, c3-c8) are the fusion results acquired by different methods).
of images. Areas that require focus have been marked with red boxes. And the detail area that needs attention is enlarged.

1) SUBJECTIVE EVALUATION
As can be seen from the enlarged detail area in the Figure 3 and Figure 4, the proposed algorithm can retain more detailed information, and there is less artificial information. In addition, its results have good contrast, which meets the needs of human vision. In comparison, the FusionGAN method has serious lack of details and is relatively blurred as a whole. As shown in Figure 3 (a3), the trees in the background can no longer see detailed features, and as shown in Figure 3 (c3), the contrast of some fused images is too high and the visual effect is poor. The fusion results obtained by the ResNet method have a low contrast of significant targets. Although the fusion results of GANMcC have obvious target contrast, they also have the phenomenon of loss of details. The fusion images obtained by the GTF method preserve the edge structure well, but the internal details are also blurred. The overall performance of the CNN method is better than the other four compared methods, but compared with the fusion results of the proposed algorithm, some areas are overexposed, resulting in poor visual effects, as shown in Figure 4 (b4). Therefore, the proposed algorithm can transfer the significant information from the source image into the fusion image, and can obtain the best visual performance in terms of contrast and detail texture, which means the proposed method performs well subjectively.

2) OBJECTIVE EVALUATION
Visual evaluation analysis is usually subjective, so it is also essential to use different indicators to evaluate the fusion results objectively. The results are shown in Figure 5. According to the data in the figure, it can be seen that the proposed method has outstanding performance in various indicators, especially the indicators EN, MI and SSIM always perform better. In all quantitative evaluations, there are only a few places that are not optimal, but this does not affect the advantages of our method.

In addition, in order to enhance the reliability of the experimental results, we select 20 pairs of image fusion results for quantitative experiments, and calculates the average value of each index in 20 pairs of images of different algorithms. The results are shown in Table 1. The data in the table also show that the proposed algorithm has significantly higher objective evaluation index values than other algorithms, which further proves the effectiveness of the proposed algorithm.

To sum up, for the fusion of infrared and visible images, the proposed algorithm has good performance both subjectively and objectively.

C. ABLATION EXPERIMENTS
1) ABLATION EXPERIMENT OF GENERATOR MODEL
In the generator model proposed in this paper, the contrast attention module and the visible image cascading part are added. The contrast attention module is designed to make the fused image have more contrast information of the infrared image. After the cascading of visible image, the fused image has obvious advantages in preserving the detail texture. Therefore, in order to verify the effectiveness of the proposed model, corresponding ablation experiments were carried out, including four cases, namely: (1) The experiment of removing the contrast attention module and the visible image cascading part at the same time; (2) The experiment of removing the contrast attention module and retaining only the part of the visible image cascading; (3) The experiment of removing the part of the visible image cascading and retaining only the contrast attention module; (4) Both are retained.

Figure 6 shows the results of the ablation experiment. It can be seen that in the first case, the contrast and detail retention of the fused image are not good, and even some areas have the phenomenon of missing details. In the second case, the retention of detailed texture information is improved compared to the first case, but the contrast information is still not obvious enough. The third case is much more contrasty than the first case, but the detail retention is still insufficient. The fourth case is the best performer in terms of visual performance, contrast, and detail retention.

In addition, we use six evaluation metrics including EN, SF, CC, MI, SSIM, and Q_{ABF} to conduct quantitative experiments on the fusion results in different situations, and the obtained results are shown in Table 2. As can be seen from Table 2, the results are consistent with Figure 6. The six index values of Case 1 are the lowest, indicating that the fusion result is the worst in this case. Some index values in Case 2 and Case 3 are improved compared to the values in Case 1, but still

### Table 1. Average quantitative evaluation results of the fused image.

| index | FusionGAN | CNN   | ResNet | GTF   | GANMcC | Proposed |
|-------|-----------|-------|--------|-------|--------|----------|
| EN    | 6.286     | 6.637 | 6.508  | 6.542 | 6.426  | 6.761    |
| SF    | 3.233     | 3.664 | 3.41   | 3.846 | 3.392  | 4.079    |
| CC    | 0.475     | 0.531 | 0.507  | 0.555 | 0.491  | 0.57     |
| MI    | 0.953     | 1.291 | 1.523  | 1.465 | 1.035  | 1.643    |
| SSIM  | 1.079     | 1.362 | 1.342  | 1.374 | 1.362  | 1.484    |
| Q_{ABF} | 0.372 | 0.387 | 0.344  | 0.424 | 0.432  | 0.478    |
not very good. The six index values of the fourth case are all the highest, which also shows that the improved part added in this paper is effective.

