

## Research Article

# Heterogeneous Gray-Temperature Fusion-Based Deep Learning Architecture for Far Infrared Small Target Detection

Junhwan Ryu  and Sungho Kim 

*Department of Electronic Engineering, Yeungnam University, Republic of Korea*

Correspondence should be addressed to Sungho Kim; [sunghokim@ynu.ac.kr](mailto:sunghokim@ynu.ac.kr)

Received 1 March 2019; Revised 23 April 2019; Accepted 10 July 2019; Published 22 August 2019

Guest Editor: Sidike Paheding

Copyright © 2019 Junhwan Ryu and Sungho Kim. 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.

This paper proposes the end-to-end detection of a deep network for far infrared small target detection. The problem of detecting small targets has been a subject of research for decades and has been applied mainly in the field of surveillance. Traditional methods focus on filter design for each environment, and several steps are needed to obtain the final detection result. Most of them work well in a given environment but are vulnerable to severe clutter or environmental changes. This paper proposes a novel deep learning-based far infrared small target detection method and a heterogeneous data fusion method to solve the lack of semantic information due to the small target size. Heterogeneous data consists of radiometric temperature data (14-bit) and gray scale data (8-bit), which includes the physical meaning of the target, and compares the effects of the normalization method to fuse heterogeneous data. Experiments were conducted using an infrared small target dataset built directly on the cloud backgrounds. The experimental results showed that there is a significant difference in performance according to the various fusion methods and normalization methods, and the proposed detector showed approximately 20% improvement in average precision (AP) compared to the baseline constant false alarm rate (CFAR) detector.

## 1. Introduction

The problem of the robust detection of small targets is an important issue in surveillance applications, such as infrared search and track (IRST) and infrared (IR) remote sensing. Information about the objects that can be obtained from the image is extremely limited due to the small target size. In particular, targets located on a long distance have a low signal-to-clutter ratio (SCR) and eventually have an adverse effect on the detection performance. In addition, because of the small target size, it is relatively vulnerable to noise of the surrounding environment, such as sun glint, sensor noise, cloud, etc., making it difficult to detect accurately.

The problem of detecting small targets has been directed mainly at using the most suitable filter among the many filters available or to design a new filter. To solve the problem of a fixed filter, which does not reflect the size change according to the movement of the target, studies have been carried out to consider the scale. Moreover, studies have been conducted on using the classifier together with the conventional machine

learning based method. On the other hand, because of the characteristics of the hand-crafted, the small target is confined to a specific environment and severe noise prevents its detection.

This paper proposes a small target detection method based on deep learning capable of end-to-end training. The network structure and training strategy are inspired by the single shot multibox detector (SSD) [1], and the network structure is transformed to a single-scale because it deals only with small targets. The proposed network learned a small target dataset that was constructed directly with the various background clusters. By learning various backgrounds of the sky, this study solved the problem caused by the uncertain heterogeneous background, which was a problem in previous research. This study also compared the result of the fusion of radiometric temperature data by directly constructing raw infrared data as well as gray-scale data that is generally used as the input of a detector network. In addition, the performance was assessed and compared according to the normalization method in heterogeneous data fusion.

The contribution of this paper is summarized as follows.

- (i) A dataset targeting various backgrounds of the sky was constructed for the detection of far infrared small targets. Unlike other research areas, where open datasets exist, there is no dataset to detect and classify far infrared small targets.
- (ii) The dataset constructed in this paper includes infrared raw data. Unlike previous studies that used only intensity-based gray data (8 bits), raw data (14 bits) can be used together. Temperature information is available by applying a radiometric calibration to the raw data. The use of gray-scale and temperature data with physical meaning together as input to the network allows the use of more information and better detection results through fusion.
- (iii) A deep learning-based network for far infrared small target detection that can train and detect from end-to-end beyond conventional hand-crafted method is proposed. Using the proposed network, this study analyzed the effects of pixel-level fusion of gray-scale and radiometric temperature data and the effects of efficient normalization methods for data fusion.

The remainder of this paper is organized as follows. Section 2 briefly introduces previous studies related to the detection and recognition of small targets. Section 3 outlines the proposed method. Section 4 introduces the experimental results and datasets. Finally, Section 5 reports the conclusions.

## 2. Related Works

Object detection is an important research area of computer vision. Among them, the detection of small targets is a challenging problem because of the limited information. The research directions to solve this problem can be classified broadly into the traditional machine learning-based methods and deep learning-based methodologies, in which recent studies will be conducted.

One of the traditional methodologies is the filter-based method [2–9]. First, previous studies [2–5] examined the filter itself. For example, Barnett [3] evaluated a promising spatial filter for point target detection in infrared images and used a median subtraction filter. Schmidt [5] examined a modified matched filter (MMF) composed of a product of a nonlinear operator called an inverse Euclidean distance and a least-mean-square (LMS) filter to suppress cloud clutter. Studies on adaptively improved filters have been conducted [6–8]. Yang et al. [7] proposed a Butterworth high-pass filter (HPF) that can adaptively determine the cut-off frequency. Zhao et al. [9] proposed another method using a filter to fuse the results of several filters with different directions. Other methods [10–15] were based on the contrast mechanism of the human vision system (HVS). Qi et al. [10] were inspired by the attention mechanism to produce a color and direction-based Boolean map to fuse, and Chen et al. [11] proposed a method of obtaining a local contrast map using a new local contrast measure that measures the degree of difference between the current location and neighbors. After that, a target is detected

with an adaptive threshold inspired by the contrast. Han et al. [12] increased the detection rate through size-adaptation preprocessing and calculated the saliency map using the improved local contrast measure, unlike the conventional method using only the contrast. Deng et al. [13] improved the contrast mechanism by the weighted local difference measure, and a method that applies a classifier was proposed [14]. Han et al. [15] proposed a multiscale relative local contrast measure to remove the interference region at each pixel.

Another approach was to solve the size variation problem that occurs when the target moves [16–18]. For example, Kim et al. [17] proposed a Tune-Max of the SCR method to consider the problem of scale and clutter rejection inspired by the HVS. In the predetection step, target candidates maximizing Laplacian-scale space images are extracted and in the final-detection step. The scale parameters were adjusted to find target candidates with the largest SCR value. This method has shown good performance, but it consists of complicated steps.

The following methodologies [19–21] deal with methods for making the best use of features. Dash et al. [19] proposed a feature selection method that can use features efficiently in a classifier rather than directly relating to the problem of detecting a small target. Kim [20] analyzed various target features to determine which feature is useful for detecting small targets and proposed a machine learning-based target classification method. Bi et al. [21] used multiple novel features to solve the problem of many false alarms (FAs) that occur when existing methods consistently use single metrics for complex backgrounds. A total of seven features were used and a method to identify the final target through a classifier was proposed.

A range of machine learning-based methodologies can be used for small target detection [22–32]. Gu et al. [23] proposed a method to apply a constant false alarm rate (CFAR) detector to the target region after suppressing the clutter by predicting the background through a kernel-based nonparametric regression method. Qi et al. [29] proposed a directional saliency-based method based on observations that the background clutter has a local direction and treat it as a salient region-detection problem. The existing methods still raise the problem of not separating the background completely. Zhang et al. [30] used an optimization approach to separate the target from the background.

