electronics

Article

Deep Learning-Based Small Target Detection for Satellite–Ground Free Space Optical Communications

Nikesh Devkota and Byung Wook Kim *

Department of Information and Communication Engineering, Changwon National University, Changwon 51140, Republic of Korea; 20227085@gs.cwnu.ac.kr
* Correspondence: bwkim@changwon.ac.kr

Abstract: Free space optical (FSO) channels between a low earth orbit (LEO) satellite and a ground station (GS) use a highly directional optical beam that necessitates a continuous line-of-sight (LOS) connection. In this paper, we propose a deep neural network (DNN)-based small target detection method that detects the position of a LEO satellite in an infrared image, which can be used to determine the receiver alignment for establishing the LOS link. For the infrared small target detection task without excessive down-sampling, we design a target detection model using a modified ResNest-based feature extraction network (FEN), a custom feature pyramid network (FPN), and a target determination network (TDN). ResNest utilizes the feature map attention mechanism and multi-path propagation necessary for robust feature extraction of small infrared targets. The custom FPN combines multi-scale feature maps generated from the modified ResNest to obtain robust semantics across all scales. Finally, the semantically strong multi-scale feature maps are fed into the TDN to detect small infrared targets and determine their location in infrared images. Experimental results using two widely used point spread functions (PSFs) demonstrate that the proposed algorithm outperforms the conventional schemes and detects small targets with a true detection rate of 99.4% and 94.0%.

Keywords: LEO satellite; deep learning; free space optical communication; small infrared target

Citation: Devkota, N.; Kim, B.W. Deep Learning-Based Small Target Detection for Satellite–Ground Free Space Optical Communications. *Electronics* **2023**, *12*, 4701. https://doi.org/10.3390/electronics12224701

Academic Editor: Martin Reisslein

Received: 14 October 2023
Revised: 11 November 2023
Accepted: 17 November 2023
Published: 19 November 2023

1. Introduction

Recently, free space optical (FSO) communications have attracted considerable attention from researchers and industry for satellite-to-ground station (satellite-GS) links and inter-satellite links owing to the extensive unlicensed bandwidth, low power consumption, and high data rates [1–3]. Satellite-GS FSO communication research is becoming especially relevant because it can provide high speed and reliable voice and data communication services in locations where terrestrial cellular and broadband access is not possible or where network coverage is insufficient [3,4]. The European Space Agency (ESA) successfully tested the first inter-satellite laser communication link between the SPOT-4 and ARTEMIS satellites for optical data-relay services at 50 Mbps [4]. The ETS-VI satellite and a GS in Konegi, Japan, established the first successful ground–satellite optical link [5]. NASA's Laser Communication Relay Demonstration (LCRD) [6], which launched in 2021, and the Terabyte Infrared Delivery (TBIRD) [7] CubeSat mission, which demonstrated laser downlinks at 200 Gbps from LEO satellites, have already paved the way for future optical communications missions.

FSO communication uses a highly directional optical beam with a very narrow beam divergence. Consequently, receivers in FSO links have a limited field of view (FOV) [8]. During downlink, if the optical beam emitted by a satellite is not directed at the FOV of the receiving aperture at the GS, it will dissipate into the atmosphere, resulting in a decrease in received signal power, an increase in bit error rate (BER), and communication failure. Hence, an FSO-based communication system between a LEO satellite and a GS needs

a direct line-of-sight (LOS) connection. For this, the GS must have a robust mechanism to detect and track the position of the LEO satellite. Although a two-line element (TLE) can obtain satellite position information, they are not precise enough to accurately point a GS toward its target satellite [9]. Since the orbit of LEO satellites such as nano- and pico-satellites tend to shift more quickly than those of bigger satellites, their TLE accuracy is very low [10]. Furthermore, GS-based telescopes equipped with a tracking system are more suitable for tracking slow-moving space objects such as stars or planets rather than tracking fast-moving LEO satellites [11]. Due to the narrow FOV and slow tracking, slight changes in the angle of the GS-based telescope can cause LOS connection failures for LEO satellites.

A viable option to establish an accurate LOS link for FSO-based communication between a LEO satellite and a GS is to use high-output light emitting diodes (LEDs) as light sources of LEO satellites. The LED source can appear as an artificial star, and it can be captured using a wide-FOV camera at a GS. Experiments on the detection of a satellite as an artificial star in an image for visible light communication (VLC) have been addressed in [11,12]. While the LEO satellite transmits the light signal, a camera with a wide FOV at the GS can capture images containing the LEO satellite as a small target. This small target in the image is detected and tracked to determine the accurate position of the satellite. Then, an uplink beacon laser beam is transmitted from the GS to the target satellite so that the optical transceiver present in the satellite can change its orientation towards the GS. Finally, a robust LOS link is created to initiate the data transmission.

With advancements in camera sensor technology for remote sensing, there has been growing interest in the usage of short-wave infrared (SWIR) bands [13]. The SWIR spectrum is a frequency band located between the visible and mid-infrared bands of the electromagnetic spectrum, and it has a wavelength range of about 1–3 μm [14]. Compared to visible bands, SWIR bands are less impacted by atmospheric scattering and absorption [13]; thus, a SWIR camera is used at the GS to capture LEO satellites as small targets in infrared images.

Although numerous studies have been conducted on detecting small targets from infrared images, challenging problems still remain. For long-distance imaging, the small targets in infrared images are minute in size and occupy only a few pixels of the whole image [15]. Furthermore, the signal-to-clutter ratio (SCR) of an infrared image for a long distance is small, because the intensity of infrared radiation is inversely proportional to the square of the imaging distance [15]. As a result, the contrast between the infrared small target and the neighboring background is low, making it difficult to distinguish targets from the background. Several conventional techniques, such as filtering-based techniques, local contrast-based techniques, and low-rank-based techniques, have been proposed to identify infrared small targets [15]. Moradi et al. [15] used the multi-scale average absolute gray difference (AAGD) and Laplacian of point spread function (LoPSF) for infrared small target detection. Zhang et al. [16] utilized local intensity and gradient properties (LIG) to detect small targets. These conventional techniques significantly rely on handcrafted features and fixed hyper parameters that are usually only effective against a specific background [17]. Additionally, these methods generally utilize only grayscale values to distinguish small targets, resulting in limited generalization capabilities [18].

With the advancements in deep neural networks (DNNs), significant progress has been made in the field of computer vision, especially in image classification and object detection. Traditional methods rely on handcrafted feature extraction based on prior knowledge. In contrast to traditional techniques, DNNs extract discriminative features of an object, such as edges, corners, texture, color, and shape, directly from input images using a data-driven approach. As a result, they can distinguish objects and patterns in the input data better than traditional methods.

Generally, the DNN structure includes down-sampling mechanisms to decrease the size of feature maps and iteratively widen the receptive field. However, small infrared targets require high-resolution feature maps with more details to differentiate them from background noise. Excessive down-sampling in DNN structures can lead to a loss in spatial resolution in small target detection resulting in low-resolution feature maps. This

can blur the small target or merge it with the background in low-resolution feature maps. Consequently, it becomes difficult for the network to distinguish the features of a small target from other objects and background noise. Hence, general DNN-based target detection networks using excessive down-sampling, such as YOLO [19] and RetinaNet [20], perform poorly on the feature extraction of small infrared targets [21]. A two-stage infrared small target detection technique based on a region proposal and a DNN was proposed by Fan et al. [22]. In the first stage, the small target's intensity is increased in accordance with the characteristics of local intensity, and then, potential target regions are suggested by using corner detection. In the second stage, each potential target region is fed into a DNN classifier to remove non-target regions from the images. However, this two-stage method is computationally expensive and cannot meet real-time requirements. To improve small target feature extraction capability and to reduce time consumption, Du et al. [23] proposed a DNN-based infrared dim small target detection algorithm using target-oriented shallow–deep features and an effective small anchor. The term "dim" implies that the small targets in infrared images have low contrast or are not very bright compared to the background due to a long imaging distance [23]. The proposed method in [23] combines semantically robust deep features with target-oriented shallow features to address the issue of low accuracy in detecting small targets. However, the ResNet50 backbone for their detection network does not utilize feature map attention and multi-path representation, which can improve the performance of visual recognition tasks. In addition, they used only a one-to-one concatenation technique between a shallow layer and a deep layer instead of combining all layers from the ResNet50 feature extraction network (FEN) to obtain multi-scale feature maps.

