

Article

Not peer-reviewed version

Infrared Dim and Small Target Detection Based on Background Prediction

[Jiankang Ma](#) , Haoran Guo , [Shenghui Rong](#) ^{*} , Junjie Feng , [Bo He](#)

Posted Date: 16 May 2023

doi: 10.20944/preprints202305.1075.v1

Keywords: Dim and small target detection; background prediction; image inpainting; Region Proposal Network (RPN)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: 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.

Article

Infrared Dim and Small Target Detection Based on Background Prediction

Jiankang Ma ¹, Haoran Guo ¹, Shenghui Rong ^{1,*}, Junjie Feng ² and Bo He ¹

¹ Underwater Vehicle Laboratory, School of Information Science and Engineering, Ocean University of China, Qingdao 266000, China; majiankang@stu.ouc.edu.cn (J.M.); guohaoran@stu.ouc.edu.cn (H.G.); bhe@ouc.edu.cn (B.H.)

² SINOPEC Research Institute of Safety Engineering Co., Ltd., State Key Laboratory of Safety and Control for Chemicals, 339 Songling Road, Qingdao 266100, China; feng-jj@hotmail.com (J.F.)

* Correspondence: rsh@ouc.edu.cn

Abstract: Infrared dim and small target detection is a key technology of various detection tasks. However, due to the lack of shape, texture, and other information, it is a challenging task to detect dim and small targets. Recently, since many traditional algorithms ignore global information of infrared images, they generate some false alarms in complicated environments. To address this problem, in this paper, a coarse-to-fine deep learning-based method is proposed to detect dim and small targets. Firstly, a coarse-to-fine detection framework integrating deep learning and background prediction is applied for detecting targets. The framework contains of coarse detection module and fine detection module. In coarse detection stage, Region Proposal Network (RPN) is employed to generate masks in target candidate regions. Then, for further optimize the result, inpainting is utilized to predict the background by the global semantic of images. In this paper, inpainting algorithm with a mask aware dynamic filtering module is incorporated into the fine detection stage to estimate background in candidate target areas. Finally, compared with existing algorithms, the experimental results indicate that the proposed framework have effective detection capability and robustness for complex surroundings.

Keywords: infrared dim and small target detection; background prediction; image inpainting; Region Proposal Network (RPN)

1. Introduction

Infrared dim and small target detection technology has an essential role in target detection field [1]. Due to long-distance and immature infrared imaging technology, the target is usually tiny (less than 80 pixels [2]) with a low signal-to-clutter ratio (SCR) and lacks shape and texture features. In addition, the background and noise in the complex environment will easily overwhelm targets. Therefore, infrared dim small target detection is a difficult and challenging task.

In order to extract targets from complex and variable backgrounds, many techniques and knowledge are applied to infrared dim small target detection algorithms. In the early stage, some methods based on background prediction [3–5] detect small targets by the difference between the infrared image and the predicted background obtained by local information of the image. Since background predictions obtained in the local domain are not accurate, so they have lower performance on complex backgrounds. For improving the quality of the target, some methods focus on features of targets to suppress background and enhance target [6–10]. Due to the salience of local features (local contrast, local gradient, etc.), these approaches effectively improve SCR of the target and suppress clutters. However, targets are not detected facing the image with buildings and highlight. Sparse matrix methods [11–13] for small targets detection concern non-local correlation to separate small targets. They transform the problem of target detection into that of sparse matrix separation, but owing to the limitations of the sparse model, many false alarms are retained. Recently, deep-learning-based methods [14–16] segment the target and background based on image features extracted by the deep

neural network. The deep network extracts deep and low features of the target, so the detection accuracy is significantly improved. Totally, the above methods have different disadvantages in the process of detecting small targets. Nevertheless, many false alarms still exist in the results of many algorithms on account of the lack of global or local information.

In this paper, a background prediction-based method with coarse and fine detection is researched. Traditionally, background prediction-based methods for infrared dim small targets detection estimate background by the local correlation of pixels through filters and transformation. Though these methods have better performance on smooth and simple background, due to ignoring the global information of the infrared image, they cannot detect the target when facing with complex scenes like buildings and highlight noises. In contrast, deep learning can adapt to varying scenarios by learning a large member of training samples. Additionally, deep learning has the capability to accomplish various tasks by understanding the global semantic information of images [17] such as small targets detection [18], image inpainting [19] and face recognition [20]. Inspired by deep learning, a background prediction method utilizing image inpainting is proposed to detect dim and small targets in complex scenarios.

In this paper, the global information of the image is utilized to estimate the local scene in an IR image. Here, two prior assumptions of small targets are proposed to appropriately illustrate the background predictions of the paper. The assumptions are demonstrated as follows.

- (1) Because of the size of the target, the remaining image after eliminating the target have negligible impact on the image background semantics. Meanwhile, the semantics are able to predict the background in small targets areas.
- (2) The background can be predicted by the semantics of the remaining in suspicious target areas (building edges, highlighting noise, etc.), conversely, the predicted background in the suspected target areas are approximately the same as the original image.

From the above assumptions, the coarse and fine detection framework requires a technique to predict the background of these suspicious target areas by understanding the image. Image inpainting has exactly an ability of background prediction in marked areas. It is a classical computer vision problem. Recently, inpainting recovers the whole image via analyzing deep and shallow semantic features of the image by the neural network, which adds some reasonable inferences in the missing areas of the image [21]. A number of inpainting algorithms have the capability of handling large contiguous areas, which ignore the inner connection between isolated areas and the whole. In this paper, an inpainting method with a mask-aware dynamic filtering (MADF) [22] module is adapted to estimate background of candidate target areas. Owing to the small size of the target, the candidate target areas are correspondingly tiny. Therefore, the two prior assumptions claim that inpainting can adapt to repair tiny areas. The method with MADF addresses the problem of repairing small and isolated areas in addition to repairing large and continuous areas.

In this paper, the detection framework is a structure containing coarse and fine detection. Deep learning is utilized to generate the candidate target areas in the coarse detection module, where the existing algorithm is utilized to detect. In the paper, this module is not focused on the study and elaboration. The fine detection includes inpainting that predicts background of candidate target areas and fusing images to detect small targets. Comparative experiments on test data demonstrate that the proposed framework is capable of detecting infrared small targets. Overall, the technical contributions of the paper are summarized in the following.

- A coarse-to-fine infrared dim and small target detection framework is proposed to adapt to complex infrared image scenes. In coarse and fine detection modules, deep learning is utilized to detect candidate target areas and fine targets.
- Image inpainting method with MADF is first employed to predict the background by global semantic information in the stage of fine detection.
- Extensive experiments on public infrared small target datasets illustrate that the proposed framework has achieved better performance compared with existing methods.

The remaining sections of this paper are organized as follows. In Section 2, a brief review of related works for infrared dim and small target detection is presented. Section 3 introduces the architecture of the proposed framework in detail. Section 4 describes the experimental results and other terms relevant to the experiment. In Section 5, The discussion about the results obtained qualitatively and quantitatively is demonstrated. Conclusions are drawn in Section 6.

2. Related Works

As an important technique for infrared detection field, infrared dim small target recognition has been widely developed and applied in various detection tasks in recent years. Current methods for infrared dim small target detection are broadly classified into conventional methods and deep learning-based methods.