Over the last few decades, research has been conducted in various directions mentioned above and more studies are being conducted based on deep learning. Liu et al. [33] proposed that training a sample using a signal-to-noise ratio (SNR) with an appropriate constant value helps improve the performance over training with a randomly sampled SNR. The targets were generated and synthesized randomly and were not actual targets. Chen et al. [34] used a synthetic aperture radar (SAR) image and treated it as a convolutional neural network- (CNN-) based classification problem not a detector network. Because there is little data, it adopts a fully convolution structure except for a fully connected layer to prevent overfitting. Generative adversarial networks (GAN), which is not a general CNN-based structure, were proposed

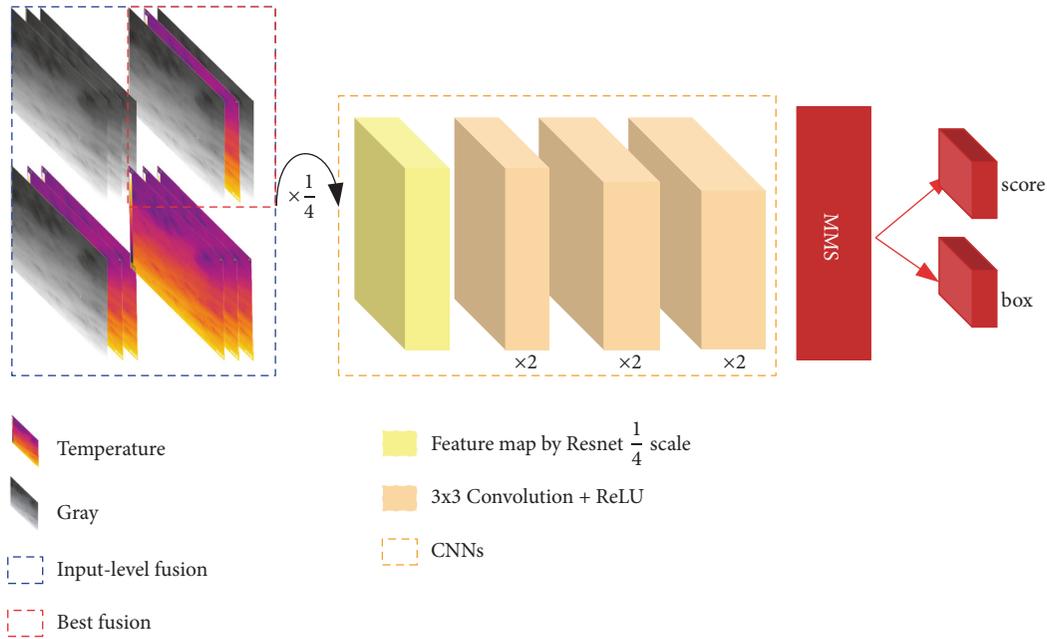


FIGURE 1: Proposed network structure and pixel-level fusion of temperature data.

[35]. The generator trains to transfer the representation of a small object similar to a large object. The discriminator, however, competes with the generator to identify the representation generated by the generator and allows the generator to have a representation that is useful for detection. Hu et al. [36] proposed a way to use the features extracted from other levels of features. Bosquet et al. [37] proposed the problem of a loss of target information as existing detector networks undergo downsampling. After several convolution layers, it assumes that the feature map has sufficient information to determine the area where the target exists and proposes a new concept called the region context network (RCN). In the feature map that passes the shallow convolution, the region with the highest possible likelihood of the target is extracted along with the context to perform a late convolution. The subsequent steps are similar to the general detector network.

Deep learning-based methodologies have been active in many areas in recent years. On the other hand, the problem of detecting small targets has not been actively researched because not only are there no publicly available datasets that can be verified, the information available from the image is limited and it is difficult to produce a situation where a dataset can be constructed.

### 3. Proposed Method

This section introduces the proposed network structure for the detection and fusion of small targets in the far-infrared region and compares the intensity-based gray-scale data with the radiometric temperature data obtained from the constructed data. This section also introduces the normalization method to fuse heterogeneous data.

*Proposed Network Architecture.* The proposed network was inspired by the SSD and uses a single-scale feature structure instead of a multiscale feature structure, which is an advantage of SSD because only small targets of up to 20 pixels are handled. The blue dashed line in Figure 1 represents input data and four cases where pixel-level fusion is possible. In addition, the first feature map is a feature map that passes through Resnet-34 [38], the base network. Subsequently, it goes through six convolution layers, and the detection result is obtained by removing redundant detection through the non-maximum suppression (NMS) in the last feature map. In Figure 1, x2 represents two convolutional blocks, so there are six convolutional layers in total. To minimize the loss of information, Resnet-34 was used up to a  $\frac{1}{4}$  scale. Bounding box regression and score prediction for obtaining the final detection results have the same structure as the general object detection network but the NMS standard is somewhat relaxed because of the small target size. For training, the learning rate is set to 0.0001 and is a fully convolutional structure consisting only of a 3x3 convolution layer. The optimization method uses Adam optimizer [39], and He initialization [40] is used.

*Comparison of Fusion Methods.* The blue dotted box in Figure 1 shows the pixel-level fusion method for the fusion of an intensity-based gray-scale and radiometric temperature data. The gray-scale data has one channel and the radiometric temperature data is also made up of one channel, so the heterogeneous data can be concatenated in channel direction. Another common method is to try the feature-level fusion method. Hou et al. [41] used the late-fusion method and it is a method of concatenating feature maps using RGB and gray-scale data as inputs to different networks with the same structure. On the other hand, this paper used the

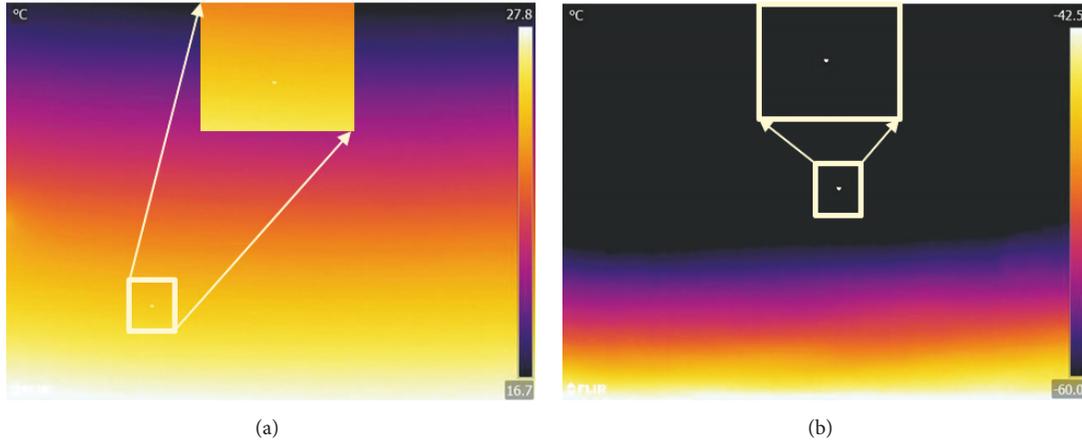


FIGURE 2: Comparison of the seasonally distorted target temperature data. (a) Midsummer in August, and (b) a midwinter in February.

pixel-level fusion method because the feature-level fusion method has not been detected properly. In addition to the pixel-level fusion method, which proposes a range of combinations based on three channels, there is also a method of accumulating three gray scale data, such as RGB and one radiometric temperature data, for a total of four channels. The pretrained deep network cannot be used when this fusion method is applied. Therefore, this paper compares several fusion methods that can fuse heterogeneous data with three channels. Proper normalization methods are required because gray scale data (8-bit) and radiometric temperature data (14-bit) with different ranges of values must be fused together at the pixel-level.

*Thermal Normalization.* Radiometric temperature data should be normalized. Kim. [42] dealt with temperature data for the problem of detecting pedestrians. At this time, a normalization method was used assuming a maximum temperature of 40°C due to human thermoregulation. On the other hand, the radiometric temperature data was distorted because the experimental environment of this paper dealt only with distant small targets. As a result, even in the same sky, as shown in Figure 2, there is a significant temperature deviation in the air according to the season. The temperature difference between the target and the surrounding air is not large at mid-summer (August, Figure 2(a)), whereas the difference is 20°C or more at mid-winter (February, Figure 2(b)).