Therefore, it can be concluded that the design of the contrast attention module and the parts of the visible image cascade are complementary, and they work together to make the generator produce the expected fusion results.

2) ABLATION EXPERIMENT OF LOSS FUNCTION

In this work, two loss functions are adopted to update optimization network parameters, namely content loss, and adversarial loss. Content loss controls the extraction of information types, and adversarial loss controls the balance of information content. Specifically, in content loss, intensity loss and contour loss are included. The former controls infrared contrast information, and the latter controls texture detail information. Through the constraint on content loss, the source image can obtain complementary information, making the extracted information more detailed and comprehensive. Through the constraint on adversarial loss, complementary information can be fitted in the form of a game, making the extracted information more integrated and
balanced. To verify the effectiveness of the loss function, ablation experiments of the loss function are performed. Specifically, it is divided into five cases, namely: (1) The discriminator is removed, and only the content loss is used to guide the optimization of the generator. (2) The content loss is removed, and only the adversarial loss is kept. (3) The intensity loss is eliminated, and the network is trained only by the adversarial loss and the contour loss. (4) The contour loss is removed and the network is only trained with adversarial loss and intensity loss. (5) Both are retained.

Figure 7 shows the results of the ablation experiment. It can be seen that when there is only content loss, the fusion results are confusing. The reason for this is that although the fusion results have the characteristic information of infrared images and visible images, due to the lack of adversarial loss, the distribution of the two kinds of information is unbalanced, and the matching performance with the source image is poor. However, the feature information in the source images is still well extracted. When only the adversarial loss is retained, the texture details and contrast information of the fusion result are not rich enough, and the fusion image is blurry. The reason for this phenomenon is that the ability to extract feature information is poor due to the lack of content loss. The third and fourth cases correspond to lack of contrast detail and lack of texture detail, respectively. Finally, when both content loss and adversarial loss are included, the fusion results not only have obvious contrast and rich texture details, but also are well-balanced in the distribution of fused images, with good visual performance.

Similarly, we use six evaluation metrics including EN, SF, CC, MI, SSIM, and $Q_{ABF}$ to conduct quantitative experiments on the fusion results of different situations, and the obtained results are shown in Table 3. The results in Table 3 are also consistent with Figure 7. It can be seen from Table 3 that the six index values of Case 5 are all the best, while the quantitative performance of the other cases is reduced to varying degrees.

From the above analysis, it can be concluded that content loss and adversarial loss are complementary, and they work together to obtain well-performing fusion results.

### TABLE 2. Quantitative evaluation results of generator model ablation.

| index | Case 1 | Case 2 | Case 3 | Case 4 |
|-------|--------|--------|--------|--------|
| EN    | 6.433  | 6.678  | 6.879  | 7.129  |
| SF    | 4.012  | 4.567  | 4.952  | 5.201  |
| CC    | 0.436  | 0.577  | 0.543  | 0.69   |
| MI    | 0.768  | 1.434  | 1.438  | 1.952  |
| SSIM  | 0.876  | 0.987  | 1.011  | 1.389  |
| $Q_{ABF}$ | 0.423 | 0.501  | 0.493  | 0.582  |

### 3) ABLATION EXPERIMENT OF REWARD FUNCTION MODULE

In the discriminator, in order to improve the discriminator’s discriminative ability and thus promote the generator to produce better fused images, a target-guided reward function module is designed. This module can more accurately discriminate the images sent by the generator. Therefore, in order to verify the effectiveness of the proposed reward function module, corresponding ablation experiments are carried out, including two cases, namely: (1) The experiment of removing the reward function module. (2) The experiment of retaining the reward function module.

![FIGURE 6. Experiment results of generator model ablation.](image-url)
Case 1 is the situation where the reward function module is removed, and Case 2 is the situation where the reward function module is retained. From the experimental results of the two pairs of pictures, it can be seen that after adding the reward function module, the fusion results have achieved better visual effects, and the fusion of the detailed features and contrast salient information of the images is better than removing the reward function.

We still use six evaluation metrics including EN, SF, CC, MI, SSIM, and $Q_{ABF}$ to conduct quantitative experiments on the fusion results, and the obtained results are shown in Table 4. It can be seen from Table 4 that the quantitative index value of Case 2 is the best in the fusion results of the two pairs of images, which is also consistent with the analysis of Figure 8. To sum up, the setting of the reward function module is very necessary.

