IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 3, 2020, accepted April 16, 2020, date of publication April 20, 2020, date of current version May 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2988931

Region-Based Removal of Thermal Reflection Using Pruned Fully Convolutional Network

GANBAYAR BATCHULUUN, NA RAE BAEK, DAT TIEN NGUYEN, TUYEN DANH PHAM, AND KANG RYOUNG PARK, (Member, IEEE)

Division of Electronics and Electrical Engineering, Dongguk University, Seoul 04620, South Korea

Corresponding author: Dat Tien Nguyen (nguyentiendat@dongguk.edu)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) through the Basic Science Research Program under Grant NRF-2019R1F1A1041123, Grant NRF-2019R1A2C1083813, and Grant NRF-2020R1A2C1006179.

ABSTRACT In general, an image obtained from a thermal camera often has a mirror reflection or shadow reflected off the ground around an object, which is referred to as thermal reflection. Sometimes the thermal reflections are connected to their objects in images, which makes it difficult to detect or recognize the object only. Thermal reflections sometimes occur on the wall near an object and are detected as another object when they are not connected to the object. Furthermore, the size of thermal reflection and pixel value significantly vary with the medium of the reflected range and the surrounding temperature. In these cases, the patterns and pixel values of thermal reflection and the object become similar and difficult to distinguish. However, there are insufficient studies on removing the thermal reflection of various kinds of objects in diverse environments. Therefore, in this paper, we propose a pruned fully convolutional network (PFCN)-based method for removing the thermal reflection of an object using the surrounding information when image transformation is performed only within the region of an object. When experiments were conducted using self-collected databases (Dongguk thermal image database (DTh-DB) and Dongguk items & vehicles database (DI&V-DB)) and open databases, the method proposed herein exhibited more outstanding performance in removing thermal reflection when compared with the state-of-the-art methods.

INDEX TERMS Thermal image, image transform, thermal reflection removal, pruned fully convolutional network.

I. INTRODUCTION

Typically, a long-wavelength infrared (LWIR) camera, which is often used in surveillance systems, can measure electromagnetic radiation of wavelengths 8–12 μm [1]. Most of the thermal radiation generated from an object or body is infrared radiation, and the LWIR camera is commonly used to measure such heat information. Hence, an LWIR camera is also referred to as a thermal camera. A thermal camera can make objects at close and far distances visible in dark surroundings without using an additional illuminator. Figure 1 shows the thermal camera, a visible light camera, thermal images and the respective visible light images. However, as shown in Figure 2, there are thermal reflections (the areas of dotted line) such as shadows or mirror reflections on the ground surface near the object in the images obtained using a thermal camera in both indoor and outdoor environments. The performance of object detection or recognition algorithms is degraded due to such thermal reflections. However, very few studies have been conducted on the removal of thermal reflection. Therefore, we propose a novel method for the removal of thermal reflections by conducting image transformation only within the specific region of thermal images. Recently, various image transformation algorithms have been developed for deep-learning-based image processing tasks. In particular, image-to-image translation methods based on generative adversarial network (GAN) have been showing high accuracy. Normally, an entire image is transformed when transforming an object in an image. However, the accuracy is reduced in such a method, as the background region is also transformed in addition to the object being transformed. Thus, a method for increasing the accuracy of transformation is proposed. In the method, transformation operation is conducted only within the region of an object that has been

The associate editor coordinating the review of this manuscript and approving it for publication was Qingli Li.

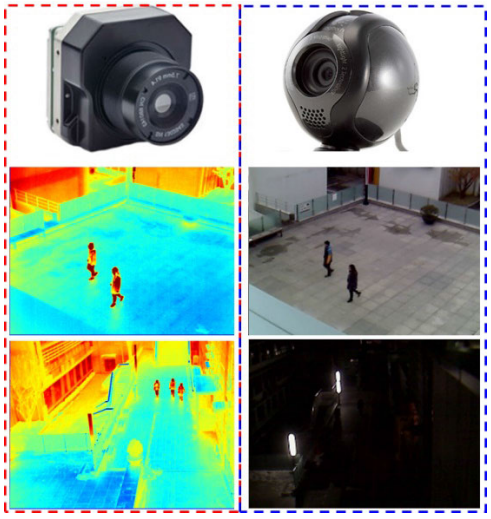

FIGURE 1. From the left to the right, the thermal camera with captured thermal images, and the visible light camera with captured images, respectively. Images captured in daytime and night, respectively, from the top to the bottom.

detected using deep learning. The surrounding information is also considered when transforming an image only within the region of an object. More specifically, thermal reflections in a thermal image are removed by transforming the heat of specific regions of the wall and surrounding floor to match the heat of the background.

The remainder of this paper is organized as follows. Previous studies are discussed in Section II, and the contributions of this study are explained in Section III. The details of the proposed method are explained in Section IV. The experiment results and comparison experiment are discussed in Section V, and lastly, the conclusion of this study is provided in Section VI.

