

Research Article

Adaptive Adjustment Object Detection Algorithm under Multiple Mechanisms Based on GAN

Zemin Qiu , Feng Wang , and Zhihong Pan

School of Information Science, Xinhua College of Sun Yat-Sen University, Guangzhou, China

Correspondence should be addressed to Zemin Qiu; qiuzemin@xhsysu.edu.cn

Received 16 April 2021; Revised 13 May 2021; Accepted 31 May 2021; Published 19 June 2021

Academic Editor: Chenxi Huang

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

Target tracking is prone to problems such as target loss and identity jump when the target is occluded and the attitude is changed. In order to solve this phenomenon, this paper proposes the adaptive adjustment object detection algorithm under multiple mechanisms based on GAN. This algorithm introduces a gradient penalty mechanism to the discriminator and uses the relative discriminator structure to reconstruct the discriminator, so as to improve the discriminatory ability of the discriminator. Then, through the feedback mechanism, the obtained data is fed back to the generator network in time, and the genetic mechanism is used to speed up the positioning of the key areas of the image. Experimental results show that compared with other existing algorithms, this algorithm can effectively locate and distinguish under different environments. And, the target still maintains a high resolution. When the target is occluded, it can effectively avoid the phenomenon of target loss and identity jump.

1. Introduction

Target detection is one of the important research directions in the field of computer vision. It integrates technologies such as positioning, recognition, and classification and is widely used in fields such as intelligent monitoring and human-computer interaction. In recent years, the target detection algorithm has made great breakthroughs. The representative target detection algorithms can be divided into two main categories. One is the R-CNN algorithm (R-CNN, Fast R-CNN, Faster R-CNN, etc.) [1, 2]. The algorithm needs to generate the target first. Candidate frame (target location), and then, classify and regress the candidate frame. The other is one-stage algorithms such as Yolo and SSD [3–5], which use the convolutional neural network (CNN) to extract image features and establish detection models to classify and locate targets. After comparison, the second type of algorithm is faster, but the accuracy is lower. In practical applications, the target in the image often has problems such as being occluded and deformed, making the extracted target feature insignificant or lacking part of the feature information, leading to classification errors and tracking failures,

and then prone to the problem of identity jump. Therefore, we are comparing the advantages and disadvantages of existing algorithms; this paper adopts the generative confrontation network in the second type of algorithm. This model is a unique network structure that captures potential data distribution. The true and false confrontation error design based on game theory construction makes it possible to deal with different types of tasks [6, 7].

Based on the idea of generative confrontation, this paper proposes a multimechanism adaptive target detection algorithm. First, through the organic combination of the punishment mechanism, feedback mechanism, and scoring mechanism, it effectively solves the problems of difficulty in training and model collapse of the generative confrontation network. Then, use the relative discriminator structure to construct the discriminator to improve the probability of the discriminator in the process of fighting competition. The result is sent to the generator network through the feedback mechanism, and finally, the individual with the optimal value is obtained through the scoring mechanism, so as to quickly and effectively locate the key area of the image. After experiments, the experimental results are basically in line

with expectations. The target detection algorithm can effectively locate and distinguish different objects and still maintain a high resolution for the occluded target, while maintaining a low number of missing alarms and false alarms.

2. Generative Adversarial Network (GAN)

The generative adversarial network (GAN) was proposed by Goodfellow. It uses two convolutional neural networks using game training to generate samples similar to the original picture [8–10].

By observing the structure of the generative confrontation network in Figure 1, GAN is mainly composed of two parts: generation network (G) and discriminant network (D). And, it uses a competing mechanism [11–13]; the purpose is to enable the generator to generate data similar to the real data distribution, which made the GAN achieve the effect of being fake. Aiming at the problem of data prediction, the generator is trained to learn the data distribution and predict the image distribution data. Generated data distribution and real data distribution are shown in Figure 2.