Traditionally, infrared dim and small target detection methods are generally divided into background prediction-based, local feature-based and sparse matrix-based methods. In the initial stage, background prediction-based methods are employed to detect dim and small the infrared target, such as mean filter [23], top-hat transform [24]. Recently, high-order statistic [3], effective background model in the Fourier domain [4] and improved bilateral filtering [5] are adopted into background prediction to detect small targets. Since the background predictions obtained via local correlation are inaccurate, these methods detect many false alarms in complex scenarios. Some researchers utilize a wide variety of features (local contrast, local gradient information, etc.) as prior information to detect targets. Chen et al. [6] proposed a method based on local contrast measure (LCM) to enhance small targets and suppresses the background by local statistics. The multiscale patch-based contrast measure (MPCM) [25] develops the LCM to further increase the contrast between target and background. The weighted strengthened local contrast measure (WSLCM) [26] considers the characteristics of the target and the background and the difference between them. There are many methods originating from LCM like multiscale tri-layer local contrast measure (TLLCM) [9], multidirectional derivative-based weighted contrast measure (MDWCM) [27], strengthened robust local contrast measure (SRLCM) [28], etc. Some researchers focus on other local characteristic. Local intensity and gradient properties (LIG) [29] characterizes two local properties of small targets from the perspective of intensity and gradient to obtain targets. Average absolute gray difference (AAGD) compensated every single weak-spot is applied in IR small target detection [30]. Local features can be applied well to enhance small targets; nevertheless, with the lack of target-background association of local features, buildings and highlighted backgrounds are identified as targets in complex images. Some regard the infrared image as a matrix, transforming the detection problem into that of sparse matrix separation. Infrared patch-image model (IPI) [13] utilizing local patch construction is proposed to detect targets. Total variation weighted low-rank constraint (TVWLR) [11] and non-convex rank approximation minimization joint $l_{2,1}$ norm (NRAM) [31] develop IPI model, proposing different low-rank matrix theories to detect small targets. Sparse matrix model is established by non-local correlation of infrared images. Therefore, it is challenging for these methods to accurately detect small targets in complex backgrounds. In conclusion, background prediction-based methods process the whole image by local information, which lead to identify bright backgrounds as targets. Local feature-based methods rely excessively on prior assumptions of local information, lacking understanding of the global information of the image. Although sparse matrix-based methods detects the target by non-local information, they still have some false alarms because of the disadvantages of sparse matrixes in complicated environment. Overall, due to ignore the global semantics information or local features, above methods generally cannot adapt to the complex and changeable environment.

Considering that deep learning can adapt to different scenarios by learning to train samples, deep learning is rapidly utilized in a variety of fields. Many researchers propose deep learning networks for infrared small target detection. Ref. [32] decomposes the detection task into two sub-tasks to reduce either miss detection (MD) or false alarm (FA). Attentional local contrast networks (ALCNet) [14] fuses attention mechanisms and local contrast into neural network to detect targets.

Dense Nested Attention Network (DNANet) [33] detects small targets by the contextual information enhanced by channel-spatial attention. Attention-Guided Pyramid Context Networks (AGPCNet) [34] enhance small targets and suppress context via combining global attention and local semantic. Interior Attention-Aware Network (IAANet) [15] increases the interior relation between target and background pixels by Transformer and Region Proposal Network (RPN) to detect small targets. Local Patch Network (LPNet) [35] demonstrates an infrared small target detection network fusing local information and global attention. A robust infrared small target detection network (RISTDNet) [18] increases effectiveness and robustness of detection by incorporating handcrafted feature and neural networks. The above methods utilize neural networks to extract target features, automatically learning classified target features to detect dim and small targets. Although they have achieved superior performance, many methods have disregarded some context information and correlation between targets and background. This can lead to the loss of small targets and the exiting of false alarms. Since the down-sampling operation ignores the features of small targets, the deep network results in the failure of detection. Nevertheless, its influence on the information of the background is virtually negligible. Hence, deep learning is able to understand global semantics information. In this paper, for deep network miss the features of small targets, a coarse-to-fine detection framework based on deep learning is researched. Moreover, image inpainting repairs target candidate areas by the global information of the whole image in the fine detection stage.

Inpainting is a very important image processing technology in computer vision field. Meanwhile, inpainting is a common application of deep learning to restore the missing areas via image understanding [36]. Pathak et al. [37] proposed an image inpainting algorithm combining CNN and GAN, which reproduces the missing areas by the information around these areas. Yang et al. [38] produce high-frequency details by the joint optimization of image content and texture constraints, addressing the task of filling large holes on high-resolution images. Liu et al. [39] presented a partial convolution method to mask and renormalize the convolution, which fixes the convolution in the valid regions to address image inpainting of arbitrary mask. DeFLOCNet [40] generates the user-intended editing results via deep features guided by low-level controls (sparse sketch lines, color dots, etc.). Probabilistic Diverse GAN (PD-GAN) [19] modulates deep features of random noise to restore images from coarse to fine. There are other image inpainting methods based on deep learning, such as image inpainting detection network (IID-Net) [41], GAN for pluralistic image inpainting (PiiGAN) [42]. The above algorithms reconstruct the missing regions via understanding the semantic information of the image by neural networks. Moreover, high-resolution RGB images are repaired with or without constraints. In this paper, an inpainting method with mask-aware dynamic filtering (MADF) module [22] is employed to estimate background of infrared images in local areas. Following as two prior assumptions in Section 1, the target candidate areas are small and the background of these areas are predicted by inpainting. Compared with others, the method with MADF is able to generate convolution kernels with any size areas for image inpainting. Therefore, the method is suitable to be employed for background prediction in infrared small target images.

3. Proposed Method

In this section, the structure of the coarse-to-fine infrared detection framework is described in details. Firstly, the architecture of the proposed method for infrared dim and small target detection is introduced, which is composed of coarse detection module and fine detection module. Then, the coarse detection with Region Proposal Network (RPN) is elaborated to obtain target candidate areas. Ultimately, the design of inpainting to estimate background in fine detection stage is demonstrated.

3.1. Model Architecture

Theoretically, infrared dim and small target images are modeled as Equation 1, where an image is regarded as the sum of target and background (noise).

$$I = T + B(N), \quad (1)$$

where I , T and $B(N)$ denote the infrared image, target and the background with noise, respectively. In the paper, the method for predicting the background $B(N)$ is researched. The $B(N)$ is available by the proposed method. Furthermore, the target T is derived from Equation 1. The overall architecture of the proposed method is indicated in Figure 1.

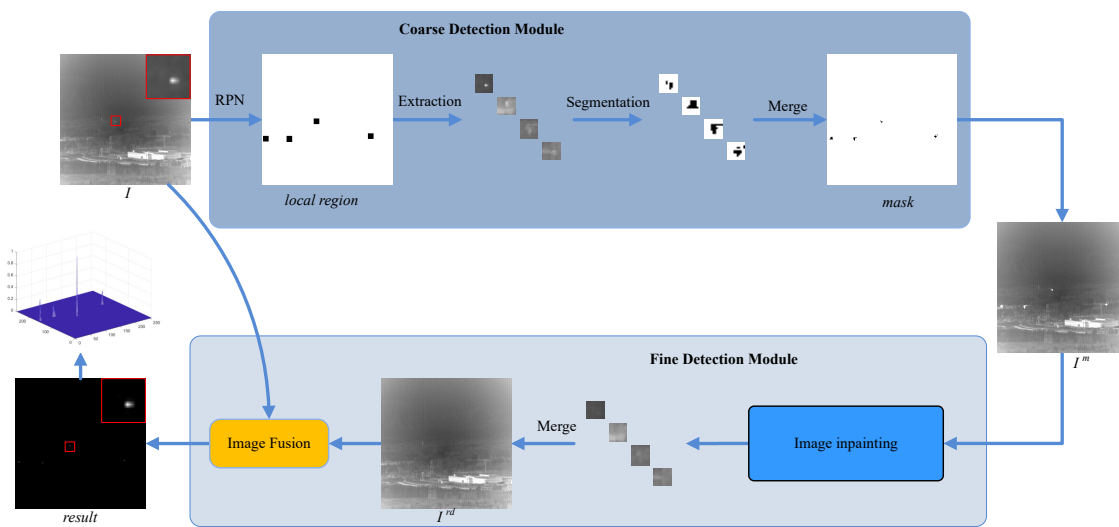


Figure 1. The architecture of the proposed coarse-to-fine framework based on deep learning. I denotes the IR image (the input of our framework). The *result* is detection result for infrared small targets in the framework. The coarse detection module and fine detection module are two parts of the proposed framework. The *mask* is calculated to mask candidate targets by the coarse detection module. I^m denotes the original image marked by masks. I^{rd} is generated by inpainting algorithm by understanding the feature of I^m . In Image Fusion module, the *result* is calculated by the difference between I and repaired image I^{rd} .

