electronics

MDPI

Article

Infrared Image Small-Target Detection Based on Improved FCOS and Spatio-Temporal Features

Shengbo Yao [1], Qiuyu Zhu [1,*], Tao Zhang [2], Wennan Cui [2] and Peimin Yan [1]

[1] School of Communication & Information Engineering, Shanghai University, 99 Shangda Road, Baoshan District, Shanghai 200444, China; yshengbo@shu.edu.cn (S.Y.); pmyan@shu.edu.cn (P.Y.)
[2] Key Laboratory of Intelligent Infrared Perception, Chinese Academy of Sciences, Shanghai 200083, China; sitp_710@mail.sitp.ac.cn (T.Z.); cuiwennan@mail.sitp.ac.cn (W.C.)
* Correspondence: zhuqiuyu@staff.shu.edu.cn

Abstract: The research of infrared image small-target detection is of great significance to security monitoring, satellite remote sensing, infrared early warning, and precision guidance systems. However, small infrared targets occupy few pixels and lack color and texture features, which make the detection of small infrared targets extremely challenging. This paper proposes an effective single-stage infrared small-target detection method based on improved FCOS (Fully Convolutional One-Stage Object Detection) and spatio-temporal features. In view of the simple features of infrared small targets and the requirement of real-time detection, based on the standard FCOS network, we propose a lightweight network model combined with traditional filtering methods, whose response for small infrared targets is enhanced, and the background response is suppressed. At the same time, in order to eliminate the influence of static noise points in the infrared image on the detection of small infrared targets, time domain features are added to the improved FCOS network in the form of image sequences, so that the network can learn the spatio-temporal correlation features in the image sequence. Finally, compared with current typical infrared small-target detection methods, the comparative experiments show that the improved FCOS method proposed in this paper had better detection accuracy and real-time performance for infrared small targets.

Keywords: infrared small-target detection; deep learning; FCOS; spatio-temporal features; maximum filter

Citation: Yao, S.; Zhu, Q.; Zhang, T.; Cui, W.; Yan, P. Infrared Image Small-Target Detection Based on Improved FCOS and Spatio-Temporal Features. *Electronics* **2022**, *11*, 933. https://doi.org/10.3390/electronics11060933

Academic Editor: Abdeldjalil Ouahabi

Received: 15 February 2022
Accepted: 14 March 2022
Published: 17 March 2022

1. Introduction

Infrared small-target detection technology has always been a research difficulty and focus in infrared image processing and has important applications in security monitoring, military reconnaissance, night driving, and other fields. However, small infrared targets occupy few pixels, are usually submerged in complex backgrounds and noises, and lack features, such as texture or color, that are required for deep learning. Therefore, the problem of infrared small-target detection is a more challenging subject. At present, infrared small-target detection technology can be divided into two categories: traditional methods and deep-learning-based methods.

Traditional infrared small-target detection methods include a series of methods based on morphological Top-hat filtering, Max–Mean, and Local Contrast Method (LMS), etc. [1–3]; however, when the target's signal-to-noise ratio is low, these methods make it difficult to determine a suitable template, which leads to reduced recall and poor detection performance. Inspired by the human visual system, many infrared small-target detection methods have been proposed, such as local contrast detection, local difference detection, and improved local contrast detection.

These methods strengthen the difference between background and infrared targets. There are other methods to separate the background and small infrared targets by finding the non-local autocorrelation of the background in the infrared image and the sparse

expression of the target and removing most of the irrelevant background, such as the Infrared Patch-Image (IPI) model [4].

In recent years, target detection methods based on deep learning have achieved good results. These target detection methods based on convolutional neural networks can effectively extract and fuse the features of the target in the image, and can train the network to learn the deep semantic features of the target. For example, RCNN [5], Fast RCNN [6], SSD [7] and other target detection methods are representative; however, these methods are all aimed at general target detection and are not suitable for small-target detection.

YOLO V3 [8], RFBNet [9], RefineNet [10] and other detection methods also have detection ability for small targets, and the detection effect of small targets is improved. RFBNet is based on the receptive field in the human visual system for target detection, which can effectively detect salient targets in pictures. RefineNet is a singleshot target detection network based on CNN and has better detection speed and detection accuracy. Although the detection effect of small targets has been improved in the above methods, the detection effect for small targets is still poor. Therefore, improving the detection accuracy of small targets has become a research hotspot in the field of computer vision.

For infrared small-target detection, in recent years, deep-learning methods that are more suitable for small targets have been studied. One type of methods is to treat small infrared targets as image noise and then transform the target detection process into an image denoising process. These methods use a neural network to denoise the image, and the denoised image is the background image. Then, the original image is subtracted by the background image to obtain the small infrared targets [11–13]. The second type of methods is to extract suitable infrared small target features for training. Since the infrared small target has no semantic information of the shape and category, the features of the small target can be extracted from other angles, such as making use of GAN networks [14,15].

In this paper, we propose an improved FCOS infrared small-target detection method, which is based on the FCOS network [16]. The main contributions of the paper are:

1. As the infrared small-target detection does not require the high-level semantics of the network, we lighten the FCOS detection network and select only the low-level infrared small target features.
2. In order to further improve the response of small targets, the traditional maximum filtering method is added, which helps to improve the detection performance of small infrared targets.
3. For the noise points in the moving infrared image, we use the image sequence as the input of the network to add time domain features, establish the connection between the images, and further eliminate the static noise points in the image.
4. We conducted a comparative experiment on the published infrared small target dataset to prove the detection performance of the method in this paper. The results show that, compared with other infrared small target detection methods, the improved FCOS method proposed in this paper has better performance for small-target detection and has better real-time performance.