Because the difference between the generated data distribution P_G and the real data distribution P_{data} cannot be calculated, therefore, the discriminator uses data sampling to calculate the divergence of P_G and P_{data} . And, use these to discriminate the difference between the two data distributions, and the objective function of the discriminator is

$$V(G, D) = E_{x \sim P_{\text{data}}} [\log D(x)] + E_{x \sim P_G} [\log(1 - D(x))]. \quad (1)$$

Among them, $E(*)$ is mathematical expectation and $D(*)$ is the output result of the discriminator. The objective function is to maximize the discriminator, equivalent to the divergence generated data and real data. The optimal generator achieves the purpose of prediction by minimizing the divergence of the distribution between the real image data and the generated data. The optimal generator can be expressed as

$$G^* = \arg \min_G \max_D V(G, D). \quad (2)$$

GAN can generate samples that are similar to the real data distribution, but it has problems such as difficulty in training, difficulty in network convergence, and mode collapse.

3. The Discriminator Network Introduces a Gradient Penalty Mechanism

In order to make the training of the model more stable and solve the problem of model collapse, this paper adds a gradient penalty to the discriminator objective function. After adding the gradient penalty, the objective function is

$$V(G, D) = E_{x \sim P_{\text{data}}} [D(x)] - E_{x \sim P_G} [D(x)] - \lambda_{gp} E_{x \sim P_{\text{penalty}}} \left[\left(\|\nabla_x D(x)\|_2 - 1 \right)^2 \right]. \quad (3)$$

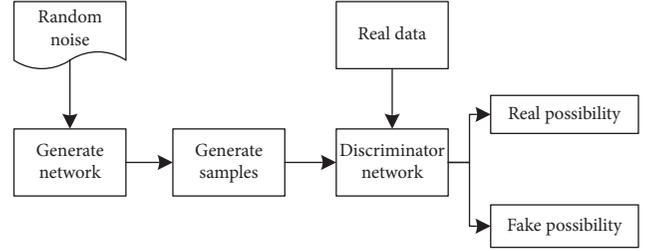


FIGURE 1: Generative adversarial network structure.

Among them, λ_{gp} is the weight parameter. P_{penalty} is the probability distribution of the sample for calculating the gradient penalty. Sampling a point in P_G and P_{data} , respectively, the method is random sampling. $\|\cdot\|_2$ is a two-norm number.

In the proposed data prediction algorithm, the loss function can reduce the difference in probability distribution between the predicted data and the real data so that the predicted data generated by the generator can more accurately fit the real data distribution. The loss function is defined as

$$L(G, D) = E_{x \sim P_{\text{data}}} [D(x)] - E_{x \sim P_G} [D(x)] - \lambda_{gp} E_{x \sim P_{\text{penalty}}} \left[\left(\|\nabla_x D(x)\|_2 - 1 \right)^2 \right]. \quad (4)$$

Adding a traditional loss function to the loss function of GAN can improve the accuracy of the data generated by the generator. The role of the generator is not only to generate data that can deceive the discriminator, but the generated data should be close enough to the real data. Since data prediction is a regression problem, adding the output result data of the discriminator to the loss function can punish the extreme error between the predicted value and the true value. Therefore, the loss function after introducing the gradient penalty is

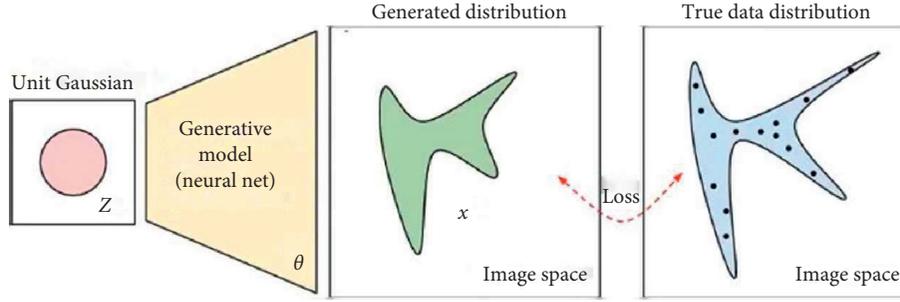
$$L_{\text{MSE}}(G, D) = L(G, D) + \lambda_{\text{MSE}} L_{\text{MSE}}(G), \quad (5)$$

where λ_{MSE} is the weight parameter, and the loss function of $L_{\text{MSE}}(G)$

$$L_{\text{MSE}}(G) = E_{x \sim P_{\text{data}}} [\|x - G(x)\|_2]. \quad (6)$$

Through research, it can be found that $L_{\text{MSE}}(G)$ is the loss function that is only extracted for the generator. Just considering the generator data to improve the accuracy of the algorithm is not enough. In GAN, the data generated by the discriminator corresponds to the generator has a guiding role.