Therefore, to overcome the limitations of previous approaches, we propose a novel DNN-based small target detection model for LEO satellite-GS FSO communications, which consists of ResNest [24] FEN architecture, a custom feature pyramid network (FPN) [25] and a target determination network (TDN). The original ResNest101-based FEN is modified via an ablation study to exploit the feature map attention mechanism and multi-path propagation without excessive down-sampling. The modification is performed through adding or removing intermediate layers in ResNest101 and limiting the down-sampling process to three. In this study, the modified ResNest-based FENs in the ablation study are named ResNestX, where X = 23, 32, 89, and 92 and represents the number of layers in the FEN network after modification. The custom FPN combines multi-scale feature maps generated via ResNest101 to create robust semantics at all scales. The custom FPN structure employed in this study uses a top-down and a bottom-up approach similar to the path aggregation network (PANET) [26]. Furthermore, we use a skip connection from the original input to the output node if they are at the same spatial resolution level. By using a skip connection, the custom FPN preserves the original information obtained from the three FEN outputs while enabling the concatenation of additional features at a low computational cost. The TDN, which consists of a classification network in parallel with a bounding box regression network, is used to identify small targets in infrared images. For the simulation of target generation using a SWIR camera, two commonly used point spread functions (PSF), i.e., a two-dimensional (2D) Gaussian function and a Moffat function, are used. Since the SWIR imaging of the target satellite is affected by atmospheric channel conditions such as clouds, turbulence, fog, light pollution, and background stars, random values of diffusion parameters are applied in PSF, and random Gaussian noise is added to the infrared image. Experimental results show that the proposed scheme can obtain small target detection accuracy up to 99.4% and 94.0% for synthetic small infrared target data generated using a 2D Gaussian function and a Moffat function, respectively.

The contributions of our study are as follows:

- To detect small targets from infrared images, we propose a DNN model that consists of a FEN, a custom FPN, and a TDN, and we evaluate the detection accuracy in various LOS link scenarios between a LEO satellite and a GS.

- Using the 2D Gaussian function and the Moffat function, we generate synthetic image datasets, which include small targets and background noise, to train the DNN model.
- In order to find the optimal combination of convolutional stages from the FEN without excessive down-sampling, we conducted an ablation study through adding or removing intermediate layers of the ResNest101-based FEN.
- The structure of the custom FPN is designed to fuse feature maps of different spatial resolutions and obtain multi-scale feature maps in which all levels, including high-resolution levels, are semantically strong.
- The TDN, which consists of a classification network in parallel with a bounding box regression network, is designed to accurately identify small infrared targets in infrared images.

The rest of this paper is organized as follows. Section 2 describes the system model of an FSO communication environment between a LEO satellite and a GS. Section 3 provides a detailed explanation of our proposed deep learning-based small target detection model. The experiments and results obtained from this study are discussed in Section 4. Section 5 provides the discussion about future work and Section 6 concludes our study.

2. System Model

In this paper, we consider an FSO communication environment between a LEO satellite and a GS. In order to establish an LOS FSO link, both the LEO satellite and the GS must be precisely aligned, owing to the directional optical beam with a very narrow beam divergence. Note that the LEO satellite is available within the LOS range of the GS for only a limited time period [3]. When real-time multimedia transmission is required, a large amount of data must be transmitted from the LEO satellite to the GS within this limited time period [5]. If the FSO signal transmitted from the satellite is not aimed directly at the FOV of the receiving aperture of the GS during this timeframe, it will dissipate into the atmosphere. This results in a decrease in received signal power, an increased BER, and communication disconnect. Hence, an FSO-based GS must establish a robust LOS connection with the satellite to initiate data transmission.

To align the receiving aperture of the GS with the satellite, telescopes equipped with a tracking system can be exploited [11]. Using TLE parameters, the tracking system aims the telescope towards the estimated position of the satellite as it moves across the sky. TLE is a standard format for encoding a satellite's orbital information, describing the orbit and position of the satellite at a particular time. Each satellite's TLE can be detected, tracked, and updated by the North American Aerospace Defense Command (NORAD) [11], which tracks and monitors satellites using radar systems and optical sensors to generate the TLE. However, this method only guides the telescope to the expected range of the satellite's position, which is not precise enough to establish the FSO communication link [11]. Note that conventional telescopes equipped with a tracking system at the ground station are more suitable for tracking slow-moving space objects such as stars or planets rather than tracking fast-moving LEO satellites. In addition, due to the telescope's extremely small FOV [4], mechanical variation or operational delay in the telescope angle might lead to missing the LOS link.

In the process of FSO communication link alignment, the SWIR camera integrated with the proposed DNN-based target detection method locates the fast-moving LEO satellite with high accuracy and precision in a wide FOV area as illustrated in Figure 1. First, the telescope at the GS is coarsely the orbital area of the LEO satellite with the help of TLE parameters. When the LEO satellite comes within LOS range of the GS, the light source in the satellite is captured by the SWIR camera at the GS, and infrared images containing the LEO satellite as a small point target are obtained. Then, DNN-based small target detection is performed to detect and track the LEO satellite position in the captured infrared images. After tracking the satellite, the GS transmits an FSO uplink beacon signal to the LEO satellite. Based on the uplink beacon signal, the LEO satellite aims its antenna towards the GS in order to precisely align itself [11]. With accurate alignment between both the

LEO satellite and the GS, data transmission for a two-way FSO communication channel is initiated.

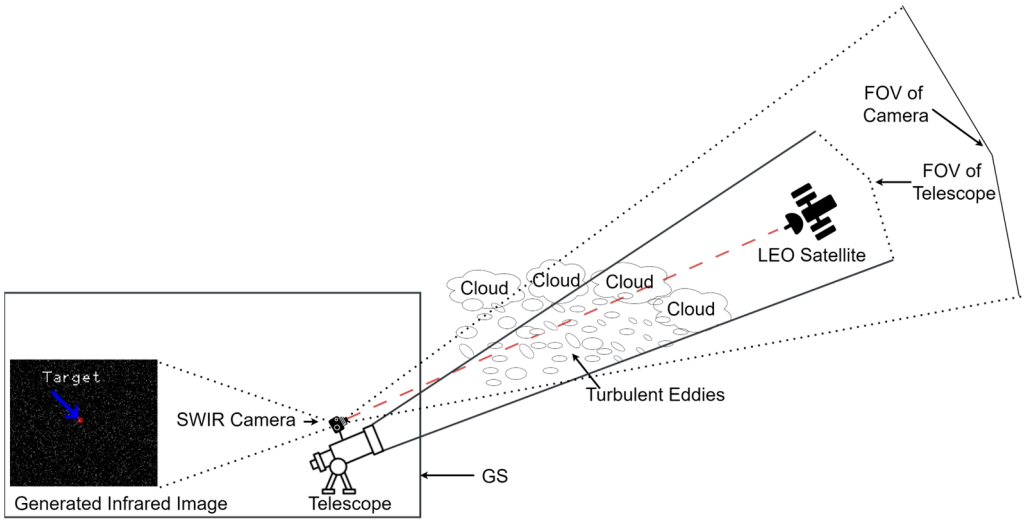

Figure 1. FSO communication environment for detecting and tracking the position of a satellite using a SWIR camera.

The intensity of infrared radiation is inversely proportional to the square of the imaging distance and is attenuated due to the channel conditions, such as clouds, fogs, and atmospheric turbulence. In addition, there might be a presence of light pollution and background stars in captured images. To successfully distinguish the target from its neighboring background in the infrared image, we developed a DNN-based target detection method that consists of a ResNest [24] architecture-based FEN, the custom FPN, and the TDN. Using this architecture, it is possible to extract multi-scale robust features of a small infrared target and successfully detect a LEO satellite. Using the position information of the small target in the image, the telescope of the GS aligns with the LEO satellite and sends an uplink beacon signal towards the satellite. Then, the satellite aligns itself precisely towards the GS using the uplink signal. Finally, a robust two-way FSO communication link can be established.

3. Proposed Method

To describe the response of an optical imaging system to a point light source, PSFs are extensively employed in astronomical science. A PSF expresses the intensity distribution of the point source of light affected by the distance and atmospheric conditions. The two PSFs that are most frequently used to model a point target are the 2D Gaussian function and the Moffat function [27]. Based on these functions, we generated synthetic image datasets for DNN model training and performance evaluation. First, the ResNest FEN architecture extracts multi-scale features of the small infrared target from the input images generated using the two PSFs. Then, the multi-scale feature maps at different spatial resolutions are combined using a custom FPN method to obtain robust semantics across all scales. Finally, the TDN containing a classification network in parallel with a bounding box regression network is used to accurately identify small targets in infrared images.

3.1. Small Target Images

Note that the infrared imaging system generally exhibits diffraction during imaging, and thus, the target satellite captured from a GS appears as a diffuse spot [23]. This can be modeled using PSFs, like the 2D Gaussian function and the Moffat function. Using both of these PSFs, we constructed infrared image datasets that include a small infrared target in every image.

3.1.1. 2D Gaussian Function

A 2D Gaussian function is often utilized to approximate the PSF because it provides a mathematical representation that closely resembles the actual blurring effects in astronomical imaging. Because this blurring effect in the image has a symmetrical nature, a small target based on the 2D Gaussian function can be described as follows:

$$T_G(x, y) = T_{max} \exp\left(-\frac{1}{2}\left[\frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2}\right]\right), \quad (1)$$

where $T_G(x, y)$ represents the pixel intensity of the infrared small target, and T_{max} represents the maximum intensity value. σ_x and σ_y represent horizontal and vertical diffusion parameters, respectively, while (x_0, y_0) represents the target's central location and (x, y) are pixel co-ordinates in the image. Note that the horizontal and vertical diffusion parameters are related to the spread shape of the PSF.