### D. COMPUTATIONAL EFFICIENCY

To test the computational efficiency of the proposed algorithm, the comparison algorithm and the proposed algorithm are placed in the same experimental environment, and then we separately calculate the average time it takes for them to complete the fusion of multiple pairs of images. Since the training time of each algorithm is too long, only the time used in the test experiment of each algorithm is calculated. Table 5 shows the average time for different algorithms to complete the fusion of 20 pairs of images.

From the results in Table 5, it can be seen that the proposed algorithm has relatively low computational efficiency.
TABLE 5. Average quantitative evaluation results of the fused image.

| Method  | FusionGAN | CNN | ResNet | GTF | GANMcC | Proposed |
|---------|-----------|-----|--------|-----|--------|-----------|
| Time/s  | 10.98     | 23.16 | 18.85  | 2.91 | 13.41  | 15.12     |

VOLUME 10, 2022

ACKNOWLEDGMENT
The authors would like to thank the editors for their support and the reviewers for their valuable comments.

REFERENCES
[1] Q. Li, G. Han, P. Liu, H. Yang, J. Wu, and D. Liu, “An infrared and visible image fusion method guided by saliency and gradient information,” IEEE Access, vol. 9, pp. 108942–108958, 2021.
[2] J. Han and B. Bhanu, “Fusion of color and infrared video for moving human detection,” Pattern Recognit., vol. 40, no. 6, pp. 1771–1784, 2007.
[3] S. R. Schnelle and A. L. Chan, “Enhanced target tracking through infrared-visible image fusion,” in Proc. 14th Int. Conf. Inf. Fusion, Jul. 2011, pp. 1–8.
[4] M. Kerkech, A. Hafiane, and R. Canals, “Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach,” Comput. Electron. Agric., vol. 174, Jul. 2020, Art. no. 105446.

[5] H. Li, W. Ding, X. Cao, and C. Liu, “Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing,” Remote Sens., vol. 9, no. 5, p. 441, 2017.

[6] Z. Wang, Y. Wu, and Q. Niu, “Multi-sensor fusion in automated driving: A survey,” IEEE Access, vol. 8, pp. 2847–2868, 2020.

[7] Z. Song, Z. Zhang, S. Yang, D. Ding, and J. Ning, “Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging,” Comput. Electron. Agric., vol. 179, Dec. 2020, Art. no. 105812.

[8] I. Misra, R. K. Gambhir, S. M. Moorthy, D. Dhar, and R. Ramakrishnan, “An efficient algorithm for automatic fusion of RISAT-1 SAR data and ResourceSat-2 optical images,” in Proc. 4th Int. Conf. Intell. Hum. Comput. Interact. (HCI), Dec. 2012, pp. 1–6.

[9] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, “Multiresolution-based image fusion with additive wavelet decomposition,” IEEE Trans. Geosci. Remote Sens., vol. 37, no. 3, pp. 1204–1211, May 1999.

[10] Q. Zhang and M. D. Levine, “Robust multi-focus image fusion using multi-task sparse representation and spatial context,” IEEE Trans. Image Process., vol. 25, no. 5, pp. 2045–2058, May 2016.

[11] H. Li, H. Qiu, Z. Yu, and Y. Zhang, “Infrared and visible image fusion scheme based on NSCT and low-level visual features,” Infrared Phys. Technol., vol. 76, pp. 174–184, May 2016.

[12] Y. Zuo, J. Liu, G. Bai, X. Wang, and M. Sun, “Airborne infrared and visible image fusion combined with region combination,” Sensors, vol. 17, no. 5, p. 1127, May 2017.

[13] Y. An, Z. Hao, and Z. Huo, “Infrared and visible image fusion based on two-scale decomposition and improved saliency detection,” J. Phys., Conf., vol. 1856, no. 1, Apr. 2021, Art. no. 012058.

[14] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, “Image fusion with convolutional sparse representation,” IEEE Signal Process. Lett., vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[15] H. Li, X.-J. Wu, and J. Kittler, “Infrared and visible image fusion using a deep learning framework,” in Proc. 24th Int. Conf. Pattern Recognit. (ICPR), Aug. 2018, pp. 2705–2710.

[16] 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, Aug. 2019.

[17] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, “DDGCAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion,” IEEE Trans. Image Process., vol. 29, pp. 4980–4995, 2020.