2. Related Work

2.1. Small-Target Detection Methods Basd on Deep Learning

At present, a variety of deep-learning methods have been proposed to solve the problem of small-target detection. These include:

1. The methods based on multiscale feature learning, whose main idea is to learn different scale objects separately, which mainly solves the problem of the small target itself with few discriminative features. This type of method is represented by FPN, and its main idea is to integrate the underlying spatial information and high-level semantic information to enhance the target characteristics [17].
2. The methods based on the receptive field, whose representative is the Trident Network. The idea is that small targets require smaller receptive fields [18], and large targets need larger receptive fields, and then dilated convolutions with different dilation rates

are used to form different feelings that are responsible for detecting three branches of different scale objects.

3. The GAN-based methods use GAN to generate high-resolution images or high-resolution features. For example, in [19], the trained detector is used to obtain the subimage containing the target, and then the generator is used to generate the corresponding high-definition image, discriminator is responsible for judging whether the generated image is real or fake, and at the same time acts as a detector to predict the category and location of the target.
4. The context-based methods use the relationship between the environment information of the small target and other easy-to-detect targets to assist the detection of small targets, such as Relation Network [20] implicitly modeling the relationship between two targets through Transformer, and use this relationship to strengthen the characteristics of each object.
5. The methods based on the dataset itself. For example, in Stitcher [21], the proportion of the loss of the small target in the total loss is used as the feedback signal. When the proportion is less than a certain threshold, the four pictures are combined into one picture as the input for the next iteration, which is equivalent to increasing the number of small targets. Augmentation [22] solves this problem directly by simply copying and pasting.

In addition, for small-target detection, one can also design special training strategies to improve the performance of detecting small target. SNIP [23] and SNIPER [24] are improved for image pyramid training methods, by detecting large targets on small scale images, detect small targets on large scale images to ensure that the target size of the input classifier is consistent with the pretrained scale on ImageNet, thereby, improving the performance of small-target detection.

2.2. Infrared Small-Target Detection Methods

For the methods based on image denoising, a denoising self-encoding network [11] is used to denoise infrared images. In [11], an end-to-end infrared small-target detection model based on denoising self-encoding network was proposed. The model includes encoding and decoding. The network can finally obtain the background image without the target, and then the original image is subtracted by the background image to obtain the position of the small target.

In TBCNet [12], the target extraction module (TEM) and semantic constraint module (SCM) are proposed, which are used to extract small targets and classify target extracted during training. ISTNet [13] is an end-to-end CNN-based infrared small-target detection method. The image filtering module (IFM) module effectively enhances the response of infrared small targets and suppresses the background response, and also incorporates adaptive receptive field fusion module and spatial attention mechanism.

For the methods based on the extraction of suitable infrared small target features for training, since the infrared small target has no semantic information of the shape and category, the features of the small target can be extracted from other angles. In [14], the adversarial generative network is used to balance the missed detection and false alarm. The network has two generations.

One is responsible for reducing the missed detection rate and the other is responsible for reducing the false alarm rate. The average of the results generated by the two generators is used as the final segmentation result. Ref. [15] proposed using a GAN network to determine the location of small infrared targets, where the generation network generates small infrared targets, and the classification network discriminates small infrared targets.

2.3. FCOS Target Detection Algorithm and Its Limitation for Infrared Small-Target Detection

In this paper, we design an improved small-target detection network based on the FCOS algorithm. The FCOS algorithm does not require a preset anchor frame and directly performs regression prediction on each pixel of the target. As there is no calculation of the

anchor frame, the training speed can be greatly improved, and the setting of hyperparameters can be reduced. Therefore, the FCOS algorithm training is more stable, and the training efficiency is improved.

FCOS is a pixel-by-pixel approach for target detection tasks, and small targets have relatively few pixels. The downsampled of the neural network is very easy to cause information loss, which in turn affects the accuracy of the model. For target detection, the single-stage detection algorithm is faster than the two-stage detection algorithm. Considering the real-time requirement of the infrared small-target detection algorithm, it is reasonable to select a single-stage target detection algorithm.

The core idea of FCOS is to predict the target category and target frame to which each point in the input image belongs. After canceling the preset anchor, FCOS completely avoids the complex calculations related to the anchor, such as the calculation of the IOU during training, and more importantly, it avoids all anchor-related hyperparameters, which are usually very sensitive to the final detection performance.

In addition, because the number of preset anchors is large, the negative samples will be far more than the positive samples, which will lead to the problem of imbalance between the positive and negative samples. Therefore, based on the anchor-free FCOS algorithm, the training result is more stable, and less affected by the hyperparameters, which greatly improves the training efficiency.

The FCOS network structure is shown in Figure 1. It shows that the FCOS network structure is based on the ResNet backbone network and uses the FPN structure to perform feature fusion on the three feature layers C3, C4, and C5 in the backbone network to obtain P3, P4, P5, and P5 feature layer is downsampled to obtain P6 and P7, which can adapt to targets of multiple scales.