Therefore, this paper adopts a feedback mechanism for the data output by the discriminator network based on the introduction of a gradient penalty mechanism in the generator. This improves the accuracy of the loss value $L_{\text{MSE}}(G)$. The discriminator in this article is constructed with a relative discriminator structure to estimate the probability that the real image is more real than the fake image and then promote the discriminator to obtain better probability results in the process of fighting competition, thereby increasing the

FIGURE 2: Generated data distribution P_G and real data distribution P_{data} .

loss of the generator value $L_{\text{MSE}}(G)$ to strengthen the network parameter optimization of the generator.

By collecting the real data and false data of the discriminator and estimating the corresponding sample expectations, the following function is obtained:

$$\begin{aligned} D_{\text{RI}}(I_{\text{real}}, I_{\text{fake}}) &= \sigma\{D(I_{\text{real}}) - E_{I_{\text{fake}}}(D(I_{\text{fake}}))\}, \\ D_{\text{RI}}(I_{\text{fake}}, I_{\text{real}}) &= \sigma\{D(I_{\text{fake}}) - E_{I_{\text{real}}}(D(I_{\text{real}}))\}. \end{aligned} \quad (7)$$

Among them, $E_{I_{\text{real}}}$ and $E_{I_{\text{fake}}}$ are the sample expectations of real data and fake data, respectively. This paper uses the minimum loss function to calculate.

For the global discriminator, the ideal values of $E_{I_{\text{fake}}}(D(I_{\text{fake}}))$ and $E_{I_{\text{real}}}(D(I_{\text{real}}))$ should be close to 0 and 1. The expected discrimination result is $D(I_{\text{real}}) = 1$ and $D(I_{\text{fake}}) = 0$. Therefore, the regression targets of $D_{\text{RI}}(I_{\text{real}}, I_{\text{fake}})$ and $D_{\text{RI}}(I_{\text{fake}}, I_{\text{real}})$ are 1 and 0.

For the generator, its role is to confuse the authenticity judgment of the discriminator as much as possible, which is opposite to the return goal of the discriminator. The regression targets of $D_{\text{RI}}(I_{\text{real}}, I_{\text{fake}})$ and $D_{\text{RI}}(I_{\text{fake}}, I_{\text{real}})$ are 0 and 1. Therefore, the adversarial function to obtain the global discriminator is defined as follows:

$$\begin{aligned} L_{\text{MSE}}(G) &= \frac{1}{2}E_{I_{\text{fake}}}\{D_{\text{RI}}(I_{\text{fake}}, I_{\text{real}})^2\} \\ &+ \frac{1}{2}E_{I_{\text{real}}}\{[D_{\text{RI}}(I_{\text{real}}, I_{\text{fake}}) - 1]^2\}. \end{aligned} \quad (8)$$

In this paper, by setting the feature-preserving loss function, the confrontation function $L_{\text{MSE}}(G)$ of the global discriminator is input into the above loss function for feature error calculation; finally, the loss function is defined as follows:

$$L(G, D) = E_{x \sim P_{\text{data}}}[D(x)] - E_{x \sim P_G}[D(x)] - \lambda_{gp} E_{x \sim P_{\text{penalty}}}\left[\left(\|\nabla_x D(x)\|_2 - 1\right)^2\right], \quad (9)$$

$$L_{\text{MSE}}(G, D) = L(G, D) + \lambda_{\text{MSE}}\left(\frac{1}{2}E_{I_{\text{fake}}}\{D_{\text{RI}}(I_{\text{fake}}, I_{\text{real}})^2\} + \frac{1}{2}E_{I_{\text{real}}}\{[D_{\text{RI}}(I_{\text{real}}, I_{\text{fake}}) - 1]^2\}\right).$$

By observing Figures 3–5, it can be seen that, after the reconstruction of the generated network in this paper, the image is basically effectively repaired, and it has a good real consistency. Although the repaired part is somewhat unclear and may have subtle changes, but it basically does not affect the detection effect.