In order to accurately predict the background, the process of obtaining $B(N)$ is divided into two parts: predicting the background over a large area by acquiring candidate targets and estimating the background by background inference from the background outside candidate targets. Correspondingly, the detection framework consists of two components: coarse detection module and fine detection module. The coarse detection module is utilized to generate the mask that represents the candidate targets by the infrared image I . The mask is obtained by threshold segmentation on the target candidate areas generated by RPN. In fine detection stage, inpainting is as a background prediction technology to estimate the background via global semantic information in these masks. The infrared image and mask are fed to inpainting to predict the background in these target candidate areas. Since the mask corresponding with candidate targets is tiny, inpainting are able to learn the global information of background outside the mask, which reconstructs the background of the image via this information within the mask. Therefore, it is a key for detecting small targets to repair candidate targets areas. The image with mask I^m is calculated by multiplying the IR image I with mask. I^m and *mask* are both fed to fine detection module to get the repaired image I^{rd} . Furthermore, I^{rd} is considered as predicting background in areas of candidate targets. In Image Fusion module, the result of target detection is obtained by the difference between the IR image I and the repaired image I^{rd} .

3.2. Coarse Detection Module

Conventionally, many methods utilize certain manual features to detect small targets, which are able to obtain small targets with pure background. However, as they lack of distinction and connection between target and background, many algorithms for infrared small and dim target detection often suffer from high false alarm rate and poor accuracy in complex images. In order to eliminate the detection false alarm regions, a coarse-to-fine infrared detection framework is proposed. In the part, the coarse detection module is introduced to roughly detect infrared dim and small targets.

The core detection method of the module is Region Proposal Network (RPN), which classifies small targets and background by deep semantic features via proposal boxes. Region Proposal Network is initially proposed as part of Faster R-CNN in [14]. In this paper, Region Proposal Network is applied to generate target candidate regions that roughly indicate the possible targets. The architecture of the module with RPN is displayed in Figure 2. In order to better adapt to the detection of small infrared targets, RPN optimized in [15] is utilized in our module. The down-sampling operation leads to the removal of small image features during image feature extraction by neural networks. For addressing this problem, the residual network is employed to extract the features of infrared images for RPN. In target candidate regions, the mask is obtained by threshold segmentation. The mask, the marker for candidate targets, is useful for estimating the background of infrared images.

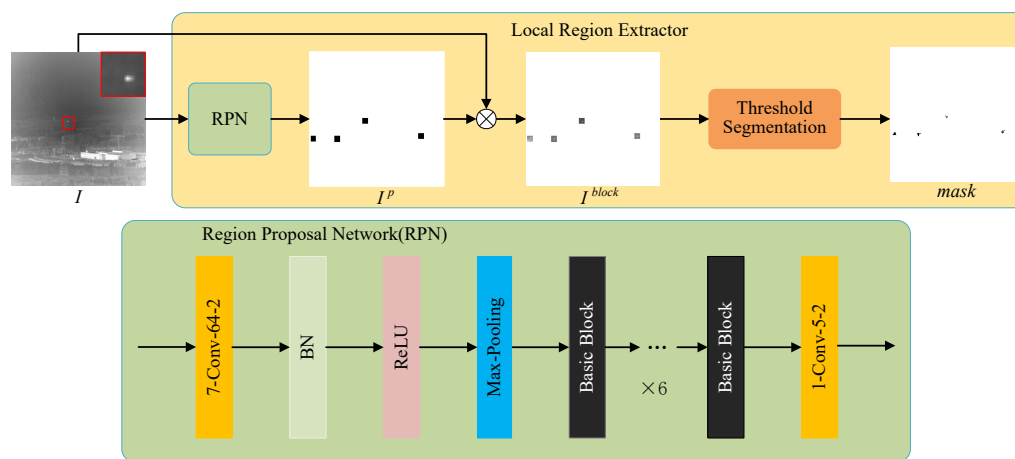


Figure 2. The architecture of coarse detection module with Region Proposal Network (RPN).

As the input of RPN, the infrared image I is processed by RPN to acquire the anchor box embodying target candidate regions I^p . Subsequent processing of coarse detection module is performed in these boxes. The original image I is chunked by candidate regions to extract primitive pixels as candidate blocks I^{block} . The $mask$ is calculated via segmenting the block by simple threshold in every block. The $mask$ value outside these blocks is set to 1. In every block of I^{block} , the $mask$ value matching with the block higher than the threshold is set as 0, otherwise, it is set as 1. After threshold segmentation, these blocks are merged as $mask$. Since the candidate targets represent areas may belong to real targets, the areas outside candidate targets are considered as background without any targets. Thus, in the $mask$, the values of the areas outside candidate targets are set to 1 to indicate the background.

3.3. Fine Detection Module

Traditionally, background prediction methods for infrared dim and target detection estimate the background by filtering and transforming, lacking of the understanding of the global information of the image, which leads detection results contain false alarm interference from the background. Hence, in the paper, aiming to obtain a better background prediction, a deep learning-based inpainting algorithm is adopted to predict background in fine detection stage. Image inpainting is a technique to restore the

missing areas by the semantic features of images. The features are extracted outside the *mask* where the target candidate areas are set 1 and others are set 0. Since infrared dim and small targets account for only less than 80 features, the candidate targets masked are very small. Therefore, the results after image inpainting can be theoretically considered as background in these areas. Generally, the background is reasonably inferred by image inpainting via understanding the original image outside the mask.

Tracing back to coarse detection module, there are many target candidate areas where it is not sufficient to accurately detect via the direct segmentation due to the interference of noise and bright background. There are two circumstances in these areas: 1) A area totally belongs to the background. Under the circumstance, the background is incorrectly considered as the target. The subsequent task is to re-judge the area as background by inpainting. 2) A area includes small targets. The key of detecting small targets is that inpainting must follow global semantic information to generate the background projections for the area.

The masks, obtaining from the target candidate regions by threshold segmentation, have small and isolated shapes. However, many existing algorithms for image inpainting repair RGB images by large and continuous masks, which lack of understanding of the semantic information of infrared dim and small target images. Thus, these algorithms are not suitable for repairing infrared dim small target images. In this paper, an image inpainting method with MADF is unitized to better predict background of infrared images. The algorithm is able to generate the convolutional kernel following the shape of masks. Thus, it has good inpainting effect on RGB images with small and isolated masks.

Figure 3 indicates the architecture of fine detection module. In the fine detection stage, inpainting employs a cascaded neural network structure with mask awareness proposed in [22]. The structure consists of encoder and decoder. The whole network structure has 7 layers with the same structure and different network parameters. Every layer includes an encoder module and two decoder modules. There are two functions of encoder: 1) extract the shallow and deep features for understanding the global features and semantics of the images by convolution layer; 2) integrate mask awareness into feature maps by the MADF module. The decoder is divided into recovery decoder and refinement decoder. The recovery decoder is designed to roughly restore missing areas of images from deep to shallow. After the recovery decoder, the feature map is optimized by the refinement decoder to generate delicate inpainting results. The image with mask is fed to the encoder with MADF to generate feature maps, applied by the decoder to recover the image. Meanwhile, these feature maps are loaded into the deep encoder layer to generate deep features. The final repaired results are obtained by the recovery decoder and the two refinement decoders. In order to ensure that image inpainting algorithm adapts to infrared dim and small target image, the cloud data set demonstrated in Figure 4 is applied to train the inpainting algorithm with MADF. Finally, the detection result is obtained by the difference of infrared image and the final repaired result in Image Fusion module.