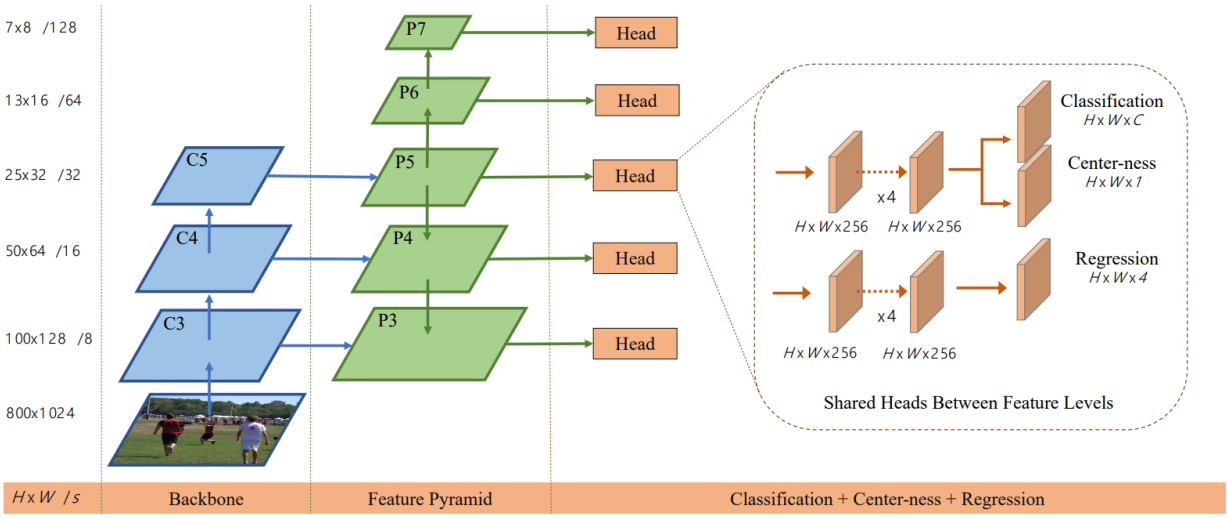


Figure 1. FCOS network structure.

There are three main types in the output layer: the classification branch, Centerness, and the regression branch, which is the main difference between the anchor-free type target detection algorithm and the anchor-based target detection algorithm.The regression part of FCOS predicts four values (l, t, r, b), which, respectively, represent the distance between a certain point in the target box and the left, top, right, and bottom of the box. Assuming that the coordinates and category of a labeled target frame B_i on an image are represented by the following formula, where the first four values are the coordinates of the upper left and lower right corners of the frame. The last value c is the category label between 1 and C (category number):

$$B_i = (x_0^{(i)}, y_0^{(i)}, x_1^{(i)}, y_1^{(i)}, c^{(i)}) \quad (1)$$

Then, the category label of each point on the input image can be determined according to whether the point is in the label box. The point outside the label box is a negative sample, and the category is set to 0; the point (x, y) in the label box is positive samples, the category target is the category label of the label box, and the regression target is the following four values:

$$\begin{aligned} l^* &= x - x_0^{(i)}, t^* = y - y_0^{(i)}, \\ r^* &= x_1^{(i)} - x, b^* = y_i^{(i)} - y. \end{aligned} \tag{2}$$

Centerness is used in FCOS to suppress the generation of low-quality bboxes.

The FCOS algorithm retains the anchorless mechanism, and introduces three strategies of pixel-by-pixel regression prediction, multiscale features, and Centerness. On the basis of the fast single-stage target detection algorithm, the detection accuracy is greatly improved. FCOS avoids the complex calculation of the anchor, uses more foreground samples for training, and the bbox's position regression is more accurate. However, for infrared weak and small targets, the detection of FCOS algorithm still has the following problems:

- As the small infrared target occupies very few pixels compared to the whole image, and the architecture of the FCOS algorithm is ResNet+FPN, after multiple down-sampling, the feature area occupied by the small target is extremely small, and it is even impossible to perceive learning. Therefore, it is more appropriate to use some high-resolution feature maps for learning, which can strengthen the neural network's perception of small target areas.
- Balance of positive and negative samples. In the infrared image, the small infrared target occupies fewer feature points, so all the pixels contained in the bbox of the target should be used as positive sample points for training to increase the number of positive samples.
- Noise points in the infrared image have a greater impact on the detection of small infrared targets, but the small infrared targets can be separated according to the characteristics of the small infrared targets in the image sequence. Therefore, it is possible to add the temporal characteristics of the image in the neural network.

3. Infrared Small-Target Detection Based on Improved FCOS

Infrared small target images have the characteristics of a small number of target pixels, blurred edges, and complex and diverse backgrounds. Therefore, for some classic target detection methods based on deep learning, it is difficult to learn the characteristics of small infrared targets, which leads to problems, such as low target detection efficiency and high false alarm rate.

In order to improve the performance of infrared small-target detection, this paper proposes an infrared small target detection method based on improved FCOS. This method uses the traditional maximum filter image preprocessing and also combines the spatiotemporal information of the image to further improve the infrared small target. The detection accuracy is high, and there is a good real-time performance. Figure 2 shows the flow of the entire detection method.