II. RELATED WORKS

The existing deep-learning-based image transformation methods can be divided into transforming only a specific region of an image and transforming an entire image. There are several GAN-based image-to-image translation methods [2]–[9]. In study [2], authors developed a two-step unsupervised learning method that transforms images between different domains by using unlabeled images without specifying any correspondence between them so as to avoid the cost of acquiring labeled data. In [3], an unsupervised image-to-image translation (UNIT) method based on GAN and variational autoencoders (VAEs) is proposed. In the paper, two limitations of the method are explained. The first limitation is that the transformation model is unimodal due to the Gaussian latent space assumption. The second limitation is that the training could be unstable due to the saddle point searching problem. In study [4], triangle GAN that can be used for semi-supervised joint distribution matching is proposed. The approach learns the bidirectional mappings between two domains with a few paired sample images. In [5], a StarGAN,

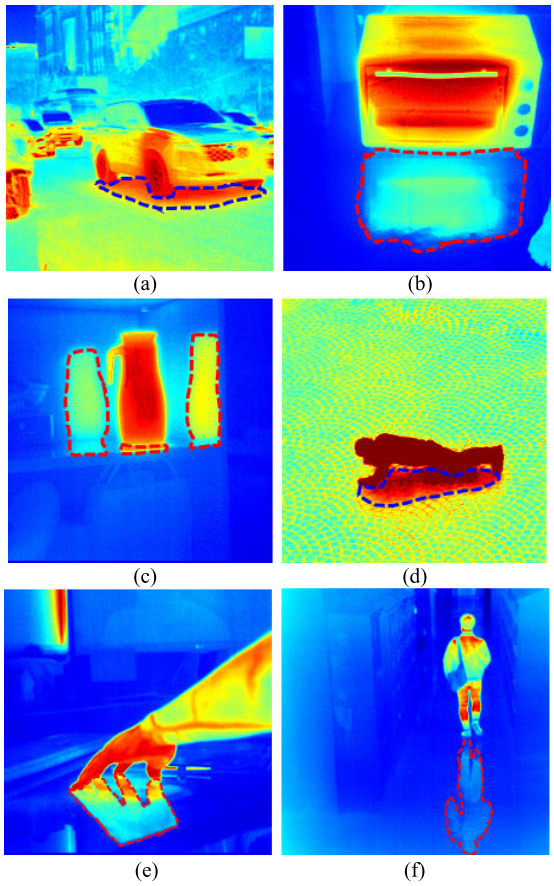

FIGURE 2. Example of the thermal reflections. (a) A vehicle; (b) a hot oven; (c) a glass bottle with hot water; (d) a man lying on an outdoor floor; (e) a hand and arm; (f) a man walking in an indoor corridor.

a scalable method that can perform image-to-image transformation for multiple domains using only a single model is proposed. The architecture of StarGAN allows simultaneous training of multiple image data sets with different domains within a single network. In [6], a method based on GAN that learns from images to discover relations between images in different domains (DiscoGAN) is proposed. The DiscoGAN can generate highly qualified images with transferred style without using any explicit pair labels and learns to relate images from very different domains. In study [7], GAN in the conditional setting is explored to design new conditional GAN (cGAN) that learns a conditional generative model. In [8], a coupled GAN (CoGAN) method for learning a joint distribution of multi-domain dataset is proposed. In contrast to the existing methods, it requires tuples of corresponding image data in different domains in the training set. CoGAN method learns a joint distribution without any tuple of corresponding image data.

In studies [2]–[8], an entire image was transformed using a deep learning network. In study [9], the network was trained by using an entire image and a corresponding mask image of objects simultaneously as inputs. They also reported that image transformation methods in previous studies fail in case of multiple objects and when the shape of an object

changes. Hence, the performance of image transformation was enhanced using the mask image. In study [10], a perceptual loss network (PLN)-based method was proposed in which image-to-image translation was performed while the image style was transferred. However, the above image-to-image translation methods involve transforming an entire image; thus, the accuracy is reduced more by the background region being transformed in addition to the region of an object than by only the region of an object being transformed. Hence, we propose a method for increasing the accuracy of image transformation by transforming only the region of an object. In this study, a method for removing thermal reflection in images obtained using a thermal camera was examined. Typically, there are two problems in thermal images, namely thermal reflection [11]–[14] and halo effect [15]–[17]. In study [11], a method of suppressing thermal reflection is proposed. In the method, the visible light reflection and the reflection of heat are experimented. Additionally, various polarizers and plates are also used, and the change in the thermal reflection according to the angle of the plates is graphically shown. In this way, a thermal reflection suppression technique considering the angles according to the plates of various polarizers and materials in the experiment is proposed. However, in this method, given that the angle varies depending on the material of the plates, the suppression performance is reduced when there are nearby floors or walls made of different materials.