3.1.2. Moffat Function

When the light signal from the LEO satellite propagates through the atmosphere, random fluctuation in the refractive index is induced. This is caused by atmospheric turbulence, and it distorts the incoming light signal in the SWIR image [28]. The PSFs produced by this distortion, also referred to as the "seeing" effect, may exhibit elongated irregular shapes [28]. The Gaussian function assumes a symmetric and smooth PSF, which fails to capture the non-uniformity and asymmetry induced by atmospheric turbulence. The Moffat function [27], however, provides a more precise approximation of the PSF in the presence of atmospheric turbulence. The small target based on the Moffat function can be described as follows:

$$T_M(x, y) = \frac{T_{max}}{\left(1+\frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2}\right)^{\beta}}, \quad (2)$$

where $T_M(x, y)$ represents the pixel intensity of the infrared small target, and β represents the shape parameter where the value ranges from 0 to 50. Diffusion parameters, σ_x and σ_y, regulate the Moffat function's size and width along the horizontal and vertical axes, respectively. β describes the variation in the shape of the small infrared target.

3.1.3. Synthetic Image Generation

Note that LEO satellite imagery is very expensive, and the cost of acquiring and processing the data is prohibitive. Therefore, we generated synthetic infrared images with the point target based on the 2D Gaussian and Moffat functions. The infrared images captured by the SWIR camera may contain background noise caused by light pollution, background stars, and image sensor noises such as thermal noise, shot noise, and dark current noise [4,11,14]. In addition, the channel conditions, such as clouds, fogs, and atmospheric turbulence, affects the image quality. Hence, we applied random diffusion parameters in the PSFs, and added random noise to the synthetic infrared images, similar to [23], considering these impacts on the detection performance.

Based on the PSF functions, we generated infrared images with a small target of 3×3 pixels. The $3 \times$ — pixel area provides adequate spatial area for capturing intensity fluctuations adjacent to the center pixel of the target. This pixel area contains sufficient intensity differences between the small infrared target and its neighboring background. Figure 2 shows an example of the generated small target images including a small target using the 2D Gaussian function. It can be seen from Figure 2a,b that the generated infrared image contains a small point target and background noise, and it is difficult to differentiate between them. Figure 2c shows the enlarged 3×3 target area, which has the maximum radiation intensity at its center and intensity fluctuations adjacent to the center pixel of

the target. Note that the contrast between the infrared target and the surrounding background noise is not large because the captured intensity of infrared radiation is inversely proportional to the square of the imaging distance. Hence, DNN-based target detection that can accurately differentiate between the small infrared target and background noise is needed to locate the satellites in the captured images. Conventional methods, such as local-contrast-based and low-rank-based methods, rely on manually created features and predetermined hyperparameters that are typically only effective against a particular background environment [14]. Unlike conventional methods that highly rely on manually created features, the DNN model itself learns to recognize the most relevant features for discriminating between small targets and the background from the input data.

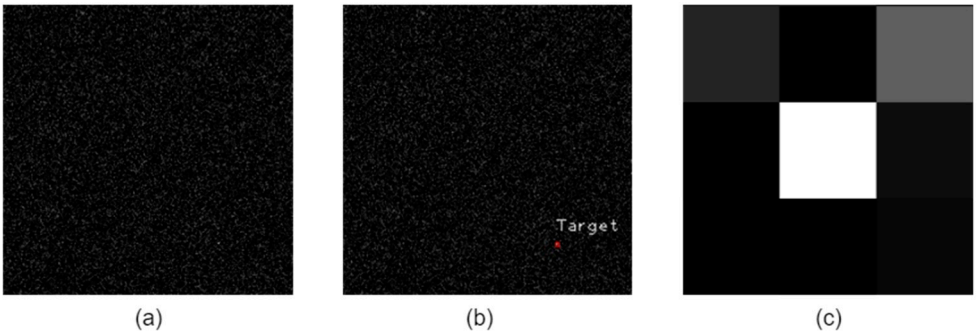

Figure 2. Synthetic infrared images including a small target considering the 2D Gaussian PSF: (**a**) synthetic infrared image; (**b**) the infrared image highlighting the 3 × 3 target area; (**c**) the enlarged 3 × 3 small infrared target area.

3.2. Target Detection Model

The DNN-based target detection model in this paper is divided into three parts, the FEN, the FPN, and the TDN, as illustrated in Figure 3. The input infrared images containing the point targets are fed to the modified ResNestX-based FEN to obtain feature maps at different spatial resolutions. Then, the feature maps with different resolutions obtained from the FEN are fused to generate a custom FPN. The custom FPN concatenates the multi-scale feature map representation obtained from the FEN to improve the proposed DNN's ability to distinguish small infrared targets from the background. Finally, the TDN, which consists of a classification subnet and box regression subnet, performs binary classification and bounding box regression to determine the position of small infrared targets.

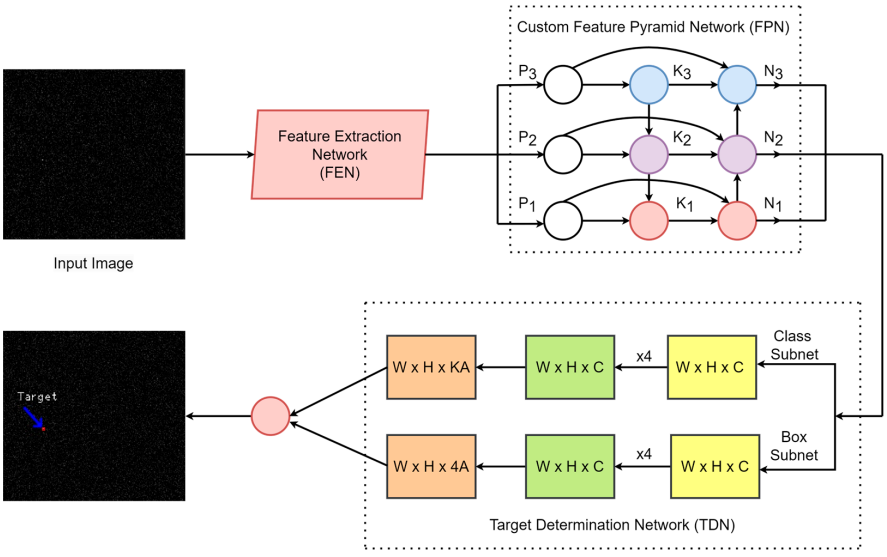

Figure 3. Overall network architecture containing FEN, custom FPN, and TDN for small infrared target detection.

3.2.1. FEN

For robust feature extraction of small point targets from generated infrared images, the ResNest101 model, which comprises a feature map attention mechanism and multi-path propagation, is used. ResNest is a variant of the ResNet architecture [29] that uses a multi-path propagation and a split-attention mechanism, splitting the input feature maps into groups and applying the attention mechanism to each group. Multi-path propagation in ResNest propagates information along multiple paths or branches within the network. ResNest generates several parallel pathways, each of which consists of a series of residual blocks, as opposed to a single pathway. These parallel pathways allow the network to learn diverse and complementary representations of the input data. This is especially effective for small target identification because it enables a DNN network to capture fine-grained characteristics and subtle patterns from multiple perspectives that may indicate the presence of a small target in an infrared image. The split-attention technique improves the ability to learn fine-grained characteristics of the target through capturing feature map interdependencies. The split attention method separates the input feature maps into many groups or branches, each of which pays attention to a different region of the feature maps. Hence, split attention enables the model to concentrate on particular spatial regions or the feature map channels that are particularly informative. In small target detection, the split-attention mechanism improves network performance via effectively suppressing unnecessary background noise and emphasizing the crucial characteristics corresponding to small targets.

ResNest101 is a specific variant of the ResNest architecture that uses 101 layers organized into several convolutional stages, i.e., C_1, C_2, C_3, C_4, and C_5, where C_i has 2^i less resolution than C_{i-1} [23]. In ResNest, C_1, consists of a simple convolutional layer, a batch normalization (BN) layer, and a rectified linear unit (ReLU) activation layer. Each stage after C_1, i.e., C_2, C_3, C_4, and C_5, consists of several residual blocks that contain multiple convolutional layers, BN layers, and ReLU activation layers. In DNN-based target detection methods, a set of predetermined reference bounding boxes known as anchor boxes are generated to compute the percentage of overlap region with the ground truth (GT) bounding box of the target. The overlap value between the anchor boxes and the GT bounding box of the target is then utilized to predict the location of the target in the image. Note that a high computational cost is required to generate anchor boxes in high-resolution feature maps obtained from the C_1 stage for predicting the location of medium and large targets of varied shapes and sizes from datasets such as COCO [30] and ImageNet [31]. Hence, they are not used in the FPN for medium and large object detection. For a small target, however, the shape and size of small infrared targets do not vary greatly, so the number of predefined anchor boxes needed to locate the target in the C_1 convolutional stage is significantly reduced [23]. In addition, the small infrared target can disappear in the latter convolutional stages, i.e., C_3, C_4, and C_5, due to excessive down-sampling. So, high-resolution features obtained from C_1 and C_2 are crucial for extracting meaningful information such as a target's size and shape for small target detection with high accuracy.