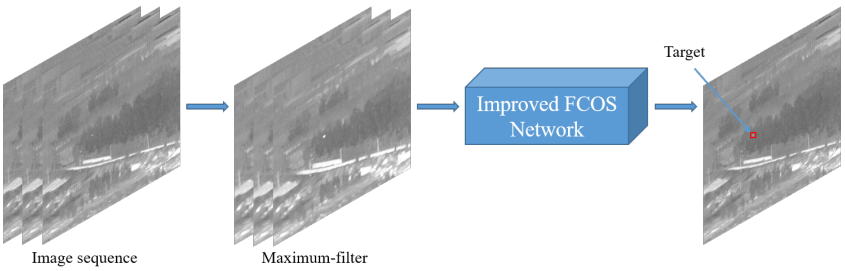


Figure 2. Improved FCOS detection method flow chart.

3.1. Image Preprocessing

Infrared small targets often appear as bright spots in the image and have high pixel values in a certain area, and thus they can be regarded as the local maximum of an image. Among the traditional image filtering methods, maximum filtering is undoubtedly the most suitable. The maximum filter can highlight the weak and small targets in the image and can enhance the characteristics of the infrared small targets. First, it is necessary to xsort the pixels in the window, then compare the center pixel and the largest pixel, take the largest pixel as the value of the center pixel, and, if the center pixel is smaller than the smallest pixel, take the smallest pixel as the center pixel.The maximum filter of a 3×3 matrix is shown in Figure 3.

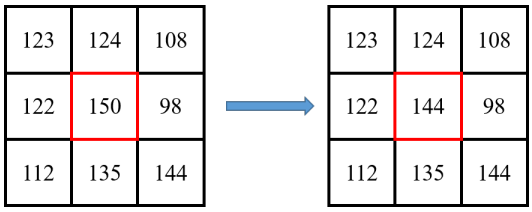


Figure 3. The maximum filtering method.

3.2. Improved FCOS Network

FCOS is a pixel-level target detection algorithm. For small targets with a small number of pixels, the down-sampling of the neural network will cause the loss of target information, resulting in the network cannot learn the effective features of the targets, which in turn affects the detection accuracy of the model. The feature information of infrared small targets is relatively simple, and the feature information is generally retained in the under-lying feature layer, so the size of the neural network can be reduced.

On the basis of the original FCOS network, C4, C5, P4, P5, P6, and P7 feature layers are removed, because too many down-sampling layers will affect the efficiency of detection. Then, only two feature layers C2 and C3 are used for feature extraction. The number of stacked channels in the two branches of classification and regression are reduced to one to reduce the computational load of the network model. The improved network structure is shown in Figure 4.

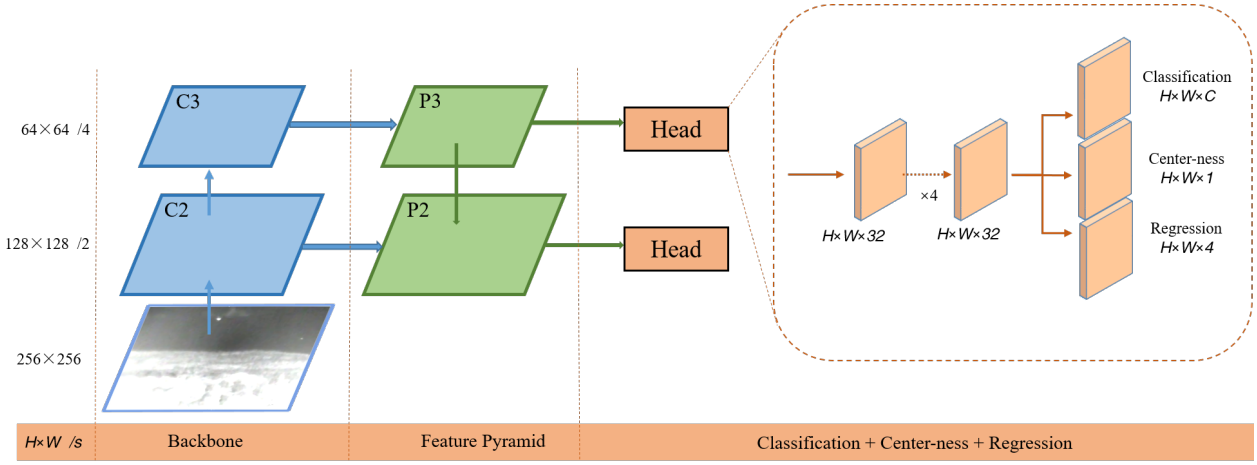


Figure 4. Improved FCOS network structure.

As can be seen from the figure, only two feature layers, C2 and C3, are input to the feature fusion layer. The step lengths of these two layers are 2 and 4, respectively, and they are in a double-sampling relationship. Then, in order to combine the high-level and low-level feature expressions, P3 is upsampled and merged with C3 to obtain feature layer

P2. Finally, category prediction and coordinate regression are performed on the two feature layers of P2 and P3.

Therefore, using only the features of two feature layers greatly reduces the size of the network and the training time. As the image size of the infrared small target dataset is 256 × 256, and the pixel size of the infrared small target is generally about 20 pixels. In the original FPN structure, the step size of downsampling is large, which will cause the loss of small target feature information. Therefore, the C4, C5, P4, P5, P6, and P7 feature layers are removed to increase the high-resolution feature expression, and improves the accuracy and recall rate of small targets. This paper also improves the four-layer convolution layer before the classification and regression branches so that the classification branches and regression branches share the network parameters, reducing the number of channels in the convolution layer, and further reducing the amount of network calculation and improving the detection speed.