4. Generating Network Scoring Mechanism

However, for the targets in the video, they are basically in a state of movement and overlap. We need to quickly and effectively locate the key areas of the image [14, 15]. Therefore, based on the above formula, this paper proposes a

scoring mechanism for the generative network in the generative confrontation network. Each iteration of the generator network is composed of three stages: mutation, evaluation, and selection so that the generator can quickly and accurately locate the key areas in the image.

4.1. Variation. Mainly for the input random noise signal z , the gradient penalty loss function of this article is used in the new round of update process. The penalty loss function $L(G, D)$ for the generating network gradient is calculated only for noise data, and finally, $L_{\text{MSE}}(G, D)$ is obtained and the network data is updated:

$$L_z(G, D) = E_{z \sim P_z}[D(z)] - E_{z \sim P_z}[D(z)] - \lambda_{gp} E_{z \sim P_z}\left[\left(\|\nabla_z D(z)\|_2 - 1\right)^2\right], \quad (10)$$

$$L_{\text{MSE}}(G, D) = L_z(G, D) + \lambda_{\text{MSE}}\left(\frac{1}{2}E_{I_{\text{fake}}}\{D_{\text{RI}}(I_{\text{fake}}, I_{\text{real}})^2\} + \frac{1}{2}E_{I_{\text{real}}}\{[D_{\text{RI}}(I_{\text{real}}, I_{\text{fake}}) - 1]^2\}\right).$$



FIGURE 3: Original image (occluded).



FIGURE 4: Generated image.

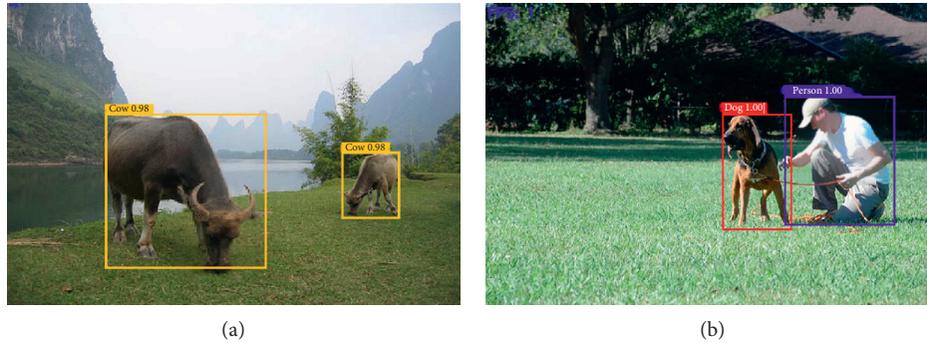


FIGURE 5: Object detection.

Among them, λ_{gp} is the weight parameter and P_z is the distribution of noise data.

4.2. Evaluation. This stage is the data evaluation stage. The network uses the above loss function and selects the best sample under the feedback of the discriminator, so as to learn the correct training sample distribution and improve the stability of model training. The evaluation mechanism is as follows.

This stage is quantified as an adaptability score, which mainly analyzes the quality of the samples generated by the network and the diversity of the generated samples. It evaluates the quality of the generated samples and the

diversity of the generated samples; the adaptability score can be expressed as

$$\mathbf{F} = \mathbf{F}_q + \gamma \mathbf{F}_d. \quad (11)$$

Among them, \mathbf{F}_q is the quality score of the generated sample and \mathbf{F}_d is the diversity score of the generated sample. It is used to measure whether the generator can fully disperse the generated samples, which can further avoid mode collapse. γ is a super parameter that balances the proportion of the \mathbf{F}_q and \mathbf{F}_d .