The C_1 convolutional stage in conventional ResNest101 only utilizes a simple convolution operation to learn the discriminative features of the small infrared target. It does not contain any residual blocks that capture more discriminative features of the small infrared target to improve detection accuracy. To overcome this limitation and to use additional residual blocks for small infrared target detection, the down-sampling process is not applied after the C_1 convolutional stage. As a result, the feature maps from C_2 have the same dimension as the feature maps obtained from the output of C_1. Finally, the feature maps obtained from C_2 are utilized as the initial high-resolution input for the FPN instead of those from C_1, which allows for the incorporation of additional residual blocks to capture more discriminative features of the small infrared target.

In addition, we conducted ablation experiments to modify the conventional ResNest101-based FEN through adding or removing intermediate layers of specific convolutional stages. The resulting FEN are named ResNestX, where X = 23, 32, 89, and 92 and represents the

number of layers in the FEN network after modification. Through this process, we can evaluate their impact on the small target detection model's performance and find the optimal combination of convolutional stages that obtains high detection accuracy. For example, when using the combination of the C_2, C_3, and C_5 convolutional stages as inputs for the FPN in the modified ResNest-based FEN, we remove the C_4 convolutional stage entirely. The resulting FEN consisting of the C_2, C_3, and C_5 convolutional stages is known as ResNest32 since it only contains a total of 32 convolution layers after removing the intermediate layers belonging to C_4.

To prevent the target from disappearing or merging with the background in the latter convolutional stages due to the excessive down-sampling of infrared images in conventional ResNest architecture, the maximum down-sampling of input infrared images in the FEN process is limited to three. Down-sampling the infrared images less than three times may result in only high-resolution feature maps that lack the semantic information needed to improve target detection accuracy. However, down-sampling the infrared images more than three times may result in a large number of low-resolution feature maps that lack the differentiating characteristics of the small infrared target. As a result, restricting down-sampling in the network to three times attempts to achieve an optimal balance between preserving the differentiating characteristics of the small infrared target and capturing semantic information for small target detection with high accuracy.

The multi-scale feature maps obtained from the modified ResNestX based FEN are denoted P_1, P_2, and P_3, respectively, as shown in Table 1. Based upon the obtained multi-scale feature maps, we devised our anchor box configuration for detecting the 3×3 target as shown in Table 2. Since the target shapes and sizes do not change, the anchor size, ratio, and scale are set to 3, 1, and 1, respectively. The obtained outputs from the modified ResNest101 based FEN are used as inputs for the FPN to create multi-scale feature maps that contain the differentiating characteristics of the small targets as well as semantic information.

Table 1. Spatial resolution of feature maps from modified the ResNestX-based FEN.

Layers	Input Map Size (Pixels)	Output Map Size (Pixels)	Down-Sampling Stride (Pixels)
P_1	256×256	128×128	2
P_2	128×128	64×64	2
P_3	64×64	32×32	2

Table 2. Anchor configuration for small infrared target detection.

Layers	P_1	P_2	P_3
Size [1]	3	3	3
Ratio	1		
Scale	1		

[1] Size is the width of the anchor box in pixels.

3.2.2. FPN

The main purpose of using a custom FPN is to create multi-scale feature maps through combining the important differentiating characteristics of small infrared targets including pixel intensity and edges found in high-resolution feature maps, i.e., P_1 and P_2, with semantic information obtained from the low-resolution feature maps, i.e., P_3. The custom FPN improves the target detection model's understanding of the overall scene through combining the higher-resolution feature maps with lower-resolution feature maps. This process enhances the model's capability to distinguish real targets from false positives, leading to higher detection accuracy [23]. Generally, the FPN structure uses a top-down

pathway and lateral connections to combine high-resolution feature maps, where semantical information is weak, with low-resolution feature maps, which have semantically strong features [25].

To create a custom FPN in the proposed scheme, we use a bottom-up pathway in addition to the top-down pathway to combine multi-scale feature maps, as was implemented in PANET [26]. In addition, we use a skip connection from the original input to the output node if they are at the same spatial resolution level. In the context of small infrared target detection, the additional bottom-up pathway facilitates the flow of the differentiating characteristics of the point target, such as pixel intensity, edges, textures, etc., from high-resolution feature maps to low-resolution feature maps. In addition, we use a skip connection from the original input to the output node if they are at the same spatial resolution level. The skip connection used in the FPN enables the concatenation of more features without adding much computation cost and preserves the original information obtained from the three outputs of FEN.

As shown in Figure 4, the top-down pathway in the custom FPN begins with the lowest-resolution feature maps, i.e., P_3 obtained from the modified ResNestX architecture. At first, the feature maps obtained from P_3 go through a 1×1 convolutional operation to ensure feature maps have the same number of initial channels, i.e., 256 channels. Then, the spatial resolution of the semantically stronger feature maps obtained from P_3 is up-sampled by a factor of 2 using 2D up-sampling. The higher resolution feature maps obtained from P_2 also go through a 1×1 convolutional operation to lower the channel dimensions. Then, the feature maps of P_2 are then combined with the up-sampled feature maps from P_3 via element-wise addition. Similarly, we combine the feature maps of P_1 and P_2 in the same top-down pathway. In the top-down pathway, the output feature maps produced by P_1, P_2, and P_3 are designated K_1, K_2, and K_3, respectively.

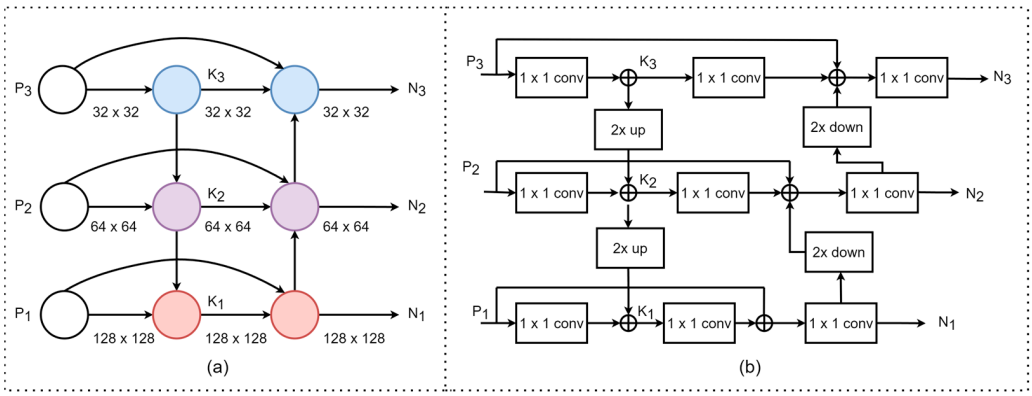

Figure 4. Network architecture of custom FPN for small infrared target detection. (**a**) Basic design of custom FPN. (**b**) Detailed architecture of custom FPN.

The bottom-up pathway begins with the highest-resolution feature maps obtained from the top-down pathway, i.e., K_1. At first, the feature maps obtained from K_1 go through a 1×1 convolutional operation. Then, they are concatenated with the feature maps obtained from the P_1 block of the modified ResNestX network. This process enables the concatenation of more features without significantly increasing the computation cost, along with preserving the original information obtained from the modified ResNestX network. The resulting feature maps again go through a 1×1 convolutional operation. The corresponding output obtained from feature map concatenation is denoted N_1. In the second step, the feature maps obtained from K_2 also go through a 1×1 convolutional operation. Then, they are concatenated with P_2 and the down-sampled N_1 to obtain N_2. The process is repeated until we obtain the feature maps from the lowest resolution layer, i.e., N_3. Through this process, we can combine low-level information obtained from high-resolution feature maps with low-resolution feature maps containing high-level information. In small target detection, this process combines the differentiating

characteristics of the point target obtained from high-resolution feature maps with the semantic information obtained from low-resolution feature maps to differentiate the point target from its neighboring background.

3.2.3. TDN

The output feature maps from the custom FPN, which are denoted N_1, N_2, and N_3, are fed into the TDN to predict the small infrared target location in the infrared images. The TDN is divided into two subnet heads, classification and box regression, as illustrated in Figure 5. Four 3×3 convolutional operations with C = 256 channel filters are applied in the classification subnet. Each convolutional operation is followed by ReLU activations. The 3×3 kernel size helps the network to capture the context of neighboring pixels while generating predictions, which could improve prediction accuracy. Then, another 3×3 convolutional operation is added in the classification subnet to output KA channel filters. The parameter "K" represents the number of target classes, and "A" represents the number of anchors per spatial location. In this study, we set K = 1, A = 1 because we consider a single target class and one anchor generated per spatial location. Finally, a sigmoid activation function is used at the end of the classification subnet to perform binary classification per spatial location.