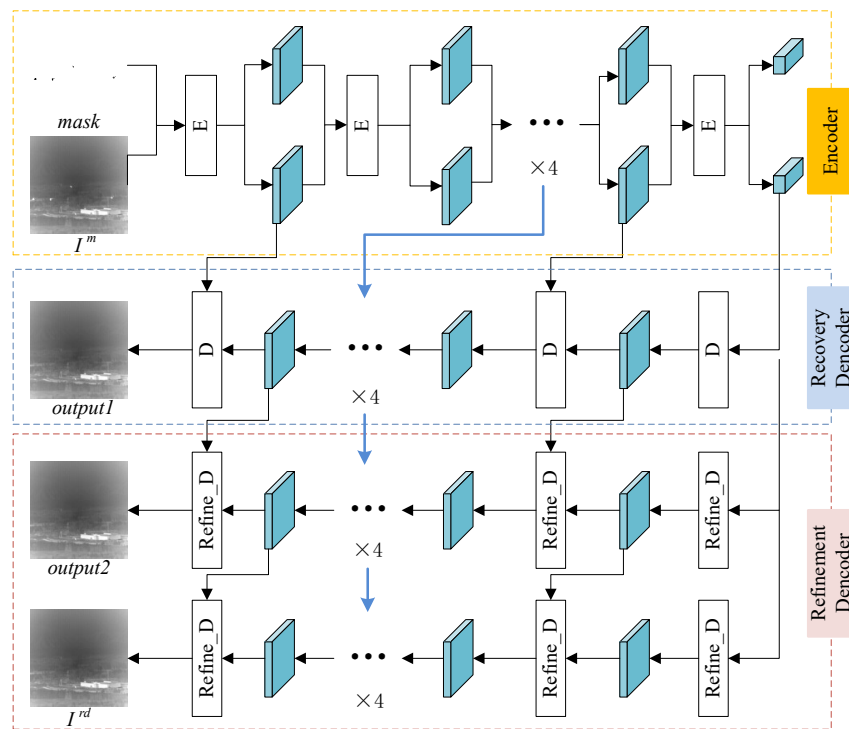


Figure 3. The architecture of the coarse detection module. The E, D and Refine_D denotes encoder lay, recovery decoder lay, and refinement decoder lay, respectively. The output1, output2 and repaired image are three output of decoder. The repaired image is the result of the finest image inpainting.

4. Results

In this section, a sky clouds dataset for training image inpainting method is introduced. The basic content of the datasets and evaluation metrics in the experiment are demonstrated. Then, in the experiment, several exiting infrared small target detection methods are compared with the proposed. Ultimately, qualitative and quantitative analysis and discussion of experimental results are presented.

4.1. Datasets

Recently, many image inpainting methods specialize in processing RGB images, leading not applicable to infrared images. For the purpose of obtaining high quality repaired results of infrared images, a dataset of pure infrared cloud images is proposed for the training phase of image inpainting. This dataset is composed of various cloud images. Images of the dataset are acquired by cropping and segmenting original images in MATLAB 2019b. These images were made into a dataset containing 43,500 infrared background images including cirrus, stratus, cirrocumulus, etc. The part of the dataset is displayed in Figure 4.



Figure 4. The part of train dataset. The dataset includes infrared images of various cloud formations.

Moreover, to evaluate the performance of the proposed framework, three public datasets are utilized, including SIRST dataset [43], IRSTD-1k dataset [44] and IRSTD640 dataset [16]. The SIRST dataset consists of 427 typical images and 480 instances of various scenes isolated from hundreds of real-world videos. It is widely applied to detect infrared dim small targets [43]. The IRSTD-1k dataset is a publicly available dataset consisting of many infrared dim small target images of complex backgrounds, including 1,000 images that have varying-shape targets and low contrast and low SCR background with clutters and noises and ground truth images corresponding to them [44]. The NUDT-SIRST dataset is composed of composite images of 5 main background scenes and a few IR images proposed in [45]. The MFIRST [32] dataset includes 9,956 training images and 100 test images. These images come from realistic infrared sequences and synthetic infrared images [43]. The MFIRST is utilized to train the model of the coarse detection module.

4.2. Implementation Details

The size of the original image is set 256×256 . To ensure that the Region Proposal Network acquires target candidate areas, the size of the anchor box is fixed as 10×10 . In the proposed framework, RPN and image inpainting method are trained respectively. RPN is trained separately by the similar loss function of [15] in the MFIRST dataset [32]. Our training model of image inpainting is constructed on the basis of the network, which has an even favorable repair effect on high resolution color images in [22]. Therefore, for the purpose of convenience, one of the pretrained models (the model for Places2 [39]) in [22] is employed as our start point of our model. The training of irregular masks is from datasets of mask in [39]. The detail parameters of the model are similar to the training process in [22]. The learning rate is set to 0.0002. We train our model for about 300K iterations. Our model was trained on a GeForce RTX 2080 GPU (8G) with batch size of 8.

4.3. Evaluation Metrics Indicators

In this paper, in order to examine the performance of the proposed framework and compare the effectiveness of different methods, as with many segmentation-based dim small target detection methods, several evaluation metrics are adopted, liking precision rate, recall rate, F-measure and receiver operating characteristic curve (ROC).

F-measure: F-measure is a primary evaluation metric in object segmentation methods [33]. It balances precision rate (Prec) and recall rate (Rec) to favorably represent the ability of precisely detecting targets with fewer false alarms. The Prec and Rec are defined as follows.

$$\text{Prec} = \frac{T_P}{T_P + F_P}, \quad (2)$$

$$\text{Rec} = \frac{T_P}{T_P + F_N}, \quad (3)$$

where T_P and F_P denote the number of successfully detected target pixels and background pixels that are incorrectly detected as targets, respectively. F_N denotes the number of ground truth target pixels that are mistakenly recognized as background. The F-measure is defined as [32] by Prec and Rec,

$$\text{F-measure} = \frac{(1 + \beta^2)\text{Prec} * \text{Rec}}{\beta^2\text{Prec} + \text{Rec}}, \quad (4)$$

where β is a constant. In the paper, it is set as 1. Thus, the F-measure is named as F1-measure (F1).

Receiver Operation Characteristics: Receiver Operation Characteristics (ROC) is a target level metric that is utilized to record and represent the tendency of the detection probability (P_d) under different false alarm rate (F_a). The detection probability (P_d) and false alarm rate (F_a) are defined as [33]:

$$P_d = \frac{T_{\text{true detections}}}{T_{\text{gt}}}, \quad (5)$$

$$F_a = \frac{P_{\text{false}}}{P_{\text{image}}}, \quad (6)$$

where $T_{\text{true detections}}$ and T_{gt} denote the number of targets that are truly detected and ground-truth targets, respectively. P_{false} and P_{image} represent the number of the false detection pixels and total pixels of images. Here, if the distance between centers of the ground truth and the predicted result is less than 4 pixels, the predicted result is regarded as a correct detection. Otherwise, the result is viewed as false.

4.4. Contrast Methods and Parameter Setting

To assess the detection performance of the proposed framework for different scenarios, in this part, we implemented comparison experiments employing different principles of comparative methods. Conventional methods include background prediction-based, local feature-based and sparse matrix-based approaches like Top-Hat [24], Local Intensity and Gradient properties (LIG) [29], Average Absolute Gray Difference (AAGD) [30], Multiscale Tri-layer Local Contrast Measure (TLLCM) [9], Non-convex Rank Approximation Minimization (NRAM) [31] and Partial Sum of the Tensor Nuclear Norm (PSTNN) [46]. In recent years, many deep learning-based infrared small target detection methods have been proposed, liking Attentional Local Contrast Network (ALCNet) [14], Dense Nested Attention Network (DNANet) [33], Attention-Guided Pyramid Context Network (AGPCNet) [34], etc. In this paper, the parameter settings of these traditional methods are chosen according to the parameters suggested in their papers, as illustrated in Table 1. For methods based on deep learning, the optimal models from the original papers are adopted as the detection network models.

Table 1. The parameters of the conventional methods.

Methods	Parameters
TopHat	structure shape: square, size: 5×5
LIG	window size: 11×11 , $k = 0.2$
AAGD	internal window scale: [3, 5, 7, 9], external window size: 19×19
TLLCM	Gaussian kernel size: 3×3 , scale: [3, 5, 7, 9]
NRAM	patch size: 50×50 , slide step = 10, $\lambda = \sqrt{\min(m, n)}$
PSTNN	patch size = 40, slide step = 40, $\lambda = 0.7$

4.5. Contrast Experiment Results

Qualitative Comparison. For comparing the performance of these algorithms by visual comparison, several images with various backgrounds are selected as the comparative data. Figure 5 illustrates these raw infrared images, where targets are marked by red boxes. Figure 5(1, 2, 5-7) contain complex background clutters. The background of Figure 5(1) consists of trees. In Figure 5(2), targets are filled with mountains and a number of bright backgrounds. The building noise affects infrared dim and small targets in Figure 5(6) and (7). The backgrounds in Figure 5(3), (4), (5) and (8) are flatter, which is composed of irregular clouds, smooth natural and artificial scenes. Figure 6 demonstrates the results of comparative methods. In the figure, for conventional methods, since the results are represented by the saliency map, we uniformly set the threshold to $0.3 \times \max_value$. Meanwhile, to conveniently compare the detection performance of various methods, results of these algorithms are normalized to be in the range of 0-1. In Figure 6, small targets are masked by red boxes and blue circles to represent the clutter in results.