3.3. Adding Spatio-Temporal Features

When the improved FCOS method is used to detect the infrared dataset, the bright spot-like noise points similar to the small infrared target will affect the accuracy of the small infrared target. The noise points in the infrared image are generally static relative to the background, and the infrared target is moving relative to the background, so the feature of the motion vector is added in the neural network, that is, the image sequence is inputted into network instead of a single image to increase the time domain information, as shown in Figure 5.

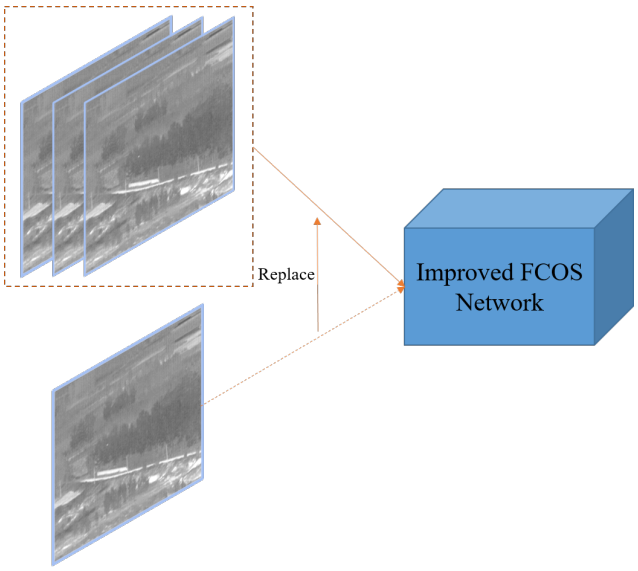


Figure 5. Comparison of changes before and after adding time domain features.

In the process of model training, the dimension of input layer in FCOS is changed, multiple time dimension channels are superimposed on the single-channel infrared grayscale image to replace the two-dimensional spatial RGB image in the original network. Experiments have found that adding time-domain features can effectively reduce the misjudgment of noise points and improve the accuracy of infrared small-target detection.

4. Experiment and Analysis

4.1. Experimental Dataset

- Published dataset of infrared image sequences

The published dataset of infrared image sequences is a small and weak aircraft target detection data set for low-altitude flight published by the National University of Defense Technology [25]. The dataset acquisition scene covers the sky, ground, and other back-

grounds as well as a variety of complex scenes. We selected two scene image sequences of pure sky and complex ground, namely sequence1 and sequence2 as the dataset for our research. Figure 6 shows some of the images in these two scene sequences.

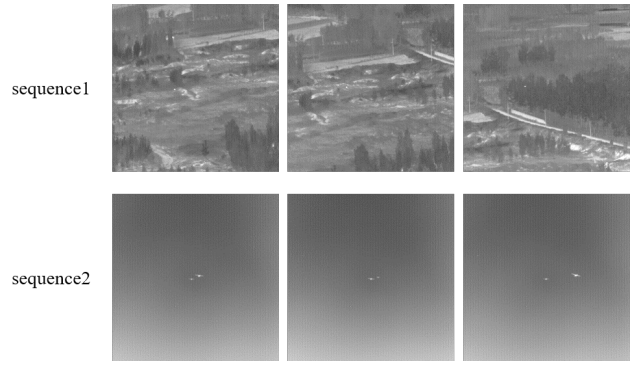
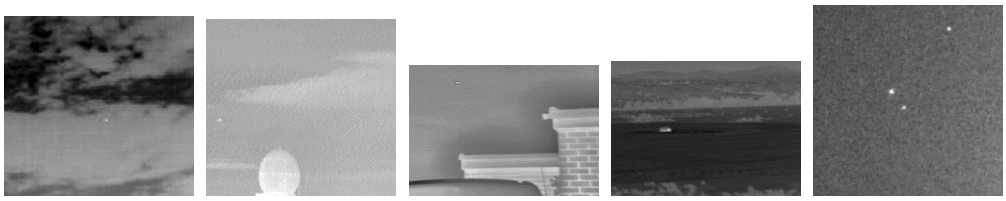


Figure 6. Examples of the published infrared image sequence dataset.

- Published single-frame infrared small target image dataset

This single-frame infrared small target dataset was proposed in an article by China Southern Airlines Dai Yimian [26]. The dataset extracts the most representative pictures from hundreds of image sequences. In order to avoid the overlap between the training set, the verification set and the test set, only one representative image is selected in each infrared sequence.

Each target is confirmed by its movement sequence to ensure that it is a real target and not pixel-level noise. About 90% of the images in the data set have only one target, about 10% of the images have multiple targets, and about 55% of the targets account for less than 0.02% (that is, in a 300 × 300 image, the target pixel is 3 × 3). A total of 65% of the targets are as bright as the background or even darker than the background. This dataset is only used in the improved FCOS network without adding time domain features. Figure 7 shows some of the images in the dataset.



(**a**) 0003# (**b**) 0031# (**c**) 0134# (**d**) 0212# (**e**) 0387#

Figure 7. Examples of a single-frame infrared small target image dataset.