At this time, both targets in Figures 2(a) and 2(b) were located in the same sky background and distance with different seasons. Owing to the distorted temperature data, the temperature of the target does not have a constant range. Therefore, the normalization method in the methods reported elsewhere cannot be used and normalized [42], as expressed in (1), to have a value of a specific range. The following were used to compare the results according to the various normalization methods: a normalization method with a specific range of  $-1$  to  $0$ ,  $-1$  to  $1$ , and  $0$  to  $1$ ; a method of normalizing the mean and standard deviation to  $0.5$ ; and a precalculated mean and standard deviation of large

scale data.  $I_n(x, y)$  is the input data and  $\min(I)$ ,  $\max(I)$  are the minimum and maximum values, respectively, for the entire input data, and subscripts  $x, y$  mean each pixel. The abbreviations,  $ub$  and  $lb$ , represent the upper and lower bounds of the normalization range, respectively. This makes  $lb$  when each pixel of the input data  $I$  is the min value, and  $ub$  when it is the max value, and the rest has a value between them. For example, if the input data should be normalized between  $-1$  and  $1$ , set  $ub$  to  $1$  and  $lb$  to  $-1$ .

$$I_n(x, y) = \frac{ub - lb}{\max(I(x, y)) - \min(I(x, y))} \times (I(x, y) - \min(I(x, y)) + lb) \quad (1)$$

## 4. Experimental Results

This section introduces the infrared small target dataset, augmentation method for training, comparison of the results with existing research directions, and various experiments.

### 4.1. Yeungnam University (YU) FIR Small Targets Dataset

*Hardware Specifications.* The FLIR T620 model in Figure 3(a) was the thermal imaging camera equipment used to build infrared small target data. FLIR T620 has a spatial resolution of  $640 \times 480$  and a temperature range between  $-40^\circ\text{C}$  and  $650^\circ\text{C}$  and stores data at 14 bits operating at 30 frames per second (FPS). Figure 3(b) presents small drones that serve as simulated targets and use the DJI's PHANTOM 4 PRO model. The model was 1.38kg, including the battery and propeller, and the size was not provided separately and was approximately  $30 \times 30 \times 16$  (cm) when measured directly. The maximum flight time was approximately 30 minutes.

*Experimental Environment and Data Acquisition.* Experiments were conducted on a specific location, and Figure 3(c) shows the flight record by Google Earth®. The yellow line indicates the accumulated path that the actual target has flown. The target was experimented in various directions and elevation angles at specific locations. When constructing



FIGURE 3: Equipment used in the experiment and experimental environment. (a) is a FLIR T620 infrared camera, (b) is a PHANTOM 4 PRO drone model used as a target, and (c) is the flight trajectory of a target.

data under these circumstances, if all sequences are used, the similarity between the adjacent frames was so large that the frame was divided into 50m frames and frames up to 1km in length. Because the near target can be detected well by the conventional deep learning-based detector, the minimum distance of the target was set to 100m and the maximum distance was set to a maximum of 1km, corresponding to the dot target. The distances used in this paper were the actual distances between the infrared camera and the target. As shown in Figure 3(c), the maximum experiment distance was 1 km and most of the yellow lines (flight trajectory) were performed at distances of less than approximately 500 m. This is because seasons other than winter have smaller targets and less contrast with the surrounding backgrounds, making it impossible to collect data from images.

*Dataset Construction.* Small infrared target datasets were constructed around 1,000 images. Owing to the problems mentioned above, most of the dataset was composed of less than 500m, mainly from winter and summer. Figure 4 shows the distance of the dataset from 100m to 900m.

*Augmentation Dataset.* Because it takes considerable time and effort to construct the data, less data can be accumulated unconditionally. Therefore, a method for increasing the number of data is needed. Because the target is small, the methods of changing the image, such as random noise and blur, are difficult to use because the signal of the target is likely to be distorted. The augmentation method used in this paper is a commonly used technique, and random crop augmentation and flip augmentation were applied. An example shown in Figure 5 performed flip augmentation for the original image (a), as shown in (b). (c) and (d) are the results of random crop augmentation for (a) and (b), respectively. The two augmentations were applied together and approximately 7,000 data were used for training.

*Label the Ground Truth.* When data was extracted from the infrared sequence file from a minimum distance of 100m to a maximum distance of 1km in 50 m increments, the maximum target size corresponded to 20 square pixels, a 1 or 2 pixels minimum. The precise location information of the target must be extracted from the constructed data. Considerable effort is needed compared to the general object label for the following two reasons. First, it is difficult to judge whether there is a target, even if it is close (within 500 m) in the case of a low contrast season or weather due to background cluster, such as clouds. Second, if the target exceeds 500m, the size of the target corresponds to several pixels; hence, it is difficult to confirm the existence of the target. Therefore, sequence data, radiometric temperature data, and intensity-based gray-scale data should be considered together. First, ground truth data is generated based on gray-scale data. If the gray-scale data is difficult to identify with the naked eye, the approximate position of the target is obtained through the sequence, and the accurate position of the target is obtained from the radiometric temperature data.

#### 4.2. Performance Evaluation of the Proposed Methods

*Performance Comparison Pixel-Level Fusion and Normalization Method.* Figure 6 shows the performance according to the normalization method and pixel-level fusion method. The gray-scale data and the radiometric temperature data showed inferior performance when they were normalized to different ranges. Therefore, radiometric temperature data and gray-scale data were fused at the pixel-level and the same normalization method was then used. As a result, it showed significant performance differences according to the normalization method. In particular, normalization with the mean and standard deviation calculated without normalizing to a specific range showed poor performance. Normalization to a specific range did not result in a significant difference

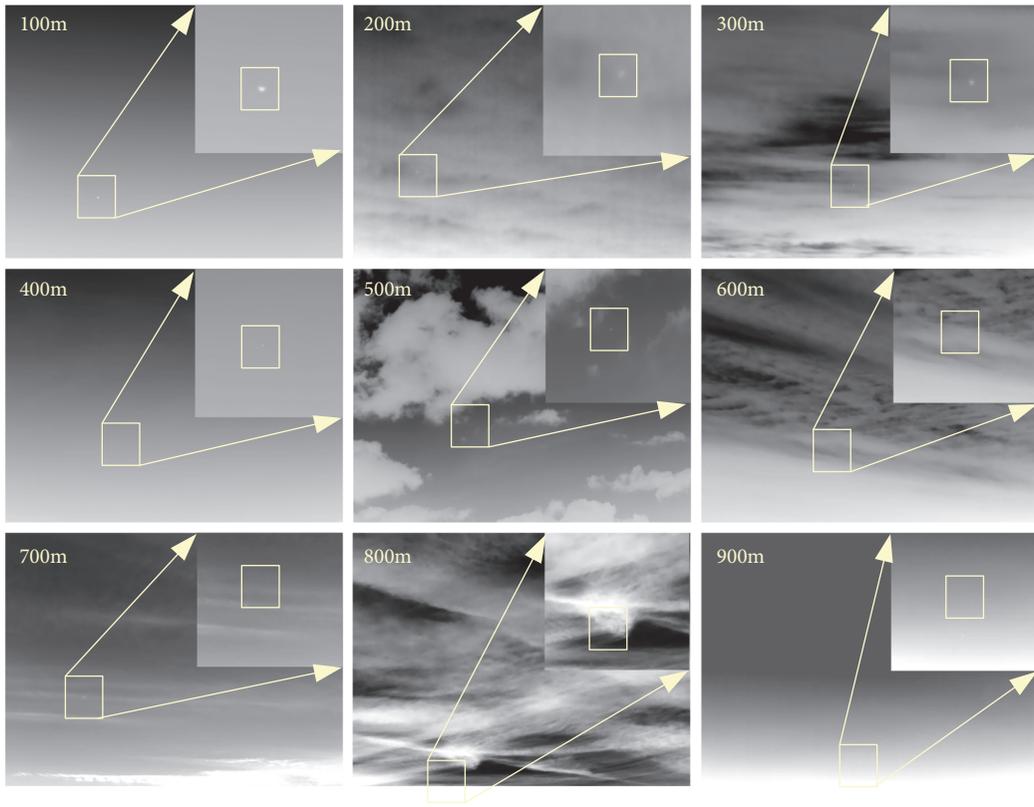


FIGURE 4: Example of infrared small target detection dataset distance from 100m to 900m.