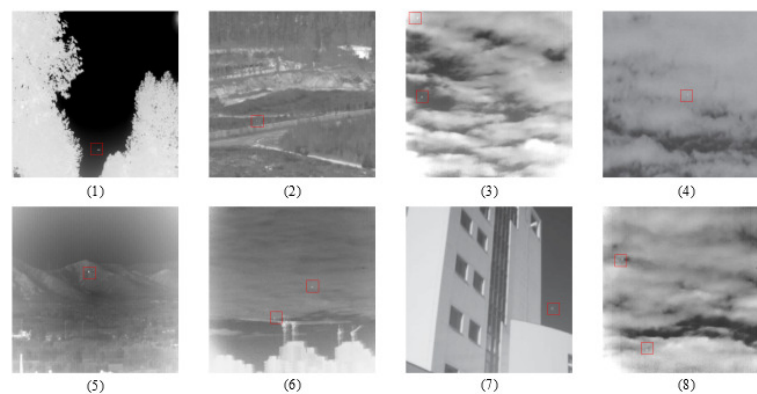


Figure 5. The illustration of representative images, which contain a variety of backgrounds. (1), (2), (5) The images of IRSTD-1k dataset[44]. (3), (6) and (8) are images of d IRSTD640 dataset[16]. (4) and (7) are two samples of SIRST dataset[43].

Figure 6(a) and (g) are the 3-D representations of the original images and the ground-truth, respectively. Intuitively, TopHat has a weak ability to detect small targets. Due to the morphological filtering of backgrounds in local regions, a large amount of background clutter is remained in results, even drowning out targets. LIG, AAGD, and TLLCM all focus on local information, so detection results are better than TopHat in visual terms. In the case of relatively smooth backgrounds, LIG and AAGD are able to detect small targets; however, there is still background noise in results. On the contrary, they achieve poor detection performance for the complicated background with building edges, complex ground and strong cloud clutters. TLLCM is able to separate small targets from the background better, except for the scenes of Figure 6(a2), (a3) and (a6). Especially, when TLLCM processes multi-target images, the second target is severely weakened, as shown in Figure 6(k3) and (k6). NRAM and PSTNN convert the problem of target detection to that of sparse matrix separation, nevertheless, in Figure 6(f1-f7) and (h1-h7), background clutters are separated as targets with a high false alarm rate, even with failure in detecting Figure 6(a4). ALCNet, DNANet and AGPCNet are the state-of-the-art deep learning-based methods with higher accuracy and lower false alarm rate compared to the traditional methods. Unfortunately, the results of ALCNet contain a lot of noise when dealing with complex backgrounds and in the Figure 6 (a3) and (a4) scenarios, it cannot even successfully detect targets. DNANet missed the real target and detects the background as the target during detecting Fig 6(a4). In the case of multiple targets, such as Figure. 6(a3), (a4) and (8), three deep learning-based methods, suffer from detection misses. Our method performs local background prediction on target candidate regions generated in coarse detection module to detect dim and small targets. In comparison, as demonstrated in Figure 6(l1-l5), our proposed algorithm can achieve the detection of targets better, although the results contain individual clutter in some scenarios.

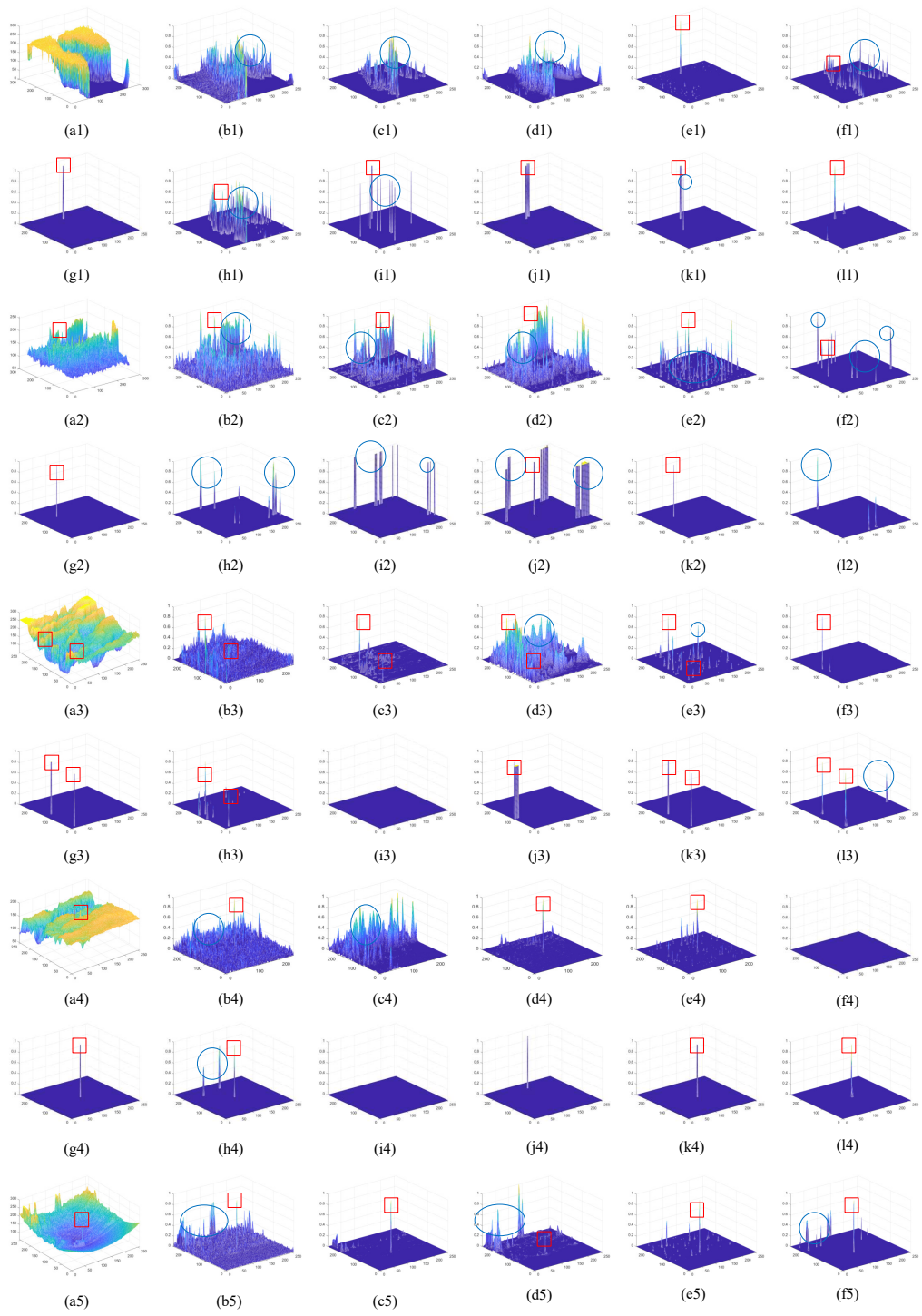


Figure 6. Cont.