In [12], a thermal reflection elimination method is proposed. The method is conducted using Mask R-CNN [18] to detect thermal reflection regions in thermal images. The method eliminates thermal reflections based on the detected regions of thermal reflections. The method changes the value of pixels only in the detected regions to increase the accuracy of the transformation. In [13], two methods are proposed such as a method that classifies the regarded material in order to estimate improved surface temperature values, and a method to detect and remove thermal reflections in thermal images. The detection method is conducted using a background subtraction algorithm. To remove a thermal reflection, the method uses weighted moving averages. In study [14], a novel reflection removal approach using polarization properties of the reflection in thermal images is proposed. The method uses four input images of different polarization angles such as 0, 45, 90, and 135 degrees for removing thermal reflections. These studies are conducted to remove thermal reflections without using deep learning.

Halo effect is explained in a documentation provided by FLIR [15]. For example, the halo effect in a thermal image is a circular region of high intensity pixels that surrounds an object. In studies [16] and [17], methods for the detection of subjects in images with the halo effect were proposed using contour-based approaches. The methods are based on background subtraction that fuses contours obtained from a thermal image and a visible light image. However, these contour-based methods are not methods for removing halo effects in images, but rather, approaches for accurately detecting subjects in thermal images with halo effects. Moreover, other existing studies that have investigated the detection [19]–[26], identification [27], [28], and recognition [29]–[31] of thermal images and the survey study [32] did not consider these two problems. Therefore, we propose an image transformation method based on the regions of thermal reflection using deep learning.

In addition, there are previous studies [33]–[42] that are conducted for image inpainting tasks which are similar to a task conducted in this study. Image inpainting technique is used to fill damaged, deteriorating, or missing parts of an image. In study [33], a method for semantic image inpainting, which generates the missing information by conditioning on the available data is proposed. The authors claim that in their method, inference is possible irrespective of how the missing information is structured, while the state-of-the-art learning-based methods require specific information about the holes in the training phase. In [34], a spatial region-wise normalization named region normalization (RN) to overcome the limitation of image inpainting problem is proposed. The mean and variance shifts caused by full-spatial feature normalization (FN) limit the image inpainting network training is presented. In [35], a method based on a deep generative model which can not only synthesize novel image structures but also explicitly utilize surrounding image features as references during network training to make better predictions is proposed. The model is a feedforward, fully convolutional neural network (FCN) which can process images with variable sizes and with multiple holes at arbitrary locations during the test time. In study [36], a generative image inpainting approach to complete images with guidance and free-form mask is proposed. The approach is based on gated convolutions learned from huge number of images without additional labelling efforts. The authors presented user sketch as an exemplar guidance to help users to remove distracting objects quickly, modify image layouts, edit faces, clear watermarks, and create novel objects in images.

In [37], a learnable bidirectional attention maps (LBAM) for image inpainting is proposed. The method used FCN to conduct image inpainting. In [38], a fined deep generative model-based method which designed a coherent semantic attention layer to learn the relationship between features of missing information in images. The method used FCN to conduct image inpainting. In study [39], an architecture named Shift-Net for image completion that exhibits high speed with promising details via deep feature rearrangement is proposed. The study presented a special shift-connection layer to the U-Net architecture. The method uses FCN with a shift layer. In [40], an image inpainting model named PEPSI which overcomes the limitation of the two-stage coarse-to-fine network using the joint learning scheme is proposed. In study [41], a two-stage adversarial model EdgeConnect that comprises of an edge generator followed by an image completion network is proposed. First, the method generates an edge information from a damaged image then combine the obtained edge information with the damaged image as inputs

TABLE 1. Summary of comparison between the proposed method and previous image transformation methods.

Category		Method	Advantage	Disadvantage
Without using a region of an image	Without using deep learning method	Thermal reflection suppression [11]	Large data acquisition, processing, and training are not required.	- Performance is low. - Performance is affected by background transform.
	Using deep learning method	Image-to-image translation [10]	Can extract suitable features in various camera settings and environments.	- Large data acquisition, processing, and training require more time. - Performance is affected by background transform.
	Using GAN-based deep learning method	Image-to-image translation [2–9]		
Using a region of an image	Without using deep learning method	Removing thermal reflection [12]	- Large data acquisition, processing, and training are not required. - Performance is not affected by background transform.	Performance is low.
	Using deep learning method	Removing thermal reflection (**proposed method**)	- It can extract suitable features in various camera settings and environments. - High performance not affected by background transform.	It requires the procedure of intensive training of deep CNN.