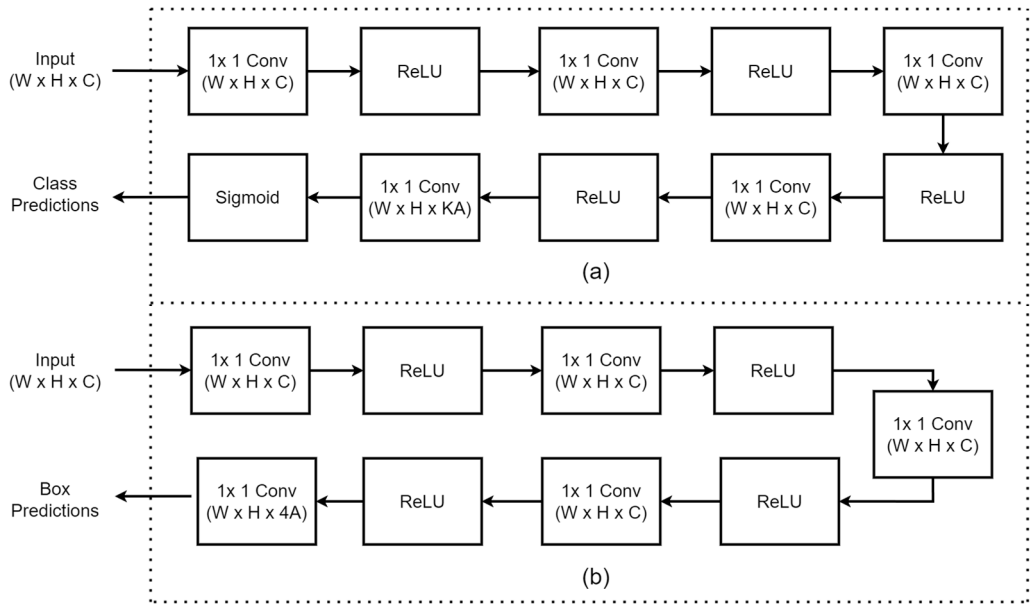

Figure 5. Network architecture of TDN containing classification subnet and box regression subnet for small infrared target detection. (**a**) Detailed architecture of classification subnet. (**b**) Detailed architecture of box regression subnet.

As shown in Figure 5, the box regression subnet has a structure that is identical to the classification subnet, with the exception that it has 4A linear outputs per spatial location. The four linear outputs of the box regression subnet can accurately predict the bounding box offsets for each anchor box per spatial location. Finally, three detection results for three outputs of the custom FPN, i.e., N_1, N_2, and N_3, are generated via the combination of classification and box regression subnetworks. The output contains the predicted target label, its location information in the input infrared image, and a prediction score. The output with the highest prediction score is chosen as the final target detection result.

4. Experiments and Results

In this section, we verify the effectiveness of our proposed small target detection model in terms of true detection rate and average precision (AP). Due to the lack of publicly available datasets with SWIR images of LEO satellites, synthetic datasets created with the

two PSFs, i.e., 2D Gaussian function and Moffat function, were used in the experiments. All experiments were conducted on PyTorch 1.12.1 in Python on a Windows 10 operating system with an Nvidia GeForce RTX 3080 Ti GPU with 64 GB memory.

4.1. Data Generation

The light source emitted by the target satellite appears as a diffused spot when captured from a SWIR camera in the GS. This phenomenon is modeled using the PSF. So, we built synthetic datasets (Data 1 and Data 2) using the 2D Gaussian function and the Moffat function to generate the small infrared target in the captured images, respectively. The target size for the synthetic datasets generated by both functions is considered to be 3×3 pixels.

The infrared images captured by the SWIR camera may contain background noise caused by light pollution, background stars, and image sensor noises such as thermal noise, shot noise, and dark current noise. In addition, the channel conditions, such as clouds, fogs, and atmospheric turbulence, affect the image quality. To present the effect of background stars, channel condition, and the sensor noise, we added random Gaussian noise to the synthetic infrared images, as performed in [23]. Because the contrast between the point target and background noise is small owing to the long-distance imaging, we applied the local SCR strategy to both Data 1 and Data 2 while adding random noise. SCR is a popular metric for determining the difference between the target and background and is frequently used to calculate detection difficulty. In general, a lower SCR indicates it is more challenging to find the infrared target in an image [32]. In this study, we set the local SCR value between 1.0 and 1.5 to reduce the contrast difference between the small target and the background in both Data 1 and Data 2 through adding random Gaussian noise. The formula for local SCR [15] is given as follows:

$$\mathrm{SCR} = \frac{\mu(\mathrm{T}) - \mu(\mathrm{B})}{\sigma(\mathrm{B})}, \tag{3}$$

where $\mu(\mathrm{T})$ is the mean pixel value of the target region, $\mu(\mathrm{B})$ is the mean pixel value of the local background region, and $\sigma(\mathrm{B})$ is the standard deviation of the background region. In our study, a 3×3 area was chosen as the target region and a 15×15 neighboring region around the target area was considered as the background region.

Data 1 using the Gaussian function was generated in the same manner used in [23] for performance comparison. While creating Data 2 using the Moffat function, we generated random values for σ_x, σ_y, and β from a uniform distribution within specific ranges. In our experiment, we used a range of 0.4 to 0.8 to randomly generate horizontal and vertical diffusion parameters, σ_x and σ_y. Similarly, β was also generated randomly from a uniform distribution having a 0.5 range between 2.5 and 3.0. The generated Moffat functions will have different spreads in the x and y directions if random diffusion parameters are used. Using these parameters, the point target in the produced image presents various amounts of sharpness or smoothness. The details of the datasets generated using the 2D Gaussian function and the Moffat function are provided in Table 3.

As shown in Table 3, in this experiment, we generated 8450 images for both Data 1 and Data 2, with each image consisting of a single infrared target at a random position, and an SCR between 1.0 and 1.5. The sample images for Data 1 and Data 2 are highlighted in Figure 6, where the images with SCR < 1.3 have lower contrast compared to image with SCR > 1.3. Hence, target detection in images with SCR < 1.3 is more challenging than in infrared images with SCR > 1.3.

Table 3. Data configuration for the small infrared target from two PSFs.

Dataset	Data Partition	Number of Images	SCR Ratio
Data 1 (Gaussian)	Training data	5070	1.0 < SCR < 1.3: 41.99% 1.3 < SCR < 1.5: 58.00%
	Test data	3380	1.0 < SCR < 1.3: 41.00% 1.3 < SCR < 1.5: 58.99%
Data 2 (Moffat)	Training data	5070	1.0 < SCR < 1.3: 41.99% 1.3 < SCR < 1.5: 58.00%
	Test data	3380	1.0 < SCR < 1.3: 41.00% 1.3 < SCR < 1.5: 58.99%

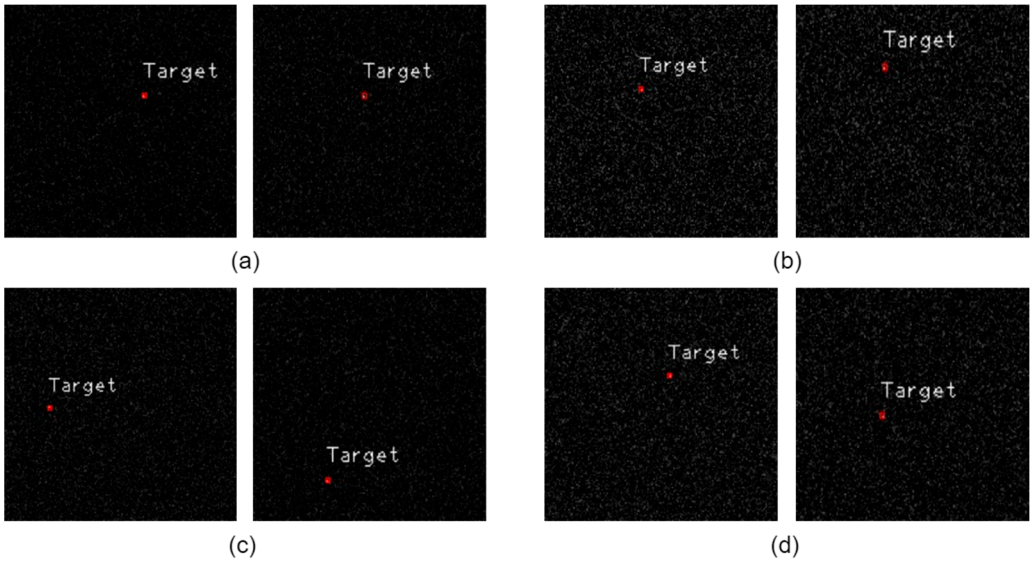

Figure 6. Sample infrared small target images generated using PSFs with varying SCR: (**a**) sample images from Data 1 with SCR < 1.3; (**b**) sample images from Data 1 with SCR > 1.3; (**c**) sample images from Data 2 with SCR < 1.3; (**d**) sample images from Data 2 with SCR > 1.3.