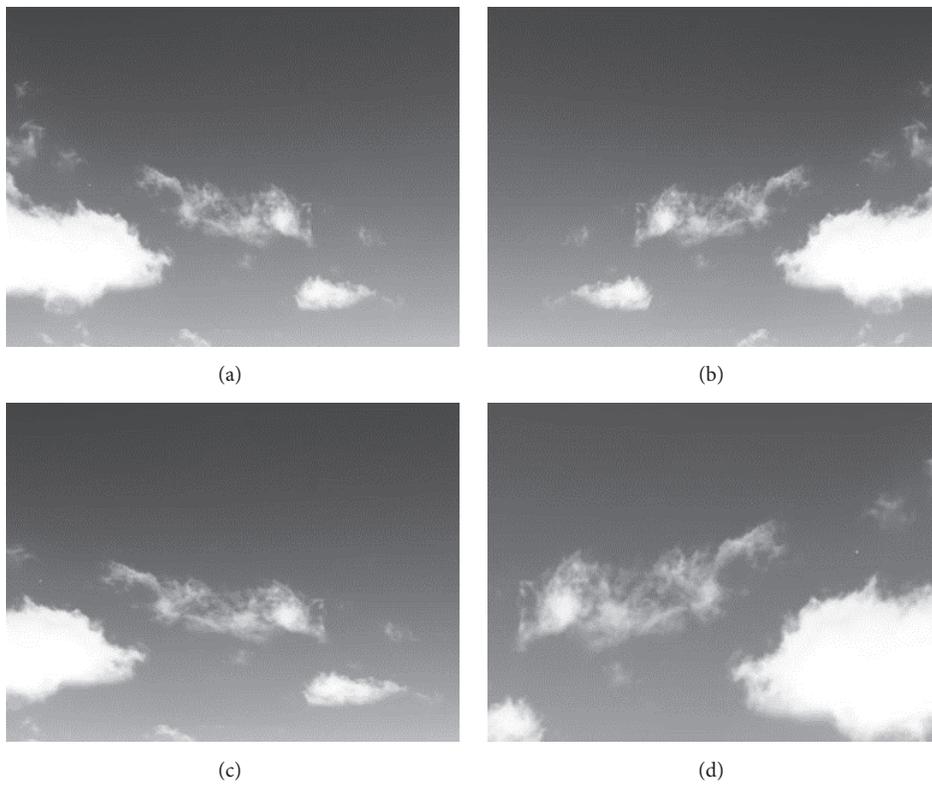


FIGURE 5: Examples of augmentation results. (a) Raw infrared, (b) flip augmented image, (c) image randomly cropped with respect to an original, and (d) random cropped image for a flip augmented image.

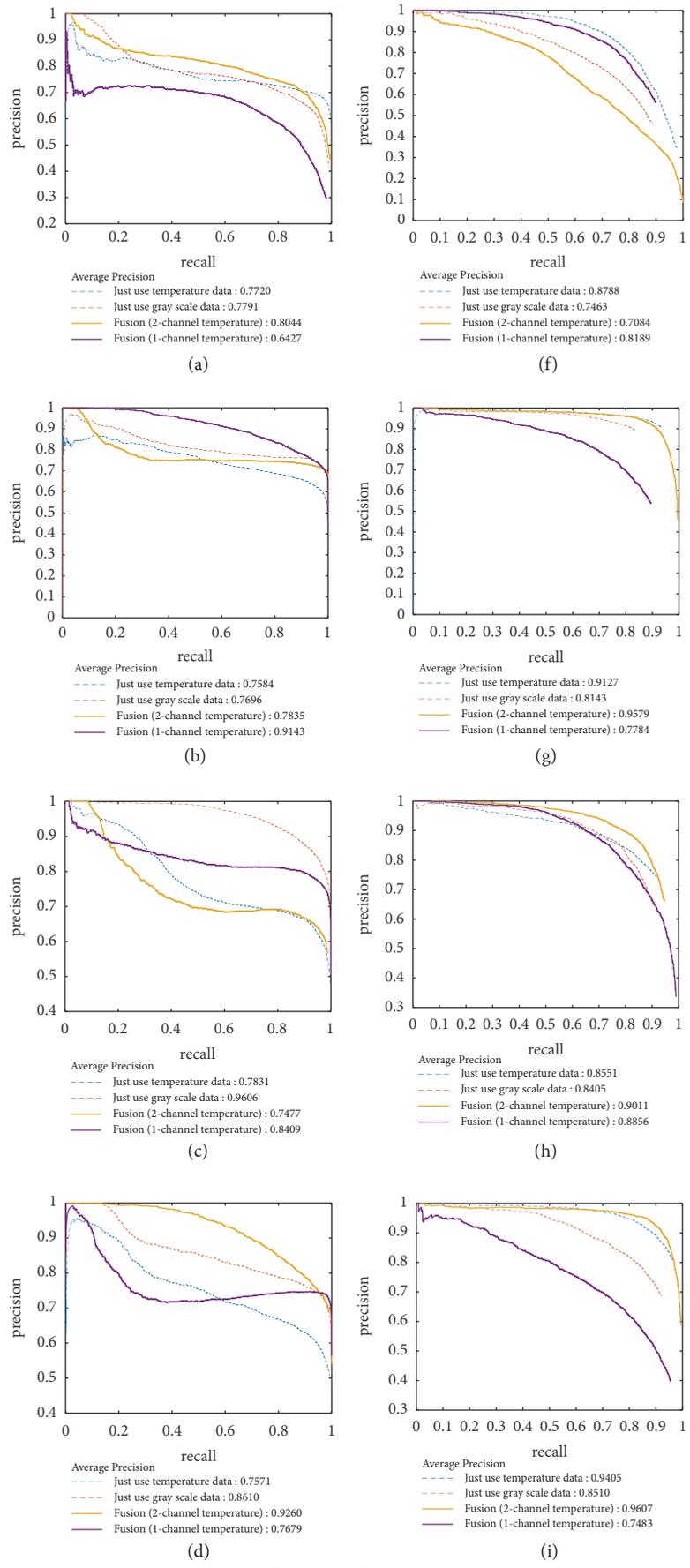


FIGURE 6: Continued.