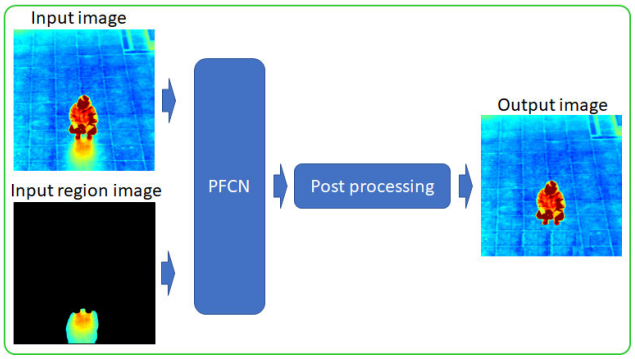

FIGURE 3. Overall flowchart of proposed method.

to second generator for desired output. In [42], a PGGAN approach is proposed. The method includes a discriminator network that combines a global GAN (G-GAN) architecture with a patch GAN.

To consider the limitation of previous works, we propose an image transformation method based on the regions of thermal reflection using deep learning. The summary of a comparison between the proposed method and previous image transformation methods is provided in Table 1.

III. CONTRIBUTIONS

This research is novel in the following four ways compared with previous works:

- This study is the first of its kind to remove thermal reflection in thermal images using deep learning.
- A general image processing method can remove thermal reflection by transforming an image only within the region where thermal reflection was detected; however, a method for transforming an image only within the detected region using deep learning does not exist currently. Therefore, we suggest a deep-learning-based method for transforming an image only within the region where thermal reflection is detected.
- In this study, a pruned fully convolutional network (PFCN), in which the heat information of surrounding walls and ground is considered, is newly proposed for transforming an image only within the area where thermal reflection is detected.
- The convolutional neural network (CNN) models for removing thermal reflection developed in this study are disclosed through [45] for an evaluation of performance by other researchers.

IV. PROPOSED METHOD

A. OVERALL PROCEDURE OF PROPOSED METHOD

In this section, the method proposed in this paper is explained in detail. In the proposed method, only specific region of an image is transformed using PFCN architecture, which is the improved version of existing FCN [46]. Figure 3 shows the flowchart of the proposed method. Moreover, a method for obtaining output images to be transformed by PFCN is further explained in section IV. *B*, whereas a method for removing thermal reflection using the output images obtained by PFCN is further explained in section IV. *D*. The thermal camera used in this study can obtain an image at the speed of 30 frames per second (fps) [47]. It can measure the temperature from -40 °C to +80 °C to make objects visible in both light and dark environments. The database (an image has the depth of 14 bits and the size of 640 × 480 pixels [12]) obtained using the thermal camera was used in the experiment. A mask region CNN (Mask R-CNN) was used to detect the approximate region (input region image in Figure 3) of thermal reflection in input images, and the detailed explanation is provided in [12].