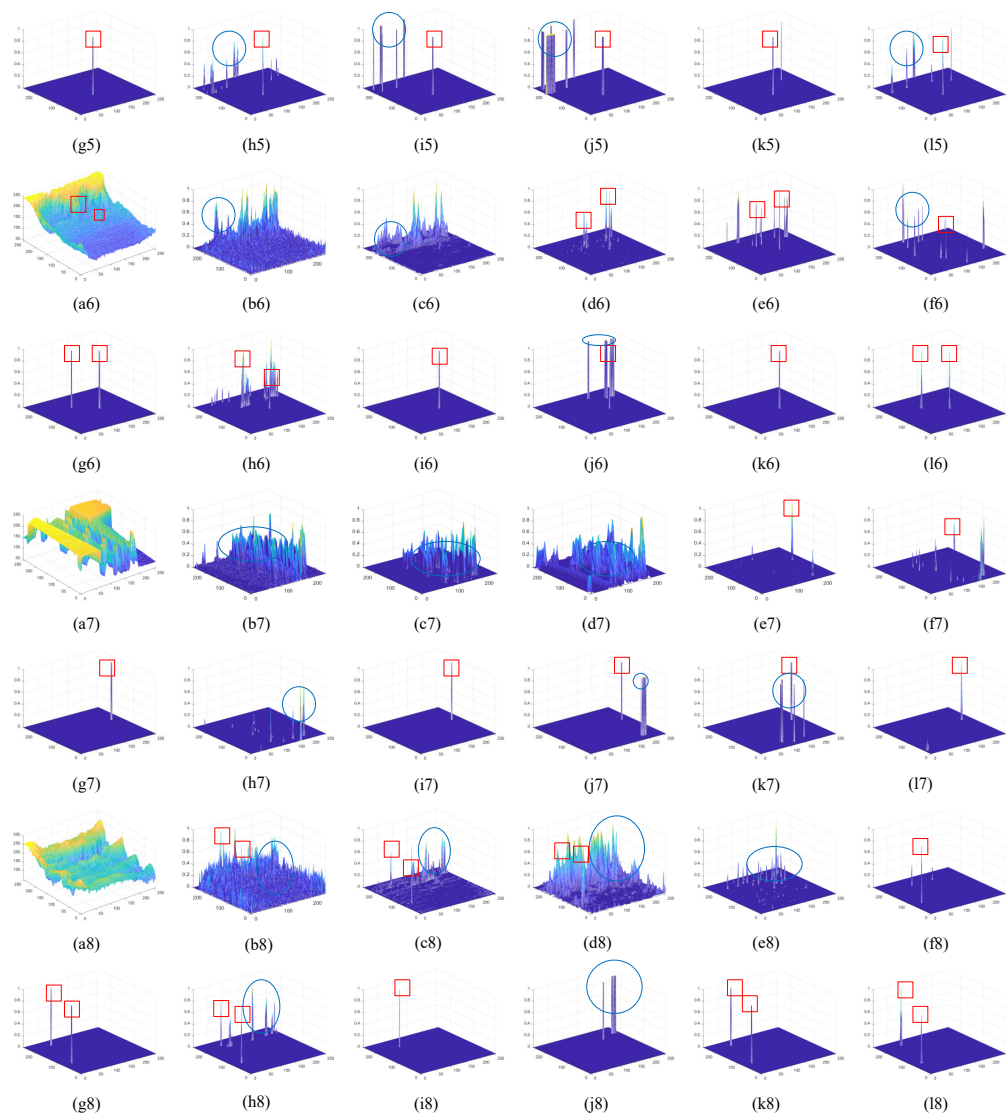


Figure 6. The processed results of the five images in Figure 5 by different methods. (a1-a8) The 3-D mesh representations of five images in Figure 5, respectively. (b1-b8) The Top-Hat. (c1-c8) The LIG. (d1-d8) The AAGD. (e1-e8) The TLLCM. (f1-f8) The NRAM. (g1-g8) The Ground-Truth. (h1-h8) The PSTNN. (i1-i8) The ALCNet. (j1-j8) The AGPCNet. (k1-k8) The DNANet. (l1-l8) The proposed method.

Quantitative Comparison. ROC curves and F1-measure-threshold curves of contrast methods on three public datasets are shown in Figure 7 and Figure 8, respectively. The horizontal axis of Figure 7 is the false alarm rate and the vertical axis is the detection rate. In Figure 8, the vertical axis and horizontal axis are the F1-measure (F1) and threshold, respectively. ROC is a metric that shows the potency of the method by applying different threshold. It can be seen that the proposed method has better effects than methods. Especially, in Figure 7(ii), the proposed algorithm obviously outperforms other comparison algorithms but lower than DNANet. Our method rapidly converges to the highest detection rate when $F_a < 2 \times 10^{-5}$, even reaching 95%. As demonstrated in Figure 7(iii), although the proposed algorithm is faster to be stable, the true positive rate is lower than the ALCNet, AGPCNet and DNANet. ForIRSTD-1k dataset, although many algorithms have a good and stable detection curve, our framework is overall second only to DNANet. In contrast, our algorithm reaches steady state faster than DNANet. F1 is an evaluation metric that balances detection rate and recall rate. The F1 curves of our framework onIRSTD-1k andIRST640 datasets are better than other comparison algorithms. Nonetheless, the F1 of ALCNet, AGPCNet, DNANet and PSTNN are higher than our framework onSIRST dataset when the threshold value is in the range of 0 to 0.15. As the threshold

value gradually increases, the F1 of our algorithm is gradually higher than the other algorithms, second only to LIG.

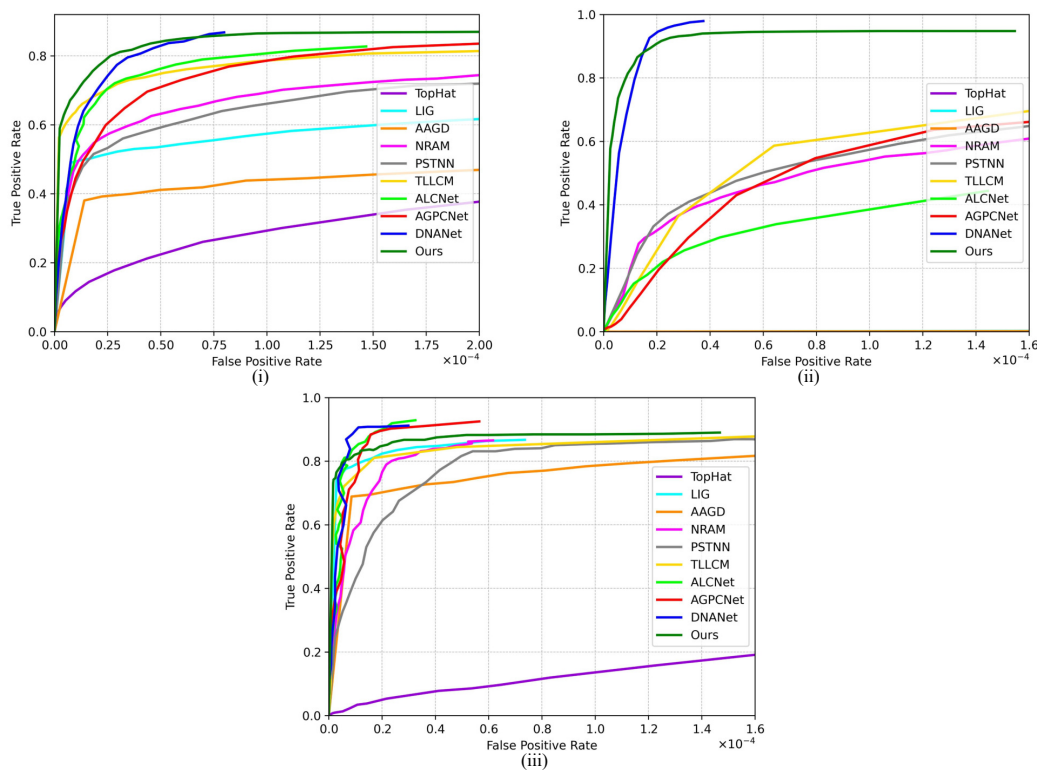


Figure 7. ROC curves of comparative algorithms on (i) IRSTD-1k dataset [44], (ii) IRSTD640 dataset [16], (iii) SIRST dataset [43].