4.2. Target Detection Condition

As mentioned earlier, the small infrared target detection model is different than the conventional target detection models for medium and large objects owing to the different down-sampling process. In addition, the intersection-over-union (IOU) method that is frequently utilized during the training of conventional target detection models to classify medium and large targets cannot be utilized for small infrared targets [33]. Hence, during the training stage of the target detection model, the predicted bounding box is considered a positive sample if the centroid of the predicted target box is located within the area of the GT; otherwise, it is referred to as a negative sample [23].

Generally, IOU and anchor boxes are employed to determine positive and negative samples in target detection networks. IOU calculates the degree of overlap between a predicted bounding box obtained from predefined anchor box generation and the actual GT bounding box of an object. An anchor is often identified as a positive sample when IOU $\geq$ 0.5 or 0.7 [33]. If the IOU method for normal objects is used to classify positive and negative samples, the point target will be classified as both a positive sample and negative sample, as shown in Figure 7. This problem is known as the sample misjudgment problem [33]. When IOU is 0.286, the anchor is considered a negative sample even if it contains the point target. When IOU is 0.5, it is considered a positive sample if a threshold of 0.5 is used. So, IOU for normal objects is not a reliable metric to classify small infrared targets. To avoid the sample misjudgment problem, we instead use the anchor centroid to determine if a sample is positive, rather than using IOU during model training. We consider a sample to be positive if the centroid of the anchor box is within the GT; otherwise, we

consider it a negative sample. Using this strategy, both the sample boxes in Figure 7 are regarded as positive samples because they both contain the point target, and their centroid belongs within the GT.

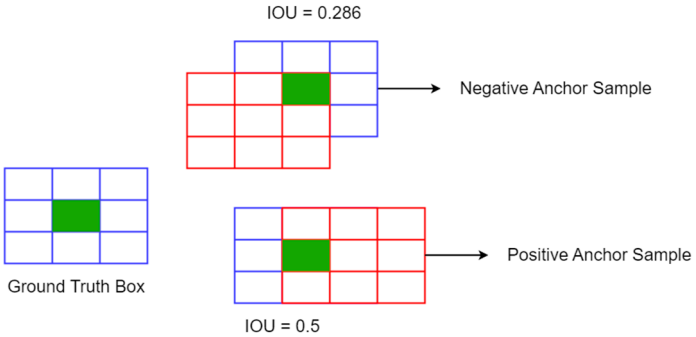

Figure 7. IOU-based evaluation for the small infrared target. The 3 × 3 blue bounding box represents the GT bounding box. The 3 × 3 red anchor on the top right-hand side is labeled a negative anchor sample because its IOU is less than 0.5, whereas the 3 × 3 red anchor on the bottom right-hand side is labeled as positive anchor sample as its IOU is equal to 0.5.

4.3. Target Detection Model Parameters

We describe the model parameters of the target detection model in Table 4. Out of the 8450 images generated for Data 1 and Data 2, 5070 were used for training and 3380 were used to evaluate model performance. Additionally, during the training stage, the data were further divided into another training dataset and a validation dataset. A validation dataset evaluates the target detection model's performance, tracks its generalization ability, and optimizes its hyperparameters. Of the 5070 images, 4000 were used for training the proposed target detection model and 1070 were used to evaluate performance and optimize its hyperparameters during training stage. Other parameters such as batch size, learning rate, and epochs were set to 8, 10^{-5}, and 15, respectively.

Table 4. Details of model parameters used in target detection model for small infrared targets.

Model Parameters	Value
Training data	4000
Validation data	1070
Test data	3380
Learning rate	10^{-5}
Batch size	8
Epochs	15

Since background regions are much larger than the target region in the image, there is an unequal distribution of target and background samples [20]. This uneven distribution is known as a class imbalance problem in target detection [20]. This problem might result in biased learning, where the target detection model focuses more on the background than on the targets. This biased learning mechanism might result in poor target detection performance because the model might not be able to learn the distinctive features of the target effectively enough to classify them with high accuracy. Hence, to address class imbalance in the target detection network, focal loss [20] was used as the loss function for classifying small infrared targets. During the training phase, focal loss helps to alleviate the impact of class imbalance through allocating larger weights to misclassified samples, particularly those from the target class [23]. This is achieved through making the target detection model pay more attention to small infrared targets via increasing the loss value while making predictions for them. In addition, Smooth L1 loss [20], which combines

the mean squared error (MSE) and mean absolute error (MAE) loss functions, is used to calculate the difference between the predicted bounding box and GT bounding box co-ordinates. Smooth L1 loss minimizes the loss obtained from bounding box regression between the predicted target box and the GT box [23].

4.4. Performance Evaluation Metrics

In this section, we discuss the evaluation metrics used to determine the performance of the proposed target detection network. The final output of the TDN determines the location of the predicted infrared target in an image and its corresponding prediction score. Depending upon the location co-ordinates and predicted score, we determine the missed detection rate, true detection rate, and false detection rate. The model is determined to have missed the target in an image if the predicted score is less than 0.5. If the predicted score exceeds 0.5, the detected result is classified as a target or a false alarm. Further, if the detected result's center is within the GT area, the detected result is the target; otherwise, it is a false alarm. The formulas for the missed detection rate (4), true detection rate (5), and false detection rate (6) are given as follows:

$$P_{md} = \frac{Num_{miss}}{Num_{all}}, \tag{4}$$

$$P_{d} = \frac{Num_{true}}{Num_{all}}, \tag{5}$$

$$P_{f} = \frac{Num_{false}}{Num_{all}}, \tag{6}$$

where Num_{miss} is the number of missed detections, Num_{true} is the total number of true detections, and Num_{false} is the total number of false detections. The summation of P_{md}, P_d, and P_f is 1. Ideally, the detection ability is better when P_d is high, while P_{md} and P_f are low. In addition, we used average precision to evaluate the accuracy of our model. In target detection networks, AP is a frequently employed evaluation metric that quantifies the accuracy of object detection algorithms. It produces a single numerical value that represents the target detection model's precision–recall trade-off. While calculating the AP, we use the standard IOU threshold of 0.5. As seen in Figure 7, with an IOU threshold of 0.5, the predicted bounding box closely aligns with the actual ground truth bounding box, resulting in a highly precise localization of the small infrared target. The centroid-based strategy can be considered an effective strategy for target detection if good detection results can be generated when using it during the training stage and when using the standard IOU in the testing stage [33].

4.5. Ablation Study

In this section, the effect of using different combinations of convolutional stages in the ResNest-based FEN is analyzed using Data 1, and the optimal convolutional stages that guarantee high detection accuracy are investigated. Finally, these optimal convolutional stages are utilized for a performance comparison using Data 2 with other DNN-based methods. Note that the main difference between the modified ResNestX, where X = 23, 32, 89, and 92, in our paper and the conventional ResNest101 model is that the input image was not down-sampled more than three times in our experiment to prevent excessive down-sampling. For example, when using C_2, C_3, and C_5, we removed C_4 entirely from the network. In addition, we did not down-sample after C_1, so that the feature maps from C_2 had the same dimension as the feature maps obtained from the output of C_1. Since C_1 and C_2 had the same dimension, the outputs feature maps of C_2 are used as the initial high-resolution output for the FPN instead of C_1 to include additional residual blocks. In addition, we utilized the optimal target-oriented shallow–deep features (TSDF) model [23] in our ablation study for performance comparison. The optimal feature fusion combinations obtained from the TSDF-based DNN [23] for the synthetic dataset is denoted

as FC_1 + FC_2 + FC_1C_3 + FC_2C_3 +FC_1C_4 + FC_2C_4 (feature fusion 1), where FC_iC_j denotes the feature fusion function.

The experimental results from the ablation study of different combinations of convolutional stages are shown in Table 5. The performance metrics for several of those combinations are true detection rate (P_d), false detection rate (P_f), missed detection rate (P_{md}), and AP. Among several candidates, the combination of ResNest92 and FPN achieved a slightly lower P_d and AP of 98.80% and 98.90%, respectively, suggesting a marginally reduced ability to detect the small infrared target compared to the other combinations. Changing the FEN to ResNest89 or ResNest23 while still using an FPN structure for feature fusion achieved slightly better results than the previously mentioned counterpart, with both obtaining a P_d of 99.10%. However, the AP from ResNest89 was comparatively worse than when we used ResNest23 along with FPN. When ResNest23 was used, the AP was found to be 99.20%, which is 0.10% better when ResNest89 was used. Finally, when ResNest32 was combined with FPN, a P_d and AP of 99.40% were achieved, while maintaining P_f at only 0.30%. When compared to the TSDF model [23] that uses ResNet50 as FEN and feature fusion 1 technique for feature combination for small infrared target detection in synthetic datasets using a 2D Gaussian function, all of the ResNest-based target detection models achieved a higher P_d and AP. Therefore, the ablation study illustrated that ResNest32 FEN with a combination of FPN had the highest P_d and AP while having the lowest P_f, indicating that this combination achieved the highest accuracy while minimizing false detections. From this ablation study, we determined that this is the combination that should be used in FEN to accurately detect the small infrared target.