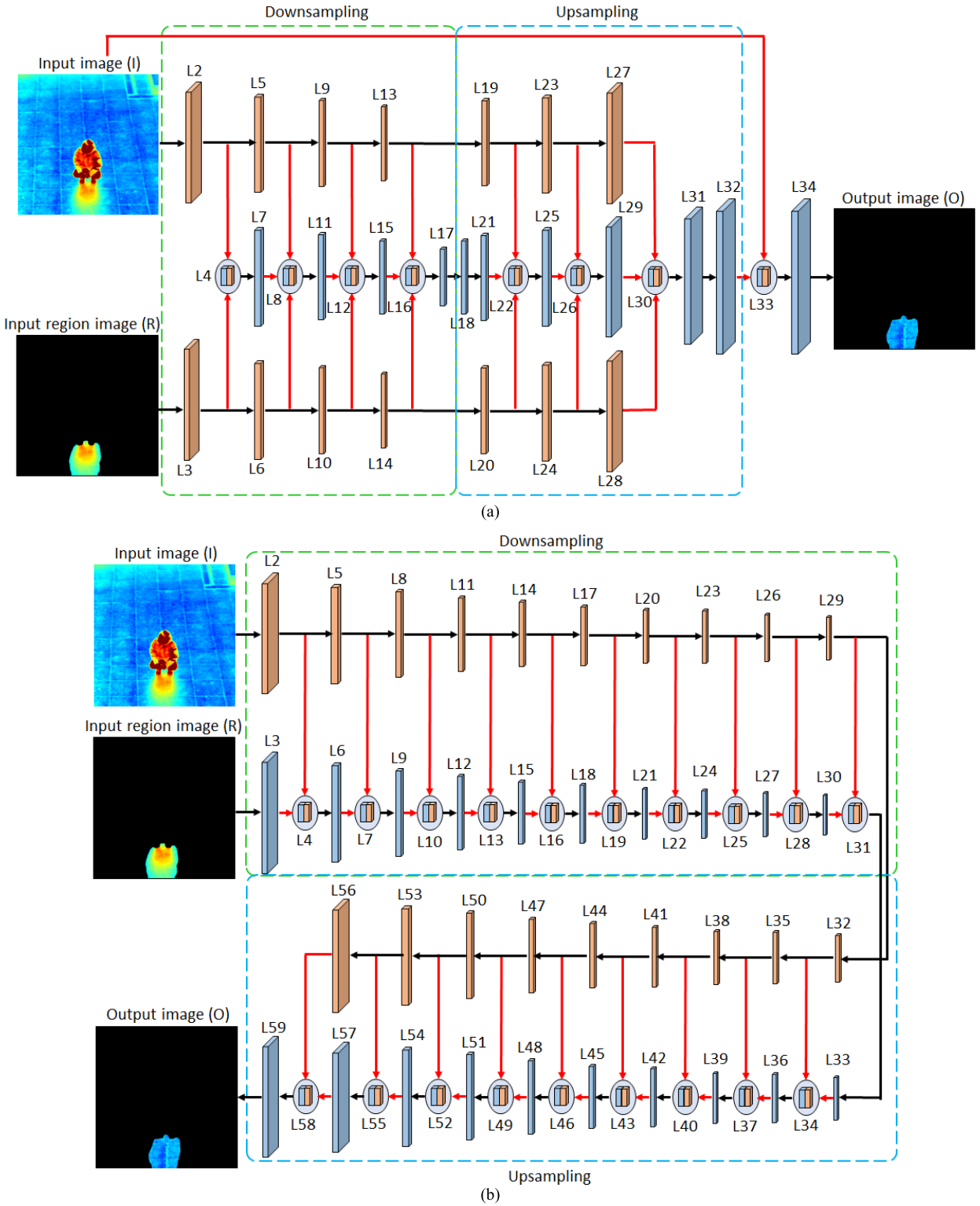

FIGURE 4. Two types of proposed FCN architectures. (a) FCN_V1; (b) FCN_V2.

B. REGION-BASED IMAGE TRANSFORMATION USING PRUNED FULLY CONVOLUTIONAL NETWORK

In this section, the proposed PFCN is explained in detail. While previous studies focused on transforming the entire original image, the proposed method increases the accuracy by conducting image transformation only within the region of the object in an image. A typical FCN architecture is used in this study as shown in Figure 4(a) or (b). The original image I

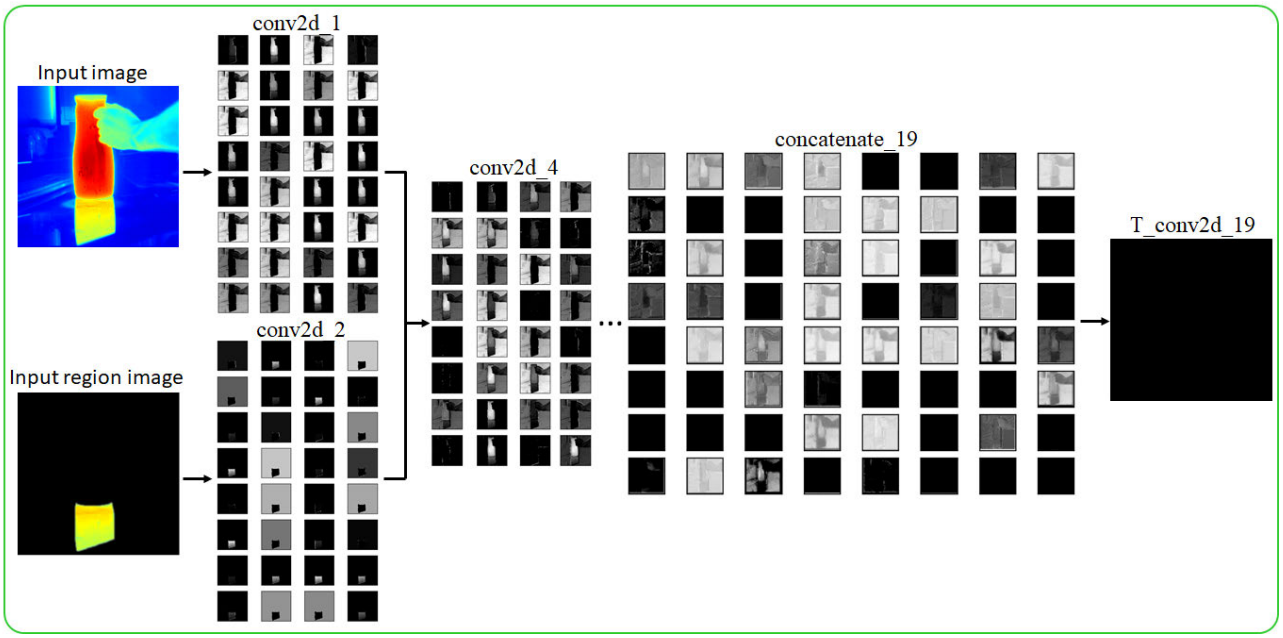

FIGURE 5. Example of extracted feature maps by using FCN_V2.