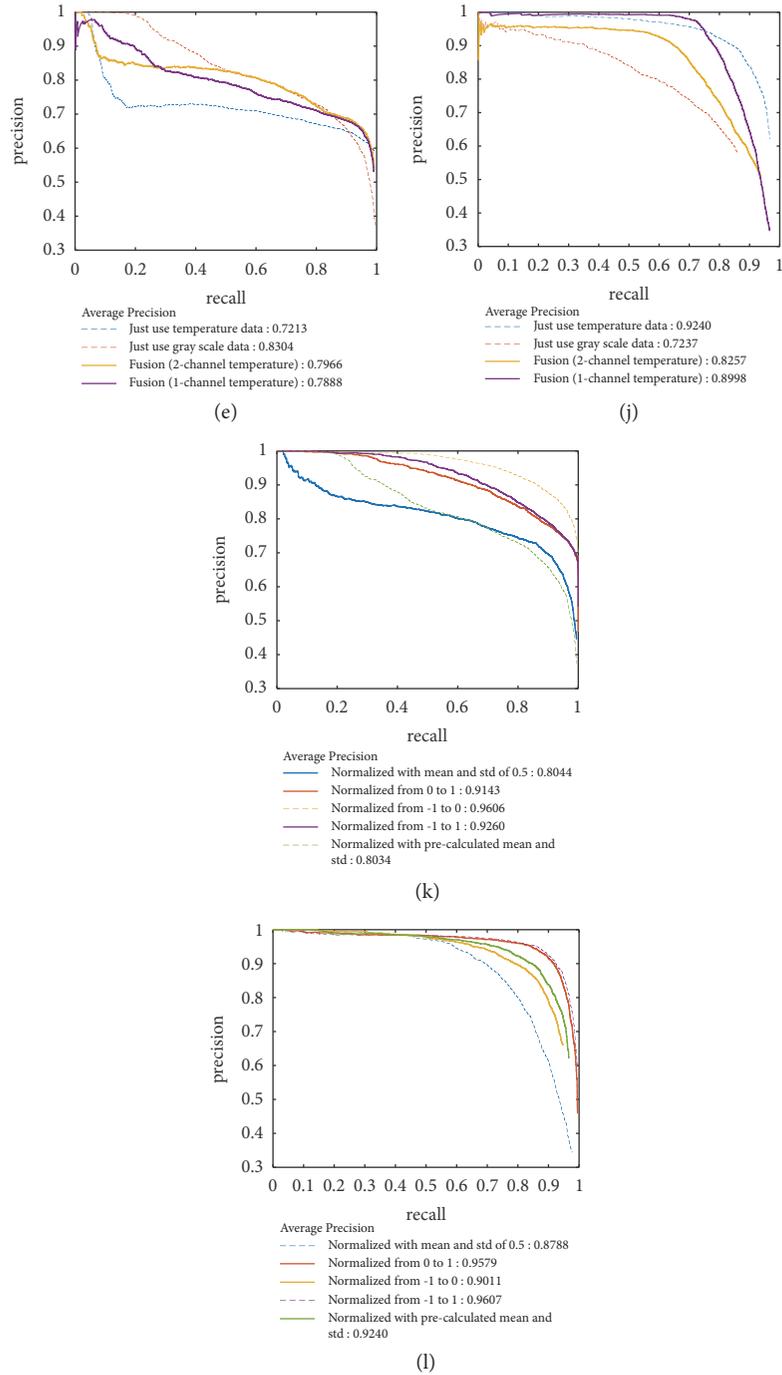


FIGURE 6: Performance comparison between the normalization and pixel-level fusion method. (a)-(e) are the results for the winter test set, and (f)-(j) are the results for the summer test set. The first row was normalized with the mean and standard deviation of 0.5; the second row is between 0 and 1; the third row is between -1 and 0; and the fourth row is between -1 and 1. The last row uses the precomputed mean and standard deviation for large scale dataset. (a)-(j) are the performance results according to the normalization method and fusion method, and (k) and (l) were obtained by collecting only the best performance of each normalization method for the summer and winter test set, respectively.

in performance between normalization methods, but overall, it was helpful to have the minimum of the normalization range to include -1. Figure 6 also shows that robust detection is possible without any significant effect on the seasonal variations.

*Experiments in a Network Optimization Perspective.* To obtain the optimized results, Table 1 compares the performance according to the network structure, batch normalization, and activation function. Because the ReLU [43] activation function does not use negative data, this study used the Leaky

TABLE 1: Performance comparison based on the network structure, batch normalization and Leaky ReLU.

# of Layers	Batch normalization	Leaky ReLU ( $\alpha:0.01$ )	Average precision
5	✓		0.7679
	✓	✓	0.7757
7	✓		<b>0.8784</b>
	✓	✓	0.7153
10	✓		0.7365
	✓	✓	0.7987
14	✓		0.6564
	✓	✓	0.6613
			0.6709
			0.6610
			0.6821
			0.6855

TABLE 2: Performance comparison based on the network structure, batch normalization and Leaky ReLU.

Normalization and fusion methods	Batch normalization	Leaky ReLU ( $\alpha:0.01$ )	Average precision
	✓	✓	+1% ~ +10%

ReLU [44] activation function with a slope factor  $\alpha$  of 0.01 and applied batch normalization. In particular, approximately 10% of the Leaky ReLU activation function was improved compared to ReLU. The performance of the table is based on the normalization method with a value between -1 and 1, and the lowest performance fusion method was used to make a clear comparison. As listed in Table 2, the AP was improved by between 1% and 10% for the various normalization and fusion methods mentioned.

*Experimental according to Fusion Method and Normalization Method.* Figure 7 shows the detection results according to the data fusion method using the fixed normalization method and Figure 8 shows the detection results according to the normalization method using the fixed data fusion method. The fixed normalization method and data fusion method use the method that showed the best performance on average. At this time, the normalization method is a method of normalizing to a value between -1 and 1, and the data fusion method is a method using two sets of radiometric temperature data.

In Figure 7, (a) is the case when only radiometric temperature data was used; (b) is for gray-scale data only; (c) is for radiometric temperature data for one channel, and (d) is for radiometric temperature data for two channels. Based on the normalization method with a value between -1 and 1, a false alarm did not occur in (d) using two radiometric temperature data, which showed the best performance and in (c) based on temperature data fusion. A false alarm occurs in (a) and (b) because it uses only single data rather than fusion-based data. On the other hand, detection was performed correctly in all four cases.

In Figure 8, (a) shows the normalization method using the previously calculated mean and standard deviation for a large scale dataset; (b) normalizes the mean and standard deviation to 0.5; (c) is the normalized value between 0 and

1; (d) is the normalized value between -1 and 0, and (e) is the detection result according to the normalized value between -1 and 1. From the detection results of (a) and (b), which performed normalization based on a specific value, it can be confirmed that although the detection is correct, many false alarms are generated and the performance is poor.

*Comparison with Existing Techniques.* Figure 9 presents a test result image from a test dataset constructed on different days and was configured to include various background clusters. Figures 9(a), 9(b), and 9(c) show the result based on the CFAR detector, high-boost (HB) method [45], and the detection results of the proposed network using the best fusion method, respectively. The CFAR detector showed 0.7621 AP, which is similar to or less than that of the deep learning-based method. The HB method works well for locating small targets, but there is a problem that the threshold parameters must be changed according to the environment changes. This paper used test datasets that were built by distance, but the maximum distance of the test dataset was only 321m because the test was done only to that distance. Robust detection is possible using the proposed deep learning-based network, even in complex and various environments, where there is a strong clutter-like cloud. In addition, robust detection is possible without being affected by seasonal changes.