[18] D. Xu, Y. Wang, S. Xu, K. Zhu, N. Zhang, and X. Zhang, “Infrared and visible image fusion with a generative adversarial network and a residual network,” Appl. Sci., vol. 10, no. 2, p. 554, Jan. 2020.

[19] Y. Liu, X. Chen, Z. Wang, J. J. Wang, R. K. Ward, and X. Wang, “Deep learning for pixel-level image fusion: Recent advances and future prospects,” Inf. Fusion, vol. 42, pp. 158–173, Jul. 2018.

[20] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, “Infrared and visible image fusion with convolutional neural networks,” Int. J. Wavelets, Multiresolution Inf. Process., vol. 16, no. 3, 2018, Art. no. 1850018.

[21] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556.

[22] 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.

[23] H. Li, J. Xi, J. Wu, and T. S. Durrani, “Infrared and visible image fusion with ResNet and zero-phase component analysis,” Inf. Phys. Technol., vol. 102, Mar. 2019, Art. no. 103039.

[24] H. Li and X.-J. Wu, “DenseFuse: A fusion approach to infrared and visible images,” IEEE Trans. Image Process., vol. 28, no. 5, pp. 2614–2623, May 2018.

[25] 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, p. 27.

[26] J. Ma, C. Chen, C. Li, and J. Huang, “Infrared and visible image fusion via gradient transfer and total variation minimization,” Inf. Fusion, vol. 31, pp. 100–109, Sep. 2016.

[27] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, and J. Wu, “Infrared and visible image fusion via detail preserving adversarial learning,” Inf. Fusion, vol. 54, pp. 85–98, Feb. 2020.

[28] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, “Least squares generative adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2794–2802.

[29] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.

[30] T. Alexander, “TNO image fusion dataset,” Data Brief, vol. 15, pp. 249–251, Dec. 2017, doi: 10.6084/m9.figshare.1008029.

[31] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, “U2Fusion: A unified unsupervised image fusion network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 502–518, Jan. 2020.

[32] J. Ma, H. Zhang, Z. Shao, P. Liang, and H. Xu, “GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–14, 2021.

[33] J. W. Roberts, F. B. Ahmed, and J. A. Van Aardt, “Assessment of image fusion procedures using entropy, image quality, and multispectral classification,” J. Appl. Remote Sens., vol. 2, no. 1, 2008, Art. no. 023522.

[34] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, “Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 1, pp. 94–109, Jan. 2012.

[35] M. Deshmukh and U. Bhosale, “Image fusion and image quality assessment of fused images,” Int. J. Image Process., vol. 4, no. 5, pp. 484, 2010.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[37] Y. Chen and R. S. Blum, “A new automated quality assessment algorithm for night vision image fusion,” in Proc. 41st Annu. Conf. Inf. Sci. Syst., Mar. 2007, pp. 518–523.
YONGKE LI received the B.S. and M.S. degrees from the Ordnance Engineering College, Shijiazhuang, China. He is currently an Associate Professor with the Department of Unmanned Aerial Vehicle Engineering, Army Engineering University, Shijiazhuang. His main research interests include pattern recognition and scene classification.

CHANGLONG WANG received the B.S. degree in radar engineering from the Electronic Engineering College, Hefei, China, in 1987, the M.S. degree in signal and information processing from the Harbin Institute of Technology, Harbin, China, in 1992, and the Ph.D. degree in electronics and information technology from Xi’an Jiaotong University, Xi’an, China, in 2006. He is currently a Professor with the Department of Unmanned Aerial Vehicle Engineering, Army Engineering University, Shijiazhuang, China. His main research interests include pattern recognition and signal processing.

YONGJIANG HU received the Ph.D. degree from the Ordnance Engineering College of the Chinese People’s Liberation Army. He is currently working with Army Engineering University. His research interests include UAV mission planning and UAV target recognition and tracking control.

FUYU HUANG was born in Hebei, China, in 1985. He received the B.S., M.S., and Ph.D. degrees from the Department of Optical and Electronic Engineering, Mechanical Engineering College, Shijiazhuang, China, in 2008, 2010, and 2014, respectively. His research interests include optical signal processing, infrared detection, and optical design.

JIANGYI YAO received the B.S. degree from the Army Aviation Academy, Beijing, China. He is currently pursuing the M.S. degree in computer science and technology with Army Engineering University, Shijiazhuang. His main research interests include computer simulation and machine learning.