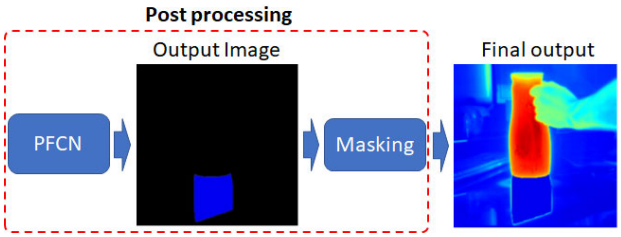

FIGURE 6. Example of the post processing and the final output image.

and input region image R of the object are used as inputs to generate the output image O. R is the image attempting to be transformed, and I is the image providing information on the surroundings.

For example, when removing a shadow of an object in the visible light image, the pixel intensity within the region is transformed to be similar to the pixel intensity of nearby ground or walls. Accordingly, I is used as an input to FCN to extract the information on the surroundings. Two different structures were experimented for the proposed method. The idea of the first structure (Figure 4(a)) is to generate O by extracting the features of I and R and combining them. The idea of the second structure (Figure 4(b)) is to update the convolution layers with the information extracted from I when transforming R to O. The concatenate layer is used when combining feature maps. Feature maps extracted from Figure 4(a) and (b) are indicated by light blue boxes and light orange boxes, respectively, whereas concatenate and convolution operations are indicated by red arrow and black arrows, respectively. L2–L34 in Figure 4(a) and L2–L59 in Figure 4(b) represent layer numbers in Tables 6 and 7 in Appendix. The details of the two structures used in the proposed method

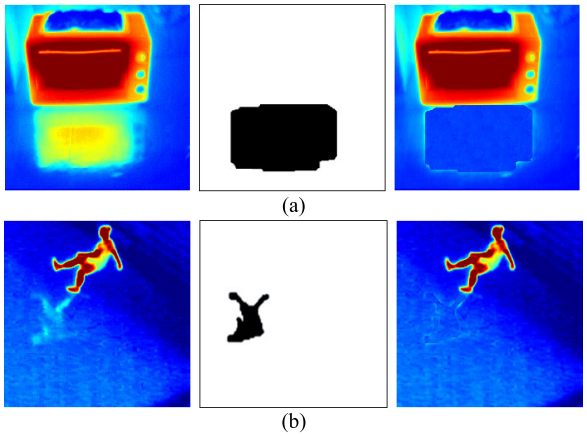

FIGURE 7. Example of removing thermal reflection. From the left to the right, original images, mask images, and final output images are presented. (a) A hot oven on a floor; (b) a man in a soccer field.

(FCN version 1 (FCN_V1) and FCN version 2 (FCN_V2)) are shown in Tables 6 and 7. The thermal images used in this study are one-channel gray scale images, not three-channel color images. In this study, a color mapping function [48] is used to map gray scale thermal images to color thermal images for accurately representing the information of the heat and surrounding temperature of the objects in images. In Tables 6 and 7 in Appendix, all the convolution layers are followed by the rectified linear unit (ReLU). In Table 6, (1×1) padding is used for conv2d_13, and (0×0) padding is used for the convolution layers. Furthermore, the filter size, stride, and padding are (1×1), (3×3), and (0×0), respectively, in Table 7.

In this study, the PFCN with enhanced performance is proposed instead of using FCN_V1 and FCN_V2 in Tables 6