4.2. Evaluation Indicators and Experimental Settings

The precision rate and recall rate of commonly used detection indicators in target detection are used in this paper. The calculation formula is as follows:

$$\begin{aligned} TP = Truepositive, TN = Truenegative \\ FP = Falsepositive, FN = Falsenegative \\ Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \end{aligned} \tag{3}$$

The number of processed frames per second (FPS) was used as the evaluation index of algorithm detection speed. The deep-learning model used in this article was implemented using the Pytorch framework. All experiments were performed on Intel(R)Core(TM)i7-7550K CPU@2.8 GHz, GPU graphics card is NVIDIA GTX2080(Pascal)/Pcle/SS, operating system is Ubuntu On 18.04 LTS server.

4.3. Comparison of Preprocessing Methods

In order to further improve the effect of the deep-learning algorithm in detecting small infrared targets, this paper combines traditional filtering methods to highlight the characteristics of small infrared targets. We selected the method of maximum value filtering and preprocessed the image before inputting to the network. As small infrared targets are often local maximums in the image, maximum value filtering can enhance the feature expression of small infrared targets.

We compare this with other filtering methods, such as Top-hat, mean filtering, contrast enhancement filtering, gradient transformation filtering methods and multiscale patch-based contrast measure (MPCM). The experimental results show that the maximum filtering method has the best detection effect for the FCOS algorithm. The comparison results of different filtering methods are shown in Figure 8, and the detection results are shown in Table 1.

Table 1. Comparison results of accuracy and recall of different filtering methods for a single frame infrared dataset and sequence1.

	Sequence1		**A Single-Frame Infrared Dataset**	
	Precision	**Recall**	**Precision**	**Recall**
Top-hat	0.610	0.818	0.775	0.964
Mean-filter	0.151	0.179	0.90	0.96
Enhance-contrast	0.947	0.955	0.958	0.990
Gradient	0.973	0.992	0.946	0.982
MPCM	0.979	0.992	0.880	0.951
Maximum	0.984	0.992	0.973	0.990

4.4. Image Sequence Length Comparison

In the method proposed in this paper, the length of the input sequence of the network model has an important influence on the detection effect of small infrared targets. Theoretically, when the time domain features are missing, the detection effect of the network will be worse, and the noise points will be misjudged as small targets, resulting in a decrease in accuracy.

As when a single image is input, it lacks the inter-frame information of the image sequence, and the network cannot fully learn the information of the temporal context. However, when the sequence length value is too large, the calculation amount of the neural network will also increase. On the other hand, it will also cause the network to overlearn the time domain features, and then decrease the algorithm's ability to recognize the infrared small target.

Table 2 shows the experimental results of using image sequences of different lengths as the network input. Experimental results show that when the image sequence is used as the network input and the sequence length is 3, the accuracy and recall rate are significantly improved, but when the length gradually increases to 5 and 7, the accuracy and recall rate start to decrease. This is because the image sequence span is too large, and the noise points have false association, which leads to the misjudgment of the algorithm.

Table 2. Comparison results of accuracy and recall of image sequences of different lengths for sequence1 and sequence2.

Length	**Sequence1**					**Sequence2**				
	1	**2**	**3**	**5**	**7**	**1**	**2**	**3**	**5**	**7**
precision	0.982	0.980	0.984	0.973	0.960	0.989	0.989	0.989	0.988	0.979
recall	0.990	0.992	0.992	0.982	0.977	0.990	0.993	0.996	0.992	0.989
FPS	32.9	31.2	36.2	28.9	22.7	58.6	43.6	41.2	24.1	15.0