Calculate the highest fitness score. And, compare it with the highest score on the network to get the global best score.

4.3. Selection. Select the best individual according to the evaluation mechanism. Others with lower scores will

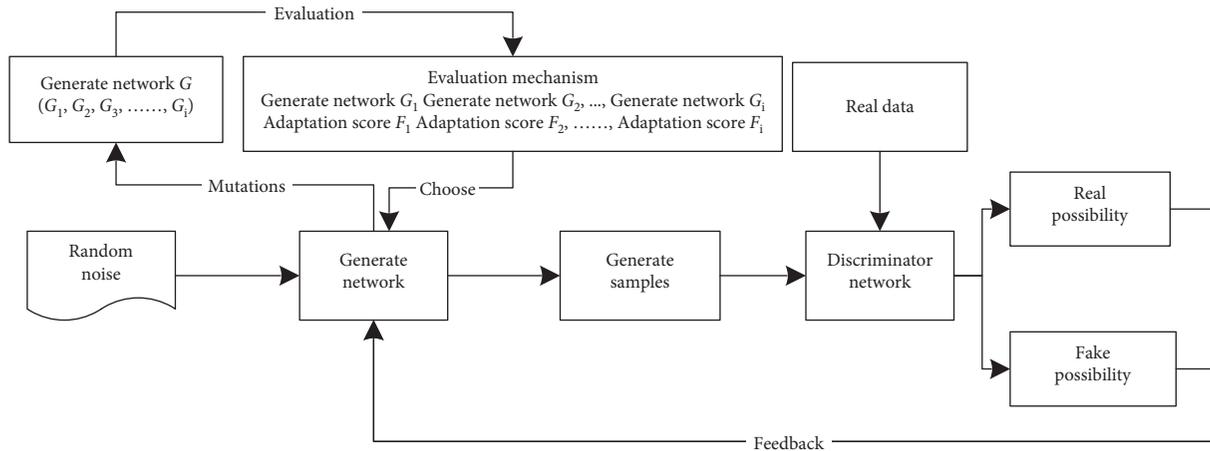


FIGURE 6: The structure of the article generation adversarial network.



FIGURE 7: Effect picture of object detection.

gradually decay so that the generated network has a certain directionality, and the best value sample is selected.

The structure of the article generation adversarial network is shown in Figure 6.

5. Simulation Experiment

This paper uses the ImageNet dataset and the MOT training dataset to test the algorithm. All experiments in this paper are carried out in the same experimental environment, and

TABLE 1: Evaluation criteria in MOTChallenge.

Measure	Better	Perfect	Description
MOTA	Higher	100%	Tracking accuracy: it is negatively correlated with the number of FN, FP, and IDs
MOTP	Higher	100%	Tracking accuracy, matching of GT, and detected bbox overlap
IDF1	Higher	100%	F1 introducing track ID
FP	Lower	0	The total number of FPs; false positives are false positives
FN	Lower	0	The total number of FNs; false negatives are also missed
IDs	Lower	0	Total number of ID changes
Hz	Higher	Inf	Processing speed, excluding the time consumption of the detector

the data obtained by the article algorithm are compared with traditional algorithms, YOLO-based target detection, Deep Sort, and literature algorithms [16] to verify the effectiveness of the algorithm.

Considering that there is a big difference between artificial occlusion and occlusion in the real world, this chapter directly simulates the actual dataset, and the result of target detection is shown in the figure below.

By observing the target detection effect map in Figure 7, we can see that the article algorithm can effectively locate and distinguish different objects and still maintain a high resolution for the occluded target, even if the characters in Figures 7(d) and 7(i) are occluded. More than half of the area can still accurately identify and mark the target.

5.1. Accuracy Test. This paper uses the evaluation criteria in MOTChallenge to evaluate the performance of the proposed algorithm and compares the measured data with traditional target detection algorithms and literature algorithms. Among them, the MOTChallenge dataset has a comprehensive data type, which includes different viewing angles, camera movement methods, and pictures taken in different climates, which can effectively evaluate algorithms. Standards are shown in Table 1.