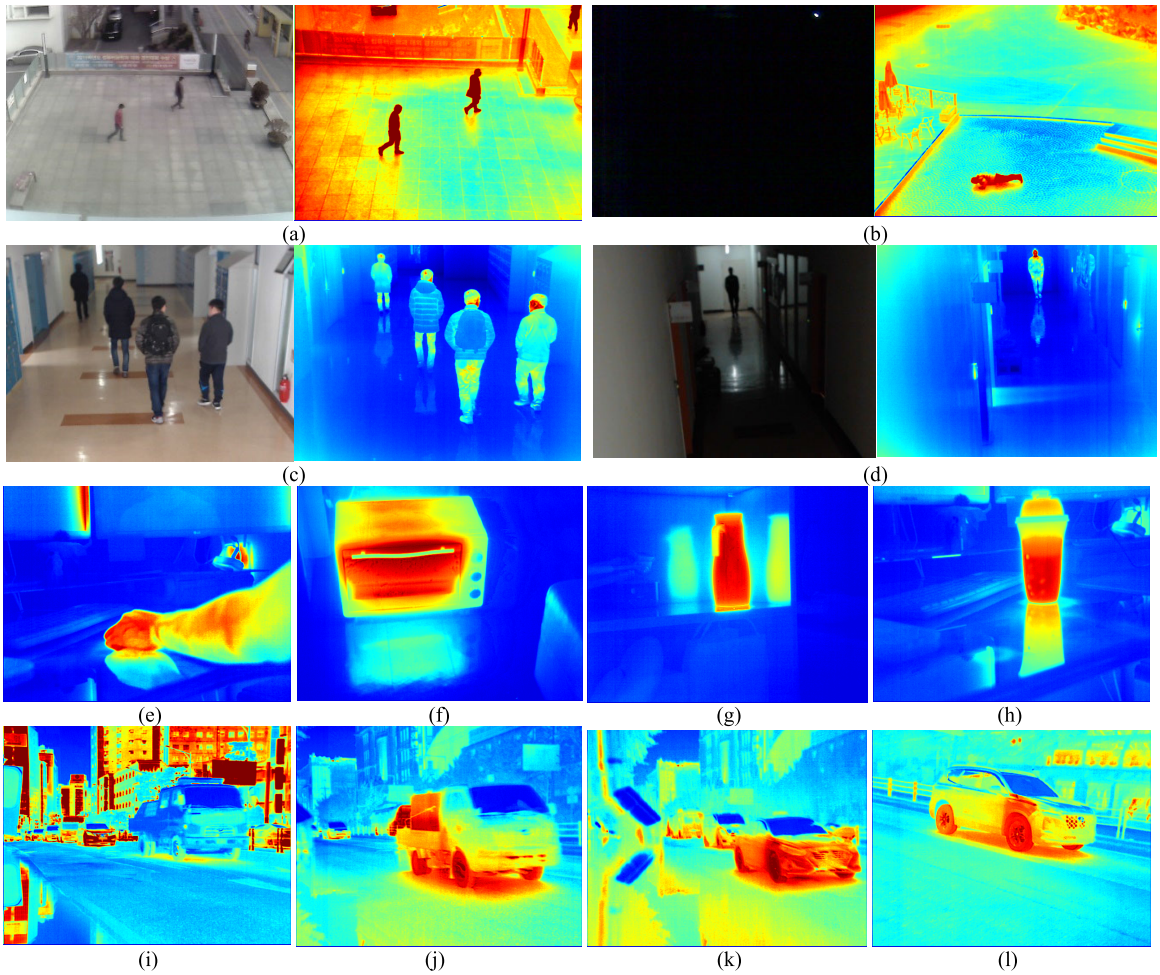

FIGURE 8. Example images of experimental database. (a) A visible light image and its corresponding thermal image captured in a bright outdoor environment; (b) a visible light image and its corresponding thermal image captured in a dark outdoor environment; (c) a visible light image and its corresponding thermal image captured in a bright indoor environment; (d) a visible light image and its corresponding thermal image captured in a dark indoor environment; (e–h) thermal images captured in an indoor environment; (i–l) thermal images captured in an outdoor environment.

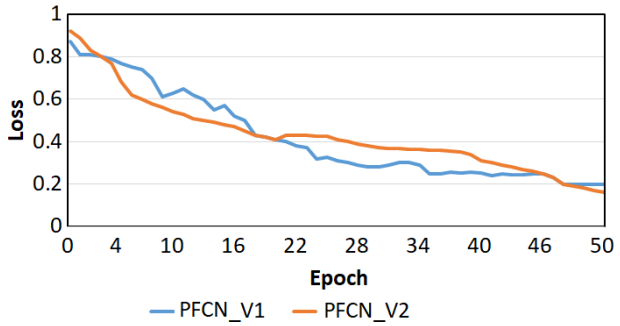

FIGURE 9. Example of training loss curves of PFCN.

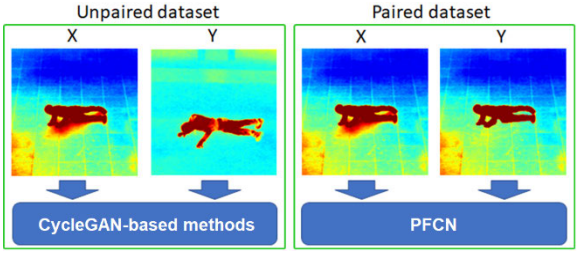

FIGURE 10. Example of training methods using paired and unpaired datasets.

and A.2 in Appendix as they are. The PFCN is a model with a reduced number of channels and parameters of the FCN based on the pruning (network surgery) [49] technique. The detailed explanation is provided in section IV.*C*. For a model obtained by training the FCN, a complete black image can be output as shown in Figure 5. This is due to the black area of the input region image in Figure 5. For example, when the input region image is input to the FCN structure, there are completely black feature maps among the feature maps extracted from the first convolution layer. Therefore, in the proposed method, the trained FCN model is pruned using the pruning function. Using the pruning function, the parameters that extract black feature maps as shown in Figure 5 are removed from the proposed structure. After fine-tuning the pruned architecture,

the expected final output image can be obtained using the structure as shown in Figure 6. The structure of the PFCN is shown in Tables 8 and 9 in Appendix.

C. DIFFERENCES BETWEEN FCN AND PFCN

The PFCN architectures proposed in this study and the existing FCN architectures have the following 3 differences.

PFCN architectures have a smaller number of channels than FCN architectures.