To further compare detection capabilities of each algorithm, we obtained the data in 2 by selecting fixed thresholds from Figure 8. The Prec, Rec and F-measure of Table 2 is calculated by the F-measure calculation method described in [32]. Here, the threshold of calculation is set to 0.15. The Prec reflects the proportion of correct results in the results detected as a target, as shown in Equation (2). The Rec denotes the ratio of the number of true results and the total number of targets, as defined in Equation (3). The F1-measure balances the performance of the Prec and Rec. Ideally, the Prec, Rec and F1-measure are both 1. In tables, the largest value has been bolded, while the secondly largest value is masked by an underscore. For IRSTD-1k and IRSTD640, the Rec and F1-measure of our proposed method are both better than other methods, but the performance of precision rate is worse than DNANet. Unfortunately, our framework is less advantageous compared to other algorithms in SIRST dataset, since most comparison methods have similar metrics performance to the proposed algorithm, as demonstrated in Table 2 SIRST. Though our detection framework is flawed in some aspects, its capability to detect dim and small targets is improved over most algorithms. Finally, through a quantitative evaluation of the metrics, we can see that our framework has a better performance than existing methods. Moreover, these metrics represent our approach can work more stably for various clutter and noisy backgrounds.

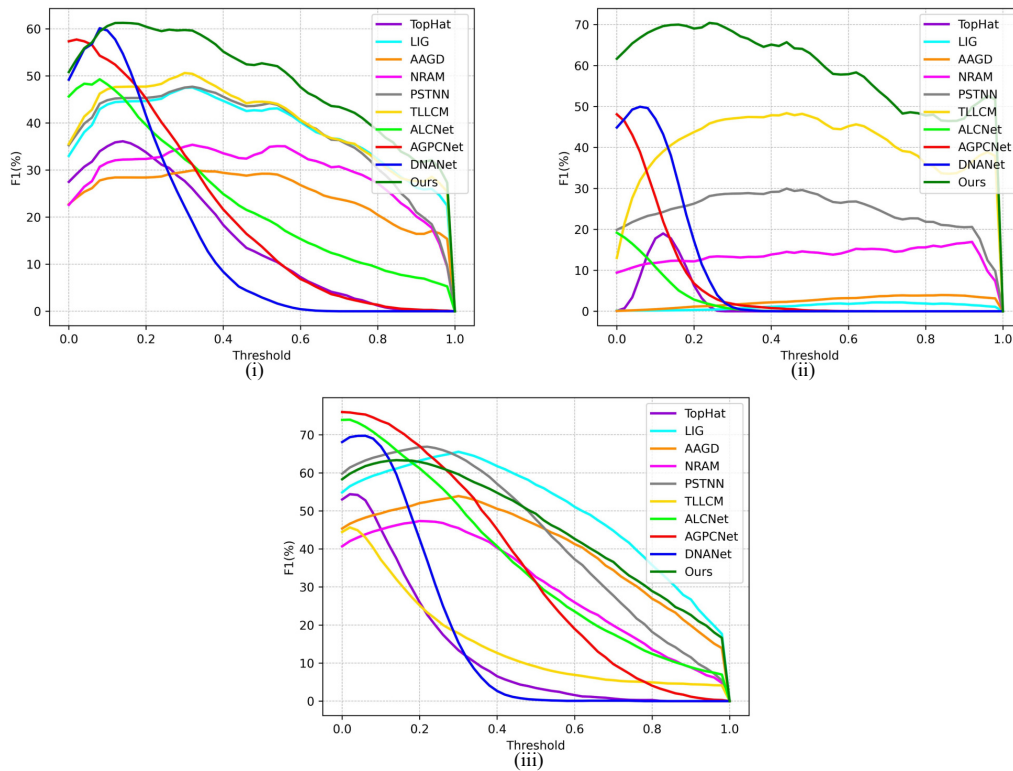


Figure 8. F1-measure-threshold curves on (i) IRSTD-1k dataset [44], (ii) IRST640 dataset[16], (iii) SIRST dataset [43].

Table 2. The average Prec, Rec and F1-measure of IRSTD-1k, IRST640 and SIRST datasets.

Methods	IRSTD-1k			IRST640			SIRST		
	Prec (%)	Rec (%)	F1 (%)	Prec (%)	Rec (%)	F1 (%)	Prec (%)	Rec (%)	F1 (%)
TopHat	42.19	52.66	34.53	41.04	18.16	16.18	65.83	29.63	34.88
LIG	53.43	59.41	47.03	30.63	46.27	29.43	85.80	66.06	69.58
AAGD	25.63	56.27	25.31	1.83	25.94	2.69	61.05	66.14	53.56
NRAM	58.99	30.11	34.35	38.22	8.43	12.27	87.05	37.88	50.10
TLLCM	60.70	56.21	51.63	39.62	63.70	41.44	74.20	23.27	32.55
PSTNN	45.52	59.17	44.82	24.83	36.97	25.16	84.84	61.70	67.70
ALCNet	60.85	38.59	44.37	15.67	4.10	6.00	87.56	55.08	65.18
AGPCNet	55.37	50.58	49.85	28.56	11.63	14.88	83.03	66.59	70.91
DNANet	81.08	42.41	52.99	86.59	22.66	34.59	89.17	43.93	57.02
Ours	67.16	65.83	61.20	71.41	76.82	69.86	85.67	54.86	63.32

5. Discussion

As mentioned earlier, we propose a coarse-to-fine infrared dim and small target detection method, which detects small targets via estimating the background of infrared images. In this section, the role of two detection stages in infrared dim and small target detection are discussed from the results.

The proposed method is composed of coarse detection module and fine detection module. On the one hand, the coarse detection module aims to set the background outside the candidate target region by detecting the candidate targets. On the other hand, it generates the *mask* for next stages. The fine detection module estimates the background inside the candidate target by that of outside these areas through an image inpainting algorithm. The experimental results on the three datasets demonstrate that the framework we proposed has a better accuracy performance on infrared dim and small targets. As Figure 6(l1-l3) and (l5-l7) indicated, the background clutter (noise) also is detected in results; however, the value of clutter is negligible in some results (l1-l3 and l7). As can be seen in Table

2, the Prec, Rec, and F1 on theIRST640 dataset are 71.41%, 76.82%, and 69.86%, respectively, which could meet the index requirements for the detection ability of many infrared scenes. The ROC curves and F1-measure-threshold curves onIRSTD-1k andIRSTD640 datasets indicate that the framework has an outstanding performance over other methods.

In this paper, to conveniently illustrate the principle, two prior assumptions are made in Section 1: 1) Due to the higher pixel value of the target than the background, the background in the small target neighborhood is estimated by the remaining image after eliminating the targets. 2) Since the pixel structures of background clutter (noise) is different that of the target, the background in suspicious target areas (building edges, highlighting noise, etc.) is predicted and similar with the origin image. The coarse detection module obtains the target frame from the depth features in which the candidate targets are generated by simple threshold segmentation. These candidate targets (real targets, suspicious targets, etc.) are fed to the next stage in the form of masks. In this stage, the mask is directly related to the performance of the detection result. As demonstrated in Figure 6(l5) and (l6), the results contain a number of obvious background noise. From experiment, the ability of generating masks is adequately presented. Following assumptions, the generation of masks is the key of the whole algorithm. However, the mask generation still has unfriendly aspects due to the limitation of training datasets and the neglect in networks for the feature of small targets. In the fine detection module, the background within these areas is predicted by infrared image outside the candidate targets. Ideally, the estimation of the background within non-targeted regions does not differ from the infrared image. Unfortunately, the results in Figure 6 indicate that there are minor flaws in the algorithm for the prediction of non-target regions. Therefore, subsequent research is still needed to optimize the background generation to improve the target detection performance.

6. Conclusions

Generally, we proposed a coarse-to-fine deep learning-based framework for infrared dim and target detection. In the coarse detection stage, RPN is utilized to generate target candidate areas, where the mask for labeling candidate targets is obtained by threshold segmentation. Inpainting, as a technology for predicting the background, is employed to restore the background of images by global semantics information in these masks, which further optimizes the performance of target detection in coarse detection. In the experiments, the proposed method is compared with other existing methods on several publicly available datasets, verifying that the proposed method outperforms the current existing methods in both subjective visual quality and objective quantitative measurements. However, since the dataset for training the image inpainting algorithm is mostly sky cloud background, the proposed lack of robust for infrared images with extremely complex backgrounds.

Author Contributions: J.M. and S.R. proposed the original idea. B.H. and J.F. provided the spaces and equipments. J.M. and H.G. collected data on infrared dim and small targets and performed the experiments. J.M. wrote the manuscript. J.M and H.G. reviewed and edited the manuscript. S.R. contributed to the direction, content, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grants No. 62001443 and the Natural Science Foundation of Shandong Province under Grants No. ZR2020QE294.