#### 4.3. How Can the Radiometric Temperature Data Be Obtained?

The radiometric temperature data can be obtained using the procedure shown in Figure 10. Variable  $x$  is the raw input data and is a 14-bit digital count. The FLIR T620 infrared camera, which receives 14-bit digital count input, internally finds for  $a, b$  corresponding to the slope and intercept of the calibration curve. This process is called a radiometric calibration. The radiance  $y$  can be obtained using  $a$  and  $b$  of

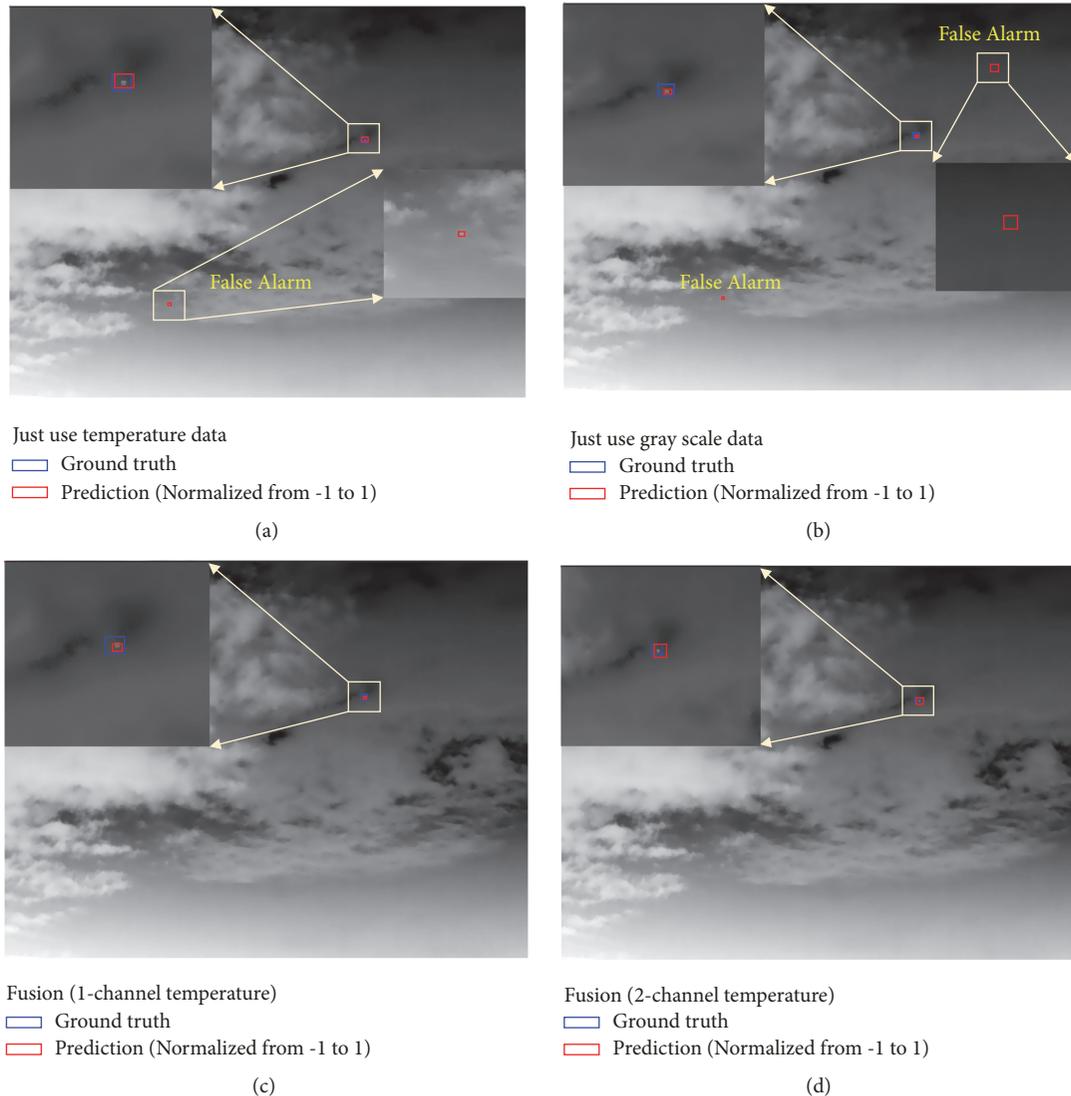


FIGURE 7: Comparison of detection results according to data fusion method. All the same normalization methods were used with values between -1 and 1. The image (a) shows the case where only the radiometric temperature data is used for three channels, (b) shows the case where only the gray scale data is used, (c) shows the case where the temperature data is fused using only one channel, and (d) is an example in which temperature data is fused to two channels.

the calibration curve and the 14-bit digital count input. The radiant energy emitted between  $T_1$  and  $T_2$ , the temperature range over which the FLIR T620 equipment operates, can be obtained by integrating the function and can be expressed in terms of  $L(\lambda)$ . This shows Planck's law as a function of the wavelength. When the radiance value corresponding to  $y$  is obtained through the calibration curve,  $L(\lambda)$  can be solved using the equation for  $L_T$  to obtain the temperature data for the input data 14-bit digital count.

## 5. Conclusions

This paper proposed a deep learning-based method for the far-infrared detection of small targets. The proposed method directly constructs datasets containing raw IR data to include

a range of backgrounds. Therefore, this study could utilize radiometric temperature data as well as commonly used gray-scale data and attempted to use this temperature data to solve the problem of a lack of information due to the small target size. Various normalization and fusion methods were examined to efficiently combine gray-scale data with radiometric temperature data. In the case of normalization, the performance was better than that using a specific value or a precomputed value for a large scale dataset rather than using a specific range. The use of data fused at the pixel-level rather than using only single data resulted in better overall performance. The seasonal performance can be detected robustly by seasonal changes. The performance of the proposed detector is similar to or better than that of the conventional detector. A comparison of the detection