PFCN architectures have a smaller number of parameters than FCN architectures.

PFCN architectures are optimized versions of FCN architectures. The optimization operation is conducted by removing low effective parameters using a pruning.

V. POST PROCESSING

Moreover, a masking operation is performed using the output image obtained by the PFCN at the post-processing step. When performing the masking operation, the thermal reflection region of the output image in Figure 6 is processed with the input image in Figure 5 as in Equation (1) to obtain the final output image in Figure 6.

$$Img_{final\ output} = Img_{input} \circ Img_{mask} + Img_{output} \quad (1)$$

In Equation (1), Img_{input}, Img_{mask}, Img_{output}, and $Img_{final\ output}$ are the input image, mask image, output image generated by the PFCN, and final output image, respectively. More specifically, the input image in Figure 5 is Img_{input}, the output image obtained by the PFCN in Figure 6 is Img_{output}, and the final output image is $Img_{final\ output}$. The pixel values of Img_{mask} are either 0 or 1 as shown in Figure 7, whereas the pixel values of the region of interest (ROI) in Img_{mask} are 0 and those of the background are 1 as shown in Figure 7. Moreover, the operator ($\circ$) is the Hadamard product (element-wise multiplication) [50], whereas the operator ($+$) is matrix addition.

VI. EXPERIMENTAL RESULTS

A. DESCRIPTION OF EXPERIMENTAL SETUP AND DATABASES

The database [12] used in this study consists of thermal images of objects at close and far distances in both dark and bright environments. The database was collected in both indoor and outdoor environments. Furthermore, the database also includes visible light images. The details of the database are shown in Tables 3 and 4, and Figures 8–11 in our previous work [12]. Figure 8 shows the examples of the images in the database. The experiment was conducted as a two-fold cross validation. Specifically, half of the data were used for training, whereas the remaining data were used for testing. Then, the training data and testing data were switched, and the experiment was repeated. The results obtained accordingly were then used to determine the average testing accuracy.

The training and testing of the algorithm proposed in this study were conducted with a desktop computer. The desktop computer is equipped with an NVIDIA graphics card (NVIDIA GeForce GTX TITAN X [51]), Intel CPU (core i7-6700 CPU @ 3.40GHz (8 CPUs)), and RAM (32 GB). The method proposed in this paper was implemented using Python-based Keras application programming interface (API) with TensorFlow backend engine [52] and OpenCV library [53].

TABLE 2. Comparison of accuracies of the proposed methods with those of the state-of-the-art methods.

Method	PSNR	SNR	SSIM
PLN [10]	11.40	2.31	0.092
CycleGAN [44]	11.42	2.34	0.096
SegNet-based removal [43]	24.71	15.63	0.939
Mask R-CNN + CycleGAN [44]	24.45	15.36	0.949
Mask R-CNN-based removal [12]	31.43	22.35	0.973
FCN_V1 [46]	10.55	1.60	0.086
FCN_V2 [46]	10.79	1.81	0.088
Proposed method (PFCN_V1)	**32.88**	**22.72**	**0.978**
Proposed method (PFCN_V2)	**32.95**	**22.89**	**0.981**

TABLE 3. Comparison of accuracies of the proposed method with those of the state-of-the-art methods with the open database.

Method	PSNR	SNR	SSIM
PLN [10]	14.34	1.99	0.055
CycleGAN [44]	14.03	1.67	0.052
SegNet-based removal [43]	28.18	15.82	0.960
Mask R-CNN + CycleGAN [44]	26.57	14.22	0.959
Mask R-CNN-based removal [12]	36.58	24.23	0.983
FCN_V1 [46]	12.58	1.09	0.045
FCN_V2 [46]	12.73	1.16	0.051
Proposed method (PFCN_V1)	**36.99**	**25.01**	**0.984**
Proposed method (PFCN_V2)	**37.23**	**25.76**	**0.988**

B. TRAINING OF PFCN MODELS

When training the proposed models, the image size, batch-size, training epoch, loss, learning rate, and optimizer are set to 224 × 224 × 1, 1, 1000, MSE (mean squared error [54]), 0.0001, and adaptive moment estimation methods (Adam) [55], respectively. MSE loss is calculated between the pixel of ground-truth image and that of restored image by PFCN as shown in Equation (2). The larger MSE loss becomes, the larger penalty is assigned to the updated weights of PFCN whereas the smaller MSE loss becomes, the larger reward is given to the updated weights, which confirms the training convergence of PFCN.

The PFCN obtained after pruning the trained FCN was fine-tuned again with 100 epochs. Figure 9 shows the training loss of each method as the number of epochs increases. As the number of epochs increased, the training loss of both methods converged. In general, for a cycle-consistent adversarial network (CycleGAN)-based method [44], unpaired reference data are used for training, whereas in the proposed method, ground-truth data for input are used. Training was performed using the paired dataset of the input and ground-truth as shown in Figure 10.