This article uses evaluation criteria: MOTA, MOTP, IDs, FP, and FN. By observing the identity jump problem of the algorithm in the target detection process. Through the test, we can obtain the data as shown in Table 2.

By observing the target tracking effect data comparison table in Table 2, we can see that the MOTA, IDs, and FP in this paper are better than other algorithms. In terms of MOTP data, the article algorithm is only 0.4% higher than the literature algorithm. It can be seen that the article algorithm can preserve better detection accuracy, while performing well in the number of false alarms (FP) and number of missed alarms (FN). It can effectively prove that the article algorithm can effectively reduce the problem of identity jumps.

By observing Figure 8, it can be seen that, under the same accuracy, the number of identity jumps of the algorithm in this paper is smaller than that of other algorithms. And, in terms of the highest accuracy, the article algorithm is about 2% higher than the literature algorithm. Observing Figure 9, one can clearly see that the

TABLE 2: Comparison table of target tracking effect data.

Algorithm	MOTA	MOTP	IDs	FP	FN
Traditional algorithm	35.9	62.5	923	14332	73122
YOLO object detection algorithm	70.7	70.9	936	8742	53386
Deep Sort	63.2	81.2	743	11261	62311
Document algorithm	71.2	82.6	257	4410	31001
Article algorithm	72.1	83.0	122	2738	34219

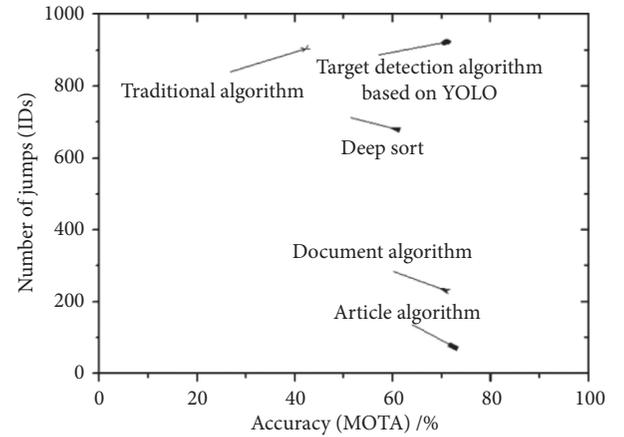


FIGURE 8: Comparison of accuracy and number of identity jumps.

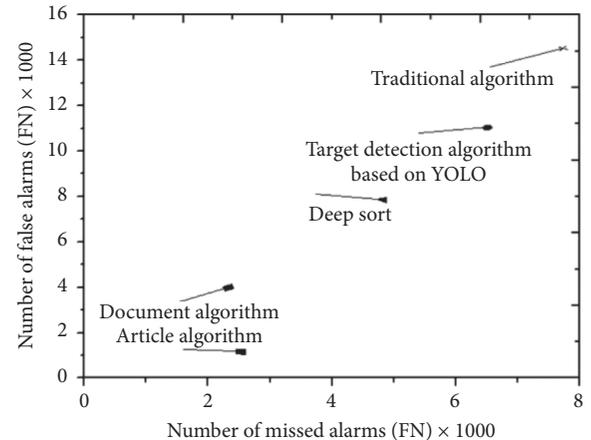


FIGURE 9: Comparison of the number of missed alarms and the number of false alarms.

algorithm in this paper is in the lower left corner of the image, and the number of missed alarms and false alarms are the lowest.

6. Conclusion

There are problems such as target loss and identity jump in current target detection. In response to these phenomena, we propose a GAN-based multimechanism adaptive target detection algorithm. The article optimizes the construction model of GAN by introducing gradient penalty and feedback and scoring mechanism into the original GAN, which achieves an effective improvement in detection accuracy. Experimental results show that compared with other existing algorithms, the article algorithm has a more stable training model, which can effectively reduce the problem of target identity jump, while maintaining high detection accuracy. Although the method is effective, there are still areas that can be improved, such as the use of a more appropriate size convolution kernel to extract features, which will continue to be studied in future works.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

The authors have equally contributed to the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This research was supported by the Scientific Research Platforms and Project of Colleges and Universities in Guangdong Province under Grant no. 2018KTSCX316, Science and Technology Project of Guangzhou under Grant no. 202002030273, and Discipline Project, Xinhua College of Sun Yat-sen University with no. 2020XZD02.