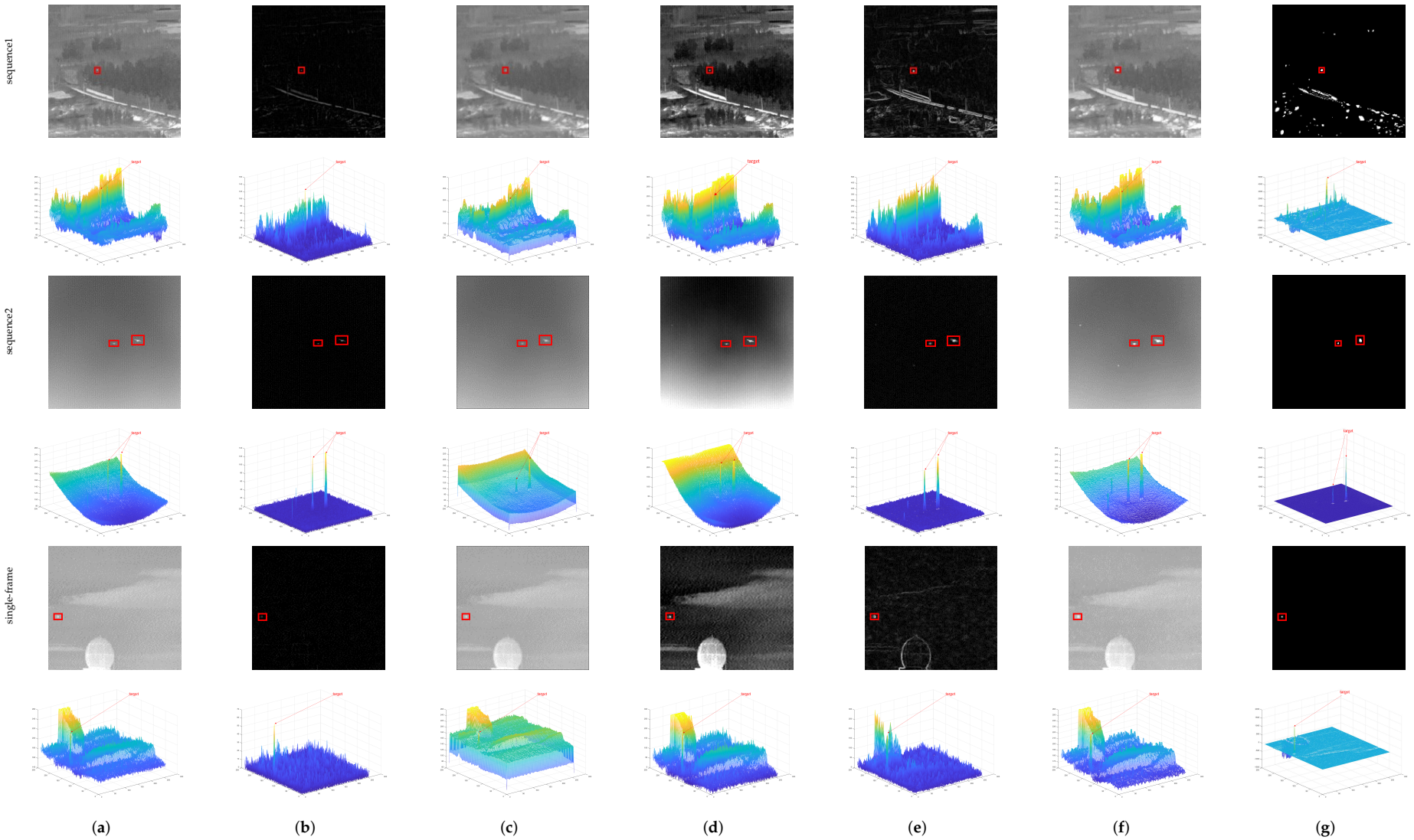


Figure 8. Results and 3D graphs of different filtering methods for multiple datasets. (**a**) Original; (**b**) Top-hat; (**c**) Mean-filter; (**d**) Contrast enhancement; (**e**) Gradient; (**f**) Maximum; (**g**) MPCM.

4.5. Compartive Experiments of Different Measures

In order to show intuitively that this method has better detection effect, and also to reflect that the fusion of deep learning and traditional methods can achieve more stable infrared small-target detection, we make use of only improved FCOS network for experiments. the comparative experiments among improved FCOS, improved FCOS + time domain features, and improved FCOS + time domain features + traditional Filtering was conducted. Table 3 shows the experimental results for the sequence dataset and single-frame dataset.

Table 3. Comparative experiment results between the improved FCOS and comprehensive method for sequence1 and the single-frame infrared dataset.

	Sequence1			**A Single-Frame Infrared Dataset**		
	Precision	**Recall**	**FPS**	**Precision**	**Recall**	**FPS**
Improved FCOS	0.967	0.987	30.5	0.966	0.986	25.5
Improved FCOS + traditional Filtering	0.982	0.990	32.9	0.972	0.992	29.0
Improved FCOS + time domain features	0.978	0.990	29.9	-	-	-
Improved FCOS + time domain features + traditional Filtering	0.984	0.992	36.2	-	-	-

The table shows that the FCOS network that only integrates the two feature layers of C2 and C3 has faster detection speed and good detection accuracy. However, due to the lack of temporal context information of the moving target, the static noise points in the image are regarded as the detection target, and thus the recall rate of this method is low. Although it could ensure a higher recall rate, it also caused missed detection. FCOS integrated with traditional methods had greatly improved accuracy and recall rate and can meet the conditions of real-time detection. Therefore, the improved FCOS method proposed in this paper had a better effect on infrared small-target detection, which greatly improved the accuracy of target detection.

4.6. Comparison of the Method in This Paper and Other Methods

We compared the proposed method with other infrared small object detection methods, including YOLO v3, RFBNet, RefineNet, ALCNet [27], Density Peaks Searching [28], and FCOS. The image sequence with a complex background was used for testing, and the results are shown in Table 4. In addition to comparing the detection accuracy of different methods, in order to prove the real-time performance of the method in this paper, the experiment also compared the detection speed of different methods in the GPU and CPU environments.

The experimental results show that the improved FCOS method proposed in this paper had the highest accuracy and had good real-time performance. The experimental results also show that merging classification and regression branches can not only improve the detection speed but also slightly improve the detection accuracy.

Table 4. Comparison of detection results between comprehensive methods and other deep-learning methods.

Detector	**YOLO v3**	**RFBNet**	**RefineNet**	**ALCNet**	**Density Peaks Searching**	**FCOS**	**Proposed**	**Proposed ***
precision	80.33	83.74	85.17	88.34	56.31	0	98.40	98.30
FPS (GPU)	35.4	34.78	27.16	7.69	-	22.72	36.2	35.5
FPS (CPU)	1.39	17.6	5.4	2.64	12.56	0.5	19.0	16.0

* Classification and regression branches are not merged.