Data Availability Statement: The data used for training and test set are available at: <https://github.com/YimianDai/sirst>, <https://github.com/RuiZhang97/ISNet>, <https://github.com/jzchenriver/IRST640> and https://github.com/wanghuanphd/MDvsFA_cGAN, accessed on 20 April 2023.

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

References

1. Rawat, S.S.; Alghamdi, S.; Kumar, G.; Alotaibi, Y.; Khalaf, O.I.; Verma, L.P. Infrared small target detection based on partial sum minimization and total variation. *Mathematics* **2022**, *10*, 671.

2. Zhang, W.; Cong, M.; Wang, L. Algorithms for optical weak small targets detection and tracking. In Proceedings of the International Conference on Neural Networks and Signal Processing, Nanjing, China, 14-17 December 2003; Volume 1, pp. 643–647.
3. Jiao, J.; Lingda, W. Infrared dim small target detection method based on background prediction and high-order statistics. In Proceedings of the International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 02-04 June 2017; pp. 53–57.
4. Zhou, A.; Xie, W.; Pei, J. Background Modeling in the Fourier Domain for Maritime Infrared Target Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2634–2649.
5. Hu, Z.; Su, Y. An infrared dim and small target image preprocessing algorithm based on improved bilateral filtering. In Proceedings of the International Conference on Computer, Blockchain and Financial Development (CBFD), Nanjing, China, 23-25 April 2021; pp. 74–77.
6. Chen, C.L.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A Local Contrast Method for Small Infrared Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 574–581.
7. Han, J.; Ma, Y.; Zhou, B.; Fan, F.; Liang, K.; Fang, Y. A Robust Infrared Small Target Detection Algorithm Based on Human Visual System. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2168–2172.
8. Han, J.; Liang, K.; Zhou, B.; Zhu, X.; Zhao, J.; Zhao, L. Infrared Small Target Detection Utilizing the Multiscale Relative Local Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 612–616.
9. Han, J.; Moradi, S.; Faramarzi, I.; Liu, C.; Zhang, H.; Zhao, Q. A Local Contrast Method for Infrared Small-Target Detection Utilizing a Tri-Layer Window. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1822–1826.
10. Xiong, B.; Huang, X.; Wang, M. Local Gradient Field Feature Contrast Measure for Infrared Small Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 553–557.
11. Chen, X.; Xu, W.; Tao, S.; Gao, T.; Feng, Q.; Piao, Y. Total Variation Weighted Low-Rank Constraint for Infrared Dim Small Target Detection. *Remote Sens.* **2022**, *14*, 4615.
12. Zhang, P.; Zhang, L.; Wang, X.; Shen, F.; Pu, T.; Fei, C. Edge and Corner Awareness-Based Spatial-Temporal Tensor Model for Infrared Small-Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, pp. 1–17.
13. 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.
14. 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.
15. Wang, K.; Du, S.; Liu, C.; Cao, Z. Interior Attention-Aware Network for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13.
16. Chen, G.; Wang, W.; Tan, S. IRSTFormer: A Hierarchical Vision Transformer for Infrared Small Target Detection. *Remote Sens.* **2022**, *14*, 3258.
17. Hariharan, B.; Arbelaez, P.; Girshick, R.; Malik, J. Object Instance Segmentation and Fine-Grained Localization Using Hypercolumns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 627–639.
18. Hou, Q.; Wang, Z.; Tan, F.; Zhao, Y.; Zheng, H.; Zhang, W. RISTDnet: Robust Infrared Small Target Detection Network. *IEEE Geosci. Remote Sens. Lett.* **2021**, pp. 1–5.
19. Liu, H.; Wan, Z.; Huang, W.; Song, Y.; Han, X.; Liao, J. PD-GAN: Probabilistic diverse GAN for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 19-25 June 2021; pp. 9371–9381.
20. Lin, Y.; Xie, H. Face gender recognition based on face recognition feature vectors. In Proceedings of the IEEE International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 27-29 September 2020; pp. 162–166.
21. Wu, H.; Zhou, J.; Li, Y. Deep generative model for image inpainting with local binary pattern learning and spatial attention. *IEEE Trans. Multimedia* **2021**, *24*, 4016–4027.
22. Zhu, M.; He, D.; Li, X.; Li, C.; Li, F.; Liu, X.; Ding, E.; Zhang, Z. Image inpainting by end-to-end cascaded refinement with mask awareness. *IEEE Trans. Image Process.* **2021**, *30*, 4855–4866.
23. 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, Denver, Colorado, 20–22 July 1999; Volume 3809, pp. 74–83.
24. Bai, X.; Zhou, F. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognit.* **2010**, *43*, 2145–2156.

25. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226.
26. Han, J.; Moradi, S.; Faramarzi, I.; Zhang, H.; Zhao, Q.; Zhang, X.; Li, N. Infrared Small Target Detection Based on the Weighted Strengthened Local Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2020**, pp. 1–5.
27. Lu, R.; Yang, X.; Li, W.; Fan, J.; Li, D.; Jing, X. Robust Infrared Small Target Detection via Multidirectional Derivative-Based Weighted Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5.
28. Li, Z.; Liao, S.; Zhao, T. Infrared Dim and Small Target Detection Based on Strengthened Robust Local Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5.
29. Zhang, H.; Zhang, L.; Yuan, D.; Chen, H. Infrared small target detection based on local intensity and gradient properties. *Infrared Phys. Technol.* **2018**, *89*, 88–96.
30. Aghaziyarati, S.; Moradi, S.; Talebi, H. Small infrared target detection using absolute average difference weighted by cumulative directional derivatives. *Infrared Phys. Technol.* **2019**, *101*, 78–87.
31. Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; Peng, Z. Infrared Small Target Detection via Non-Convex Rank Approximation Minimization Joint $l_{2,1}$ Norm. *Remote Sens.* **2018**, *10*, 1821.
32. 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, Long Beach, CA, USA, 16–20 June 2019; pp. 8509–8518.
33. Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; Guo, Y. Dense Nested Attention Network for Infrared Small Target Detection. *IEEE Trans. Image Process.* **2022**.
34. Zhang, T.; Li, L.; Cao, S.; Pu, T.; Peng, Z. Attention-Guided Pyramid Context Networks for Detecting Infrared Small Target Under Complex Background. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, pp. 1–13.
35. Chen, F.; Gao, C.; Liu, F.; Zhao, Y.; Zhou, Y.; Meng, D.; Zuo, W. Local Patch Network with Global Attention for Infrared Small Target Detection. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, pp. 1–1.
36. Wang, N.; Zhang, Y.; Zhang, L. Dynamic selection network for image inpainting. *IEEE Trans. Image Process.* **2021**, *30*, 1784–1798.
37. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2536–2544.
38. Yang, C.; Lu, X.; Lin, Z.; Shechtman, E.; Wang, O.; Li, H. High-resolution image inpainting using multi-scale neural patch synthesis. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6721–6729.
39. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 85–100.
40. Liu, H.; Wan, Z.; Huang, W.; Song, Y.; Han, X.; Liao, J.; Jiang, B.; Liu, W. Deflocnet: Deep image editing via flexible low-level controls. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 19–25 June 2021; pp. 10765–10774.
41. Wu, H.; Zhou, J. IID-Net: Image inpainting detection network via neural architecture search and attention. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1172–1185.
42. Cai, W.; Wei, Z. PiiGAN: generative adversarial networks for pluralistic image inpainting. *IEEE Access* **2020**, *8*, 48451–48463.
43. 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, 19–25 June 2021; pp. 950–959.
44. Zhang, M.; Zhang, R.; Yang, Y.; Bai, H.; Zhang, J.; Guo, J. ISNet: Shape Matters for Infrared Small Target Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 877–886.
45. Hui, B.; Song, Z.; Fan, H.; Zhong, P.; Hu, W.; Zhang, X.; Ling, J.; Su, H.; Jin, W.; Zhang, Y. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background. *China Sci. Data* **2020**, *5*, 291–302.
46. Zhang, L.; Peng, Z. Infrared Small Target Detection Based on Partial Sum of the Tensor Nuclear Norm. *Remote Sens.* **2019**, *11*, 382.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.