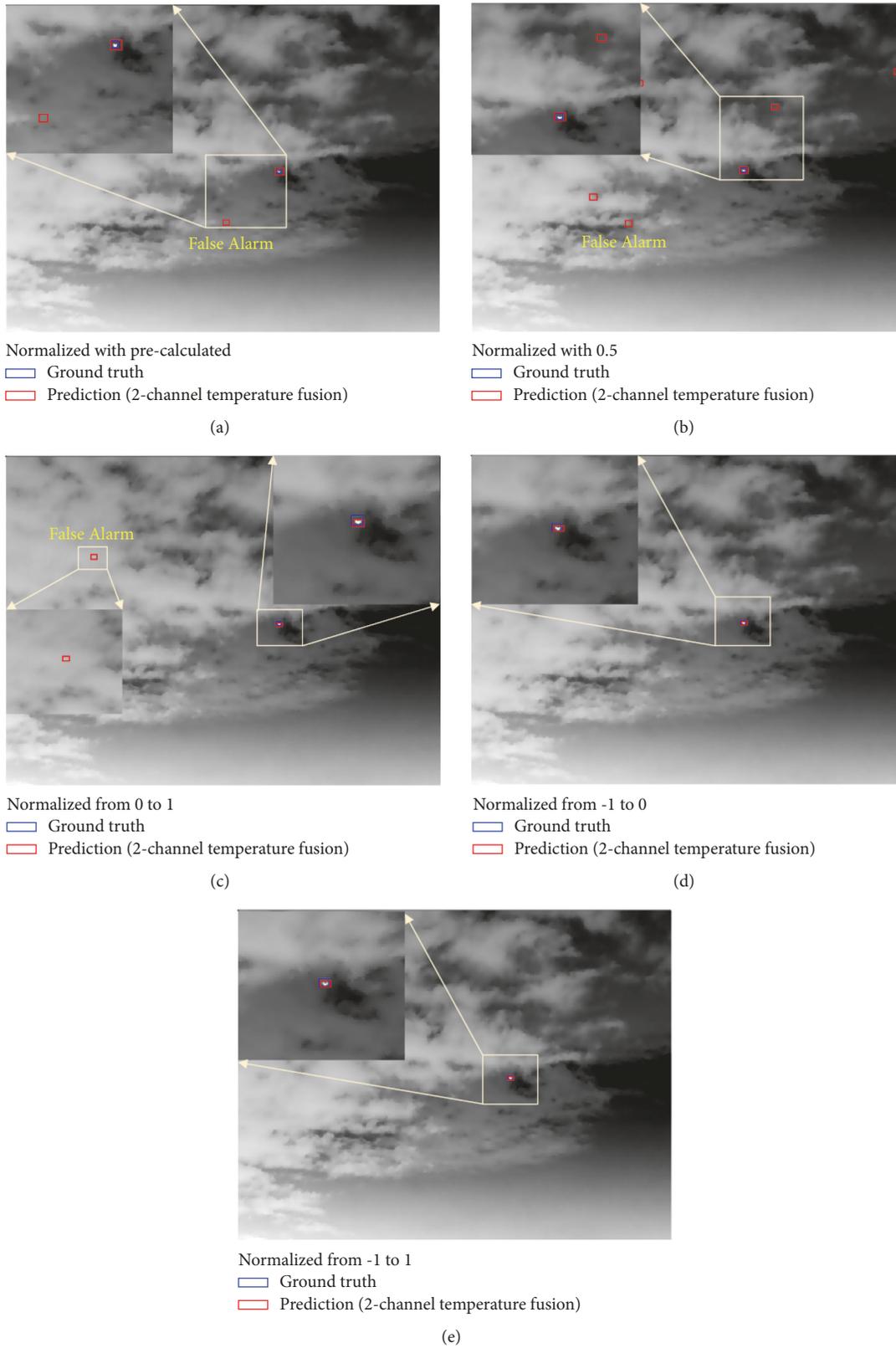


FIGURE 8: Comparison of the detection results according to the normalization method. The same fusion method was used for the two temperature channels from (a)-(e). The image (a) was normalized to a previously calculated mean and standard deviation for a large dataset; (b) was normalized to an arbitrary value of 0.5; and (c) is a value between 0 and 1. In normalization, (d) is normalized to a value between -1 and 0, and (e) is normalized to a value between -1 and 1.

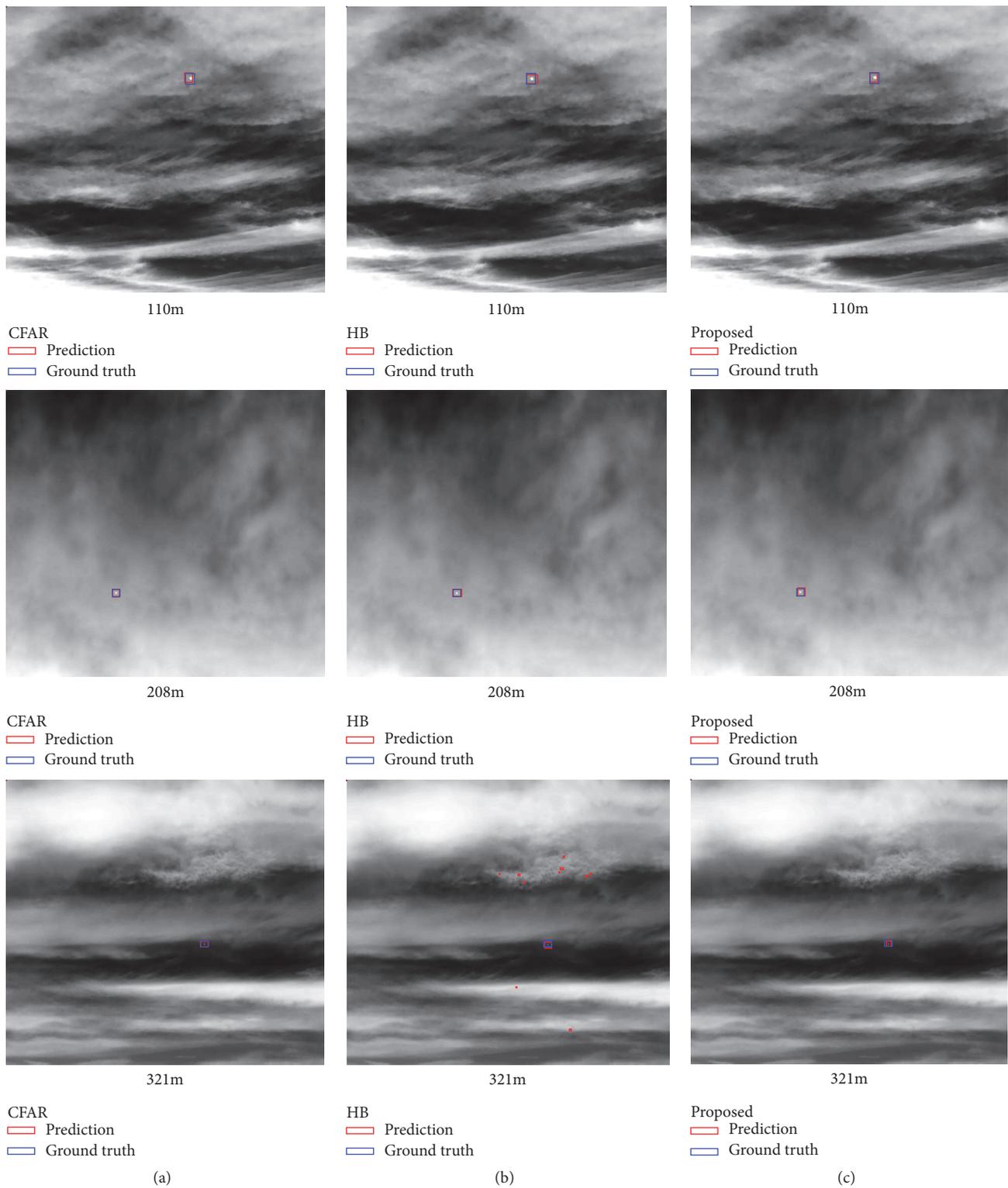


FIGURE 9: Comparison of the results of proposed deep learning based detector, conventional CFAR detector, and HB-based detector. The proposed detector is based on the fusion method using two sets of radiometric temperature data that showed the best performance and normalization method with a value between -1 and 1. In case of HB, the threshold parameter for detection at 208m was applied to 321m as it is.

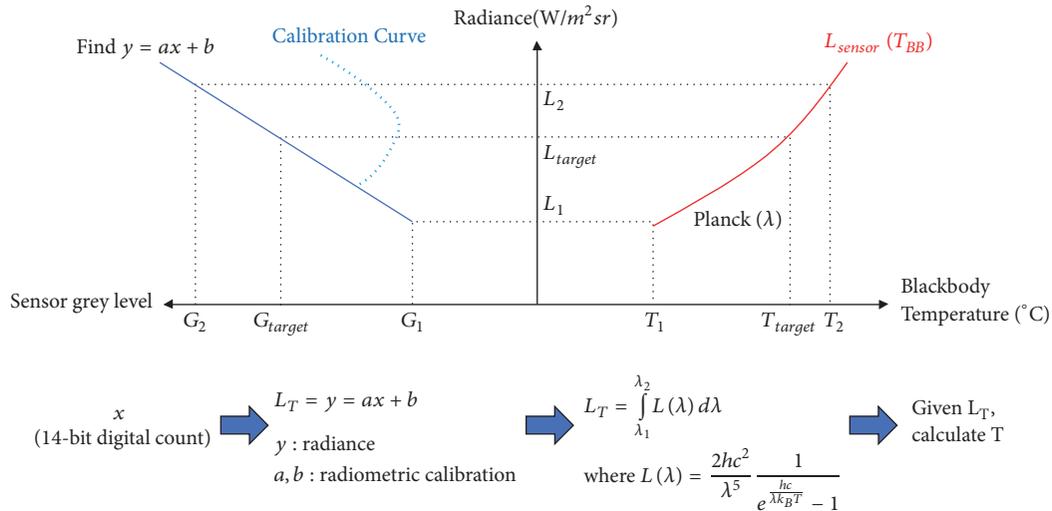


FIGURE 10: Process for obtaining radiometric temperature data. The radiance corresponding to the 14-bit digital count input data was calculated through the radiometric calibration process. Using the Planck equation to find the temperature corresponding to that radiance, the temperature is the target's radiometric temperature data.

results confirmed that the clutter can be detected robustly using the proposed deep learning-based method, even in very complicated and varying environments.