5. Conclusions

In this paper, a new infrared small-target detection method is proposed. First, based on the FCOS algorithm, an improved FCOS network for small infrared targets is proposed, which removes the network structure that is not conducive for the detection of small infrared targets. The detection is performed on the low-level feature layer of the neural network, which improves the training speed of the network. Secondly, in order to further highlight the small infrared target, the deep-learning method is combined with the traditional method, and the maximum filtering method is used to filter the infrared image, which can enhance the response of the small infrared target.

In addition, spatio-temporal features are added, and training is conducted in the form of image sequences, which enables the network to learn the relationships between target motion in images, which can effectively reduce the misjudgment caused by noise in infrared images and improve the recall rate of detection methods.

The experimental results on the published infrared dataset show that, compared with other infrared small object detection methods, this method had the best performance. The detection accuracy reached 98.4%, and the detection speed in GPU and CPU environment reached 36.2 and 19.0 FPS, respectively. Future research includes: designing better network structures and classification/regression loss functions that are more suitable for infrared small targets as well as more useful integration of traditional methods and deep-learning methods.

Author Contributions: Methodology, Q.Z. and P.Y.; software, S.Y.; data curation, T.Z. and W.C.; writing—original draft preparation, S.Y.; supervision, Q.Z.; funding acquisition, T.Z and W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the open project fund of the Key Laboratory of intelligent infrared perception, Chinese Academy of Sciences (grant number: CAS-IIRP-2030-03), and the APC was funded by CAS-IIRP-2030-03.

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

References

1. Tom, V.T.; Peli, T.; Leung, M.; Bondaryk, J.E. Morphology-based algorithm for point target detection in infrared backgrounds. In Proceedings of the Signal and Data Processing of Small Targets 1993, Orlando, FL, USA, 11–16 April 1993; Drummond, O.E., Ed.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 1993; Volume 1954, pp. 2–11. [CrossRef]
2. Deshpande, S.D.; Er, M.H.; Venkateswarlu, R.; Chan, P. Max-mean and max-median filters for detection of small targets. In Proceedings of the Signal and Data Processing of Small Targets 1999, Denver, CO, USA, 18–23 July 1999; Drummond, O.E., Ed.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 1999; Volume 3809, pp. 74–83. [CrossRef]
3. Bae, T.W.; Zhang, F.; Kweon, I.S. Edge directional 2D LMS filter for infrared small-target detection. *Infrared Phys. Technol.* **2012**, *55*, 137–145. [CrossRef]
4. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared Patch-Image Model for Small-Target Detection in a Single Image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [CrossRef] [PubMed]
5. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.
6. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
7. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
8. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
9. Liu, S.; Huang, D.; Wang, A. Receptive Field Block Net for Accurate and Fast Object Detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
10. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-Shot Refinement Neural Network for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
11. Shi, M.; Wang, H. Infrared Dim and Small-Target Detection Based on Denoising Autoencoder Network. *Mob. Netw. Appl.* **2020**, *25*, 1469–1483. [CrossRef]

12. Zhao, M.; Cheng, L.; Yang, X.; Feng, P.; Liu, L.; Wu, N. TBC-Net: A real-time detector for infrared small-target detection using semantic constraint. *arXiv* **2020**, arXiv:2001.05852.
13. Ju, M.; Luo, J.; Liu, G.; Luo, H. ISTDet: An efficient end-to-end neural network for infrared small target detection. *Infrared Phys. Technol.* **2021**, *114*, 103659. [CrossRef]
14. Wang, H.; Zhou, L.; Wang, L. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8509–8518.
15. Zhao, B.; Wang, C.; Fu, Q.; Han, Z. A novel pattern for infrared small-target detection with generative adversarial network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4481–4492. [CrossRef]
16. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9627–9636.
17. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
18. Li, Y.; Chen, Y.; Wang, N.; Zhang, Z. Scale-aware trident networks for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6054–6063.
19. Bai, Y.; Zhang, Y.; Ding, M.; Ghanem, B. Sod-mtgan: Small object detection via multi-task generative adversarial network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 206–221.
20. Hu, H.; Gu, J.; Zhang, Z.; Dai, J.; Wei, Y. Relation networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3588–3597.
21. Chen, Y.; Zhang, P.; Li, Z.; Li, Y.; Zhang, X.; Meng, G.; Xiang, S.; Sun, J.; Jia, J. Stitcher: Feedback-driven data provider for object detection. *arXiv* **2020**, arXiv:2004.12432.
22. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. *arXiv* **2019**, arXiv:1902.07296.
23. Singh, B.; Davis, L.S. An analysis of scale invariance in object detection snip. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3578–3587.
24. Singh, B.; Najibi, M.; Davis, L.S. Sniper: Efficient multi-scale training. *arXiv* **2018**, arXiv:1805.09300.
25. Hui, B.; Song, Z.; Fan, H.; Zhong, P.; Hu, W.; Zhang, X.; Ling, J.; Su, H.; Jin, W.; Zhang, Y.; et al. *A Dataset for Infrared Detection and Tracking of Dim-Small Aircraft Target under Ground/Air Background*; Science Data Bank: Beijing, China, 2019. (In Chinese) [CrossRef]
26. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Asymmetric contextual modulation for infrared small-target detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 950–959.
27. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Attentional Local Contrast Networks for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9813–9824. [CrossRef]
28. Huang, S.; Peng, Z.; Wang, Z.; Wang, X.; Li, M. Infrared Small-Target Detection by Density Peaks Searching and Maximum-Gray Region Growing. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1919–1923. [CrossRef]