References

- [1] R. Yun, Z. Changren, and X. Shunping, "Object detection based on fast/faster RCNN employing fully convolutional architectures," *Mathematical Problems in Engineering*, vol. 2018, Article ID 3598316, 7 pages, 2018.
- [2] L. Zhang, Y. Zhang, and Z. Zhang, "Real-time water surface object detection based on improved faster R-CNN," *Sensors*, vol. 19, no. 16, p. 3523, 2019.
- [3] X. Zhang, X. Dong, Q. Wei et al., "Real-time object detection algorithm based on improved YOLOv3," *Journal of Electronic Imaging*, vol. 28, no. 5, p. 1, 2019.
- [4] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," *Electronics*, vol. 9, no. 3, pp. 1–11, 2020.
- [5] C. Sun, Y. Ai, S. Wang et al., "Mask-guided SSD for small-object detection," *Applied Intelligence*, vol. 51, no. 20, pp. 1–12, 2020.
- [6] Z. Wu, C. He, L. Yang et al., "Attentive evolutionary generative adversarial network," *Applied Intelligence*, vol. 51, no. 6, pp. 1–15, 2020.
- [7] Q. Cui, H. Sun, Y. Kong, X. Zhang, and Y. Li, "Efficient human motion prediction using temporal convolutional generative adversarial network," *Information Sciences*, vol. 545, pp. 427–447, 2021.
- [8] B. Jiang, Z. Zhou, X. Wang et al., "cmSalGAN: RGB-D salient object detection with cross-view generative adversarial networks," *IEEE Transactions on Multimedia*, vol. 1, pp. 1–11, 2020.
- [9] J. Li, C. Qu, S. Peng, and Y. Jiang, "Ship detection in SAR images based on generative adversarial network and online hard examples mining," *Journal of Electronics & Information Technology*, vol. 41, no. 1, pp. 143–149, 2019.
- [10] G. Chen, L. Liu, J. Guo, Z. Pan, and W. Hu, "Semi-supervised airplane detection in remote sensing images using generative adversarial networks," *Journal of University of Chinese Academy of Sciences*, vol. 37, no. 4, pp. 539–546, 2020.
- [11] C. Wang and J. Tian, "Fine-grained inshore ship recognition assisted by deep-learning generative adversarial networks," *CAAI Transactions on Intelligent Systems*, vol. 15, no. 2, pp. 296–301, 2020.
- [12] X. Zhai, Z. Cheng, Y. Wei et al., "Compressive sensing ghost imaging object detection using generative adversarial networks," *Optical Engineering*, vol. 58, no. 1, p. 1, 2019.
- [13] Z. Liu, W. Zhang, and P. Zhao, "A cross-modal adaptive gated fusion generative adversarial network for RGB-D salient object detection," *Neurocomputing*, vol. 387, pp. 210–220, 2020.
- [14] L. Courtrai and M. T. Pham, "Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks," *Remote Sensing*, vol. 12, no. 19, p. 3152, 2020.
- [15] A. N. Shan, S.-k. Lin, J.-z. Qiao, and C.-h. Li, "Object detection via learning occluded features based on generative adversarial networks," *Control and Decision*, vol. 1, pp. 1–7, 2020.
- [16] J. Li, X. Duan, M. Xu, and G. Xue, "Salient object detection based on conditional generative adversarial networks for video," *Transducer and Microsystem Technologies*, vol. 38, no. 11, pp. 129–132, 2019.
- [17] Y. Wei, C.-q. Xu, Z.-f. Diao, and B.-q. Li, "A multi-target pedestrian tracking algorithm based on generated adversarial network," *Journal of Northeastern University (Natural Science)*, vol. 41, no. 12, pp. 1673–1679, 2020.