### Data Availability

The infrared small target data used to support the findings of this study have not been made available because of security reasons.

### Conflicts of Interest

The authors declare no conflict of interest.

### Acknowledgments

This work was supported by 2019 Yeungnam University Research Grants.

### Supplementary Materials

The supplementary file compares the detection results of the proposed detector with a constant false alarm rate (CFAR) detector, which corresponds to the baseline method. The first page compares the detection results of the proposed detector with the CFAR detector for the winter season, and the upper left represents the flight record for constructing the test demo dataset. The yellow solid line is the flight record of the actual target. The second page compares the results of the CFAR detector with that of the proposed detector by comparing the detection results for summer. The third page is a total seasonal flight record for building a test demo dataset containing both seasons. (*Supplementary Materials*)

### References

- [1] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, Cham, Switzerland, 2016.
- [2] R. C. Warren, *Detection of Distant Airborne Targets in Cluttered Backgrounds in Infrared Image Sequences [Ph.D. thesis]*, University of South Australia, 2002.
- [3] J. Barnett, "Statistical analysis of median subtraction filtering with application to point target detection in infrared backgrounds," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 1050, pp. 10–18, 1989.
- [4] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," *Proceedings of SPIE*, vol. 1954, pp. 25–32, 1993.
- [5] W. Schmidt, "Modified matched filter for cloud clutter suppression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 6, pp. 594–600, 1990.
- [6] T. Soni, J. R. Zeidler, and W. H. Ku, "Performance evaluation of 2-D adaptive prediction filters for detection of small objects in image data," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 327–340, 1993.
- [7] L. Yang, J. Yang, and K. Yang, "Adaptive detection for infrared small target under sea-sky complex background," *IEEE Electronics Letters*, vol. 40, no. 17, pp. 1083–1085, 2004.
- [8] B. Zhang, T. Zhang, Z. Cao, and K. Zhang, "Fast new small-target detection algorithm based on a modified partial differential equation in infrared clutter," *Optical Engineering*, vol. 46, no. 10, Article ID 106401, 2007.
- [9] B. Zhao, S. Xiao, H. Lu, and J. Liu, "Point target detection in space-based infrared imaging system based on multi-direction filtering fusion," *Progress in Electromagnetics Research*, vol. 56, pp. 145–156, 2017.
- [10] S. Qi, D. Ming, J. Ma, X. Sun, and J. Tian, "Robust method for infrared small-target detection based on Boolean map visual theory," *Applied Optics*, vol. 53, no. 18, pp. 3929–3940, 2014.

- [11] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, 2014.
- [12] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2168–2172, 2014.
- [13] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4204–4214, 2016.
- [14] Z. Cui, J. Yang, S. Jiang, and J. Li, "An infrared small target detection algorithm based on high-speed local contrast method," *Infrared Physics & Technology*, vol. 76, pp. 474–481, 2016.
- [15] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, "Infrared small target detection utilizing the multiscale relative local contrast measure," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 612–616, 2018.
- [16] D. J. Gregoris, S. K. Yu, S. Tritchew, and L. Sevigny, "Detection of dim targets in FLIR imagery using multiscale transforms," *Proceedings of SPIE*, vol. 2269, pp. 62–72, 1994.
- [17] S. Kim and J. Lee, "Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track," *Pattern Recognition*, vol. 45, no. 1, pp. 393–406, 2012.
- [18] H. Fu, Y. Long, R. Zhu, W. An, H. Yu, and J. Dong, "Infrared small target detection based on multiscale center-surround contrast measure," in *Proceedings of the Ninth International Conference on Graphic and Image Processing*, vol. 10615, Qingdao, China, 2018.
- [19] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 1–4, pp. 131–156, 1997.
- [20] S. Kim, "Analysis of small infrared target features and learning-based false detection removal for infrared search and track," *Pattern Analysis and Applications*, vol. 17, no. 4, pp. 883–900, 2014.
- [21] Y. Bi, X. Bai, T. Jin, and S. Guo, "Multiple feature analysis for infrared small target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1333–1337, 2017.
- [22] C. Gao, T. Zhang, and Q. Li, "Small infrared target detection using sparse ring representation," *IEEE Aerospace and Electronic Systems Magazine*, vol. 27, no. 3, pp. 21–30, 2012.
- [23] Y. Gu, C. Wang, B. Liu, and Y. Zhang, "A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 3, pp. 469–473, 2010.
- [24] Y. Li and Y. Zhang, "Robust infrared small target detection using local steering kernel reconstruction," *Pattern Recognition*, vol. 77, pp. 113–125, 2018.
- [25] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 60–72, 2016.
- [26] H. Deng, Y. Wei, and M. Tong, "Small target detection based on weighted self-information map," *Infrared Physics & Technology*, vol. 60, pp. 197–206, 2013.
- [27] Z. Wang, J. Tian, J. Liu, and S. Zheng, "Small infrared target fusion detection based on support vector machines in the wavelet domain," *Optical Engineering*, vol. 45, no. 7, Article ID 076401, 2006.
- [28] S. Kim, K.-J. Yoon, and I. S. Kweon, "Object recognition using a generalized robust invariant feature and Gestalt's law of proximity and similarity," *Pattern Recognition*, vol. 41, no. 2, pp. 726–741, 2008.
- [29] S. Qi, J. Ma, C. Tao, C. Yang, and J. Tian, "A robust directional saliency-based method for infrared small-target detection under various complex backgrounds," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 495–499, 2013.
- [30] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint  $l_2, l_1$  norm," *Remote Sensing*, vol. 10, no. 11, 2018.
- [31] Z. Chen, G. Wang, J. Liu, and C. Liu, "Infrared small target detection algorithm based on self-adaptive background forecast," *International Journal of Infrared and Millimeter Waves*, vol. 27, no. 12, pp. 1619–1624, 2007.
- [32] X. Fan, Z. Xu, J. Zhang, Y. Huang, and Z. Peng, "Infrared dim and small targets detection method based on local energy center of sequential image," *Mathematical Problems in Engineering*, vol. 2017, Article ID 4572147, 16 pages, 2017.
- [33] M. Liu, H. Y. Du, Y. J. Zhao, L. Q. Dong, and M. Hui, "Image small target detection based on deep learning with snr controlled sample generation," in *Current Trends in Computer Science and Mechanical Automation*, vol. 1, pp. 211–220, Sciendo Migration, 2017.
- [34] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.
- [35] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 1951–1959, USA, July 2017.
- [36] G. X. Hu, Z. Yang, L. Hu, L. Huang, and J. M. Han, "Small object detection with multiscale features," *International Journal of Digital Multimedia Broadcasting*, vol. 2018, 2018.
- [37] B. Bosquet, M. Mucientes, and V. M. Brea, "STDnet: A ConvNet for small target detection," in *Proceedings of the British Machine Vision Conference*, 2018.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 770–778, July 2016.
- [39] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, 2015.
- [40] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15)*, pp. 1026–1034, IEEE, Santiago, Chile, December 2015.
- [41] Y. L. Hou, Y. Song, X. Hao, Y. Shen, and M. Qian, "Multispectral pedestrian detection based on deep convolutional neural networks," *Infrared Physics & Technology*, vol. 94, pp. 69–77, 2018.
- [42] T. Kim and S. Kim, "Pedestrian detection at night time in FIR domain: Comprehensive study about temperature and brightness and new benchmark," *Pattern Recognition*, vol. 79, pp. 44–54, 2018.
- [43] V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, pp. 807–814, Haifa, Israel, June 2010.

- [44] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning*, vol. 30, 2013.
- [45] Y. Shi, Y. Wei, H. Yao, D. Pan, and G. Xiao, "High-boost-based multiscale local contrast measure for infrared small target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 33–37, 2018.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