Table 5. Ablation study on Data 1 to find optimal convolutional stages combination.

FEN	Feature Combination	P_d (%)	P_f (%)	P_{md} (%)	AP (%)
ResNest23	Custom FPN	99.10	0.80	0.10	99.20
ResNest32	Custom FPN	99.40	0.30	0.30	99.40
ResNest89	Custom FPN	99.10	0.90	0.00	99.10
ResNest92	Custom FPN	98.80	1.10	0.10	98.90
ResNet50 [23]	Feature fusion 1	98.30	1.70	0.00	95.20

4.6. Performance Comparison

To prove the performance improvements from the proposed method, we compared the best performance obtained from our ablation study to conventional methods such as Var_Diff [15], AAGD [15], LOG [15], LIG [16], NRAM [34], IPI [35], DPIR [36], and GST [37] using Data 1. Then, we utilized Data 2 to compare the proposed DNN method's performance with other DNN-based methods. Comparing several DNN-based approaches aids in analyzing how various network architectures, feature extraction methods, and feature fusion strategies impact small infrared target detection. The performance comparison for Data 1 is shown in Table 6.

When compared with the previous conventional approaches, our model was able to provide accurate localization of the small infrared target while significantly reducing the false detection rate. Compared with conventional handcrafted algorithms such as LIG, NRAM, Var_Diff, and IPI, we achieved decreases of 14.08%, 16.89%, 17.92%, and 20.41%, respectively, in the false detection rate while achieving increases in the true detection rate of 13.78%, 16.59%, 17.62%, and 20.11%, respectively. Compared to DPIR and GST, our proposed DNN method achieved even higher increases in true detection rates: 55.08% and 61.94%, respectively. False detection rates were also drastically reduced by 55.38% and 62.24%, respectively. Conventional techniques rely on simple handcrafted features to manually generate important discriminative information on the small infrared target. However, the features of small infrared targets are difficult to distinguish from the background due to their small size and low SCR when using the handcrafted conventional method.

Hence, they produced the high false detection rates shown in Table 6. In contrast to the conventional methods, the proposed DNN method automatically extracts features from the infrared small target using multiple convolutional stages of modified ResNest architecture without excessive down-sampling. The automatic feature extraction enhances the generalization capability of the proposed DNN in different environmental settings, including low-SCR conditions. In addition, the custom FPN combines the multi-scale feature maps obtained from the modified ResNest architecture to enhance the discriminative capability of the DNN. Hence, our proposed DNN method not only achieved a higher true detection rate but also significantly reduced the false reduction rate.

Table 6. Performance comparison for small target detection using Data 1.

Algorithm	P_d (%)	P_f (%)
Var_Diff [15]	81.78	18.22
AAGD [15]	81.75	18.25
LOG [15]	81.48	18.52
LIG [16]	85.62	14.38
NRAM [34]	82.81	17.19
IPI [35]	79.29	20.71
DPIR [36]	44.32	55.68
GST [37]	37.46	62.54
Proposed	99.40	0.30

The performance comparison of various DNN-based structures using Data 2 is shown in Table 7. Here, we utilized the optimal feature map combinations obtained from [22] with ResNet50 and ResNest101 FEN to compare with the optimal FEN and FPN combination obtained from our ablation study. For the performance comparison, we additionally computed the AP for the DNN-based algorithms, in addition to P_d and P_f. The optimal feature fusion combinations obtained from the TSDF-based DNN [23] are denoted $FC_1 + FC_2 + FC_1C_3 + FC_2C_3 + FC_1C_4 + FC_2C_4$ (feature fusion 1) and $FC_1 + FC_2 + FC_3 + FC_1C_4 + FC_2C_4$ (feature fusion 2), where FC_iC_j denotes the feature fusion function. These feature fusion methods were obtained using a ResNet50 architecture in [23] and were applied to ResNest101 for small target detection using Data 2. The optimal FEN model obtained from our ablation method is ResNest32. When using feature fusion 2 of the conventional ResNest101 FEN, P_d increased by 0.74% while P_f decreased by 0.74%, compared to feature fusion 2 of the ResNet50. Similarly, for the same feature fusion combination, there was a significant rise in AP from 85.10% to 92.16% when the conventional ResNest101 was used instead of ResNet50, suggesting that the ResNest101 FEN network performs better than ResNet50. The combination of ResNest32 FEN and custom FPN achieved better performance with P_d of 94.00%, P_f of 6.00%, and AP of 94.70%. Hence, based upon the experimental results on Data 2, the combination of ResNest32 FEN and custom FPN outperformed other DNN-based methods.

Table 7. Performance comparison for small target detection using Data 2.

FEN	Feature Combination	P_d (%)	P_f (%)	AP
ResNet50 [23]	Feature fusion 1	92.93	7.07	86.88
	Feature fusion 2	92.49	7.52	85.10
ResNest101	Feature fusion 1	88.14	5.36	81.22
	Feature fusion 2	93.23	6.78	92.16
ResNest32	Custom FPN	94.00	6.00	94.70

The superior performance of the proposed small target detection method can be explained by the combination of the modified ResNest architecture and the proposed custom FPN method. ResNest uses an attention mechanism that enables the network to dynamically assign different weights to multiple input features, enhancing its ability to obtain relevant information and suppress irrelevant or noisy signals, compared to the ResNet backbone. In the TSDF-based feature fusion methods [23], the authors only applied a one-to-one concatenation strategy between a shallow layer and a deep layer, rather than integrating all layers from ResNet50 to create multi-scale feature maps. Hence, they contain more limited information than the custom FPN structure used in the proposed DNN-based target detection method. In the custom FPN in our study, information obtained from the modified ResNest-based FEN is shared among all the layers using a top-down, bottom-up pathway and a skip connection before sending it to the TDN. In addition, the ResNet50 architecture used in TSDF [23] is generally used for medium and large object detection and is not specific to the small target. Through conducting an ablation study to modify the structure of ResNest101 specifically for small target detection, the proposed DNN-based small target detection method outperformed the previous DNN-based methods. Figure 8 illustrates the prediction results from Data 1 and Data 2 using our proposed DNN-based target detection model. As shown in Figure 8, our proposed DNN-based target detection model is able to detect the small infrared target in both Data 1 and Data 2. Using the proposed scheme, the GS can successfully detect the position of the LEO satellite in the infrared images, and this information is used to establish a robust FSO channel link between the GS and the target LEO satellite.

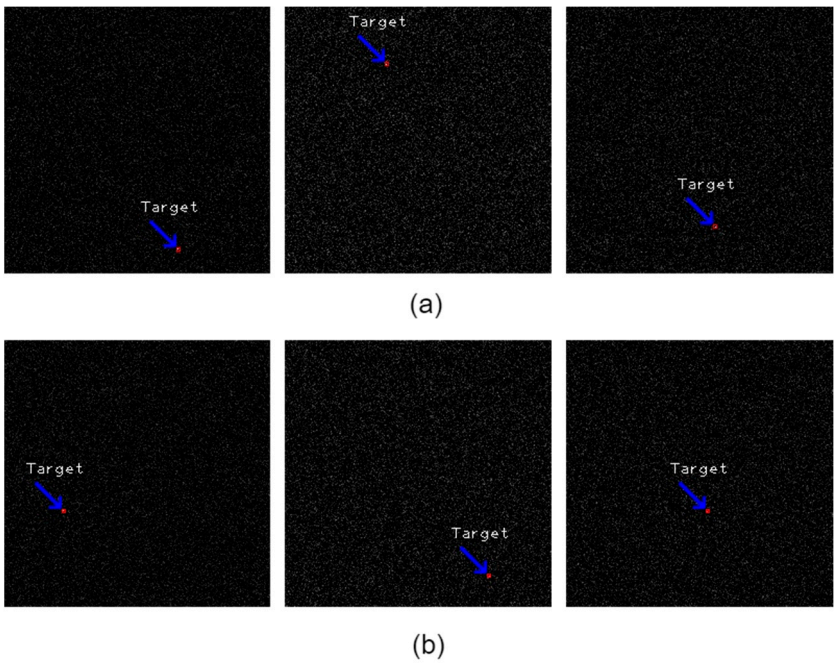

Figure 8. (**a**) Predicted bounding box visualization for small infrared targets in Data 1. (**b**) Predicted bounding box visualization for small infrared targets in Data 2.

5. Discussion

The proposed method surpassed the previous small target detection approaches for synthetic datasets Data 1 and Data 2, achieving P_d of 99.40% and 94.00%, and AP of 99.40% and 94.70%, respectively. This is the result of ResNest32's split attention and multi-path propagation, as well as the multi-scale feature maps obtained from the custom FPN.

In practical scenarios, the LEO satellite needs to be constantly tracked by the SWIR camera. As a logical extension of our study, we will develop a precise and robust tracking solution that integrates convolutional neural networks (CNNs) and transformer networks.

The backbone CNN will be responsible for extracting features from the input infrared images. The transformer network will be responsible for simultaneously detecting and tracking the infrared target using the features obtained from the CNN. To further improve the prediction time for tracking, we will leverage inverted residual networks and linear bottlenecks found in MobileNets [38] in the CNN backbone in our future work.

6. Conclusions

In this paper, we proposed a DNN-based small infrared target detection network that can detect the location of a LEO satellite in an infrared image captured at the GS and can determine the correct alignment of the receiver telescope to establish the FSO channel link. To train the DNN model, we generated synthetic datasets that contain LEO satellites as the small infrared target based upon frequently used PSFs (the Gaussian function and the Moffat function). We designed a modified ResNest-based FEN to extract features from infrared images using the optimal combination of convolutional stages. Then, we utilized a custom FPN network to combine those features and extract the differentiating characteristics of the point target, such as pixel intensity, as well as semantic information to differentiate the point target from its neighboring background. Finally, we detected the LEO satellite, imaged as a small infrared target using a TDN that consists of a classification subnet and a bounding box subnet. Experiment results using two PSF-based datasets proved that our proposed model outperformed existing small target detection schemes and was able to achieve true detection rates of 99.40% and 94.00%, respectively.

Author Contributions: Conceptualization, N.D. and B.W.K.; methodology, B.W.K.; software, N.D.; validation, N.D. and B.W.K.; formal analysis, B.W.K.; investigation, N.D.; resources, N.D.; data curation, B.W.K.; writing—original draft preparation, N.D.; writing—review and editing, N.D. and B.W.K.; visualization, N.D. and B.W.K.; supervision, B.W.K.; project administration, B.W.K.; funding acquisition, B.W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2022R1A2B5B01001543).

Data Availability Statement: The datasets created for this study are available from the corresponding author on reasonable request.

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

References

1. Yahia, O.B.; Erdogan, E.; Kurt, G.K.; Altunbas, I.; Yanikomeroglu, H. A weather-dependent hybrid RF/FSO satellite communication for improved power efficiency. *IEEE Wirel. Commun. Lett.* **2021**, *11*, 573–577. [CrossRef]
2. Le, H.D.; Pham, A.T. On the design of FSO-based satellite systems using incremental redundancy hybrid ARQ protocols with rate adaptation. *IEEE Trans. Veh. Technol.* **2021**, *71*, 463–477. [CrossRef]
3. Maharjan, N.; Devkota, N.; Byung, W. Kim: Atmospheric Effects on Satellite−Ground Free Space Uplink and Downlink Optical Transmissions. *Appl. Sci.* **2022**, *12*, 10944. [CrossRef]
4. Kaushal, H.; Kaddoum, G. Free space optical communication: Challenges and mitigation techniques. *arXiv* **2015**, arXiv:1506.04836.
5. Kaushal, H.; Kaddoum, G. Optical communication in space: Challenges and mitigation techniques. *IEEE Commun. Surv. Tutor.* **2016**, *19*, 57–96. [CrossRef]
6. Mitchell, J. 2022 NASA Optical Communications Update. In Proceedings of the 5th Annual Directed Energy Symposium, National Harbor, MD, USA, 5–6 October 2022.
7. Robinson, B.S.; Boroson, D.M.; Schieler, C.M.; Khatri, F.I.; Guldner, O.; Constatine, S.; Shih, T.; Burnside, J.W.; Bilyeu, B.C.; Hakimi, F. TeraByte InfraRed Delivery (TBIRD): A demonstration of large-volume direct-to-Earth data transfer from low-Earth orbit. In *Free-Space Laser Communication and Atmospheric Propagation XXX: 29–30 January 2018, San Francisco, CA, USA*; SPIE: Bellingham, DC, USA, 2018.
8. Kaushal, H.; Jain, V.K.; Kar, S. Acquisition, tracking, and pointing. In *Free Space Optical Communication*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 119–137.
9. Walsh, S.M.; Karpathakis, S.F.E.; McCann, A.S.; Dix-Matthews, B.P.; Frost, A.M.; Gozzard, D.R.; Gravestock, C.T.; Schediwy, S.W. Demonstration of 100 Gbps coherent free-space optical communications at LEO tracking rates. *Sci. Rep.* **2022**, *12*, 18345. [CrossRef]

10. Ly, D.; Lucken, R.; Giolito, D. Correcting TLEs at epoch: Application to the GPS constellation. *J. Space Saf. Eng.* **2020**, *7*, 302–306. [CrossRef]
11. Marbel, R.; Ben-Moshe, B.; Grinshpoun, T. Pico-Sat to Ground Control: Optimizing Download Link via Laser Communication. *Remote Sens.* **2022**, *14*, 3514. [CrossRef]
12. Tanaka, T.; Kawamura, Y.; Tanaka, T. Development and operations of nano-satellite FITSAT-1 (NIWAKA). *Acta Astronaut.* **2015**, *107*, 112–129. [CrossRef]
13. Gach, J.L.; Boutolleau, D.; Brun, C.; Carmignani, T.; Clop, F.; Feautrier, P.; Lemarchand, S.; Stadler, E.; Wanwanscappel, Y. C-RED 3: A SWIR camera for FSO application. In *Free-Space Laser Communications XXXII*; SPIE: Bellingham, DC, USA, 2020.
14. Hansen, M.P.; Malchow, D.S. Overview of SWIR detectors, cameras, and applications. In *Thermosense XXX*; SPIE: Bellingham, DC, USA, 2008.
15. Moradi, S.; Moallem, P.; Sabahi, M.F. A false-alarm aware methodology to develop robust and efficient multi-scale infrared small target detection algorithm. *Infrared Phys. Technol.* **2018**, *89*, 387–397. [CrossRef]
16. 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. [CrossRef]
17. 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**, *32*, 1745–1758. [CrossRef] [PubMed]
18. 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. [CrossRef]
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
20. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
21. Luo, H.; Wang, P.; Chen, H.; Kowelo, V.P. Small Object Detection Network Based on Feature Information Enhancement. *Comput. Intell. Neurosci.* **2022**, *2022*, 6394823. [CrossRef]
22. Fan, M.; Tian, S.; Liu, K.; Zhao, J.; Li, Y. Infrared small target detection based on region proposal and CNN classifier. *Signal Image Video Process.* **2021**, *15*, 1927–1936. [CrossRef]
23. Du, J.; Lu, H.; Hu, M.; Zhang, L.; Shen, X. CNN-based infrared dim small target detection algorithm using target-oriented shallow-deep features and effective small anchor. *IET Image Process.* **2021**, *15*, 1–15. [CrossRef]
24. Zhang, H.; Wu, C.; Zhang, Z.; Zhu, Y.; Lin, H.; Zhang, Z.; Sun, Y.; He, T.; Mueller, J.; Manmatha, R. ResNeSt: Split-Attention Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022.
25. 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.
26. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
27. Mojžíš, F.; Jaromir, K.; Jan, Š. Point spread functions in identification of astronomical objects from Poisson noised image. *Radioengineering* **2016**, *25*, 169. [CrossRef]
28. Trujillo, I.; Aguerri, J.A.L.; Cepa, J.; Gutiérrez, C.M. The effects of seeing on Sersic profiles—II. The Moffat PSF. *Mon. Not. R. Astron. Soc.* **2001**, *328*, 977–985. [CrossRef]
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
30. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014.
31. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
32. Hsieh, T.-H.; Chou, C.-L.; Lan, Y.-P.; Ting, P.-H.; Lin, C.-T. Fast and robust infrared image small target detection based on the convolution of layered gradient kernel. *IEEE Access* **2021**, *9*, 94889–94900. [CrossRef]
33. Du, J.; Lu, H.; Zhang, L.; Hu, M.; Chen, S.; Deng, Y.; Shen, X.; Zhang, Y. A spatial-temporal feature-based detection framework for infrared dim small target. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 112. [CrossRef]
34. Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; Peng, Z. Infrared small target detection via non-convex rank approximation minimization joint l2,1 norm. *Remote Sens.* **2018**, *10*, 1821. [CrossRef]
35. 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]
36. 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]

37. Gao, C.-Q.; Tian, J.W.; Wang, P. Generalised-structure-tensor-based infrared small target detection. *Electron. Lett.* **2008**, *44*, 1. [CrossRef]
38. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B. Searching for MobileNetV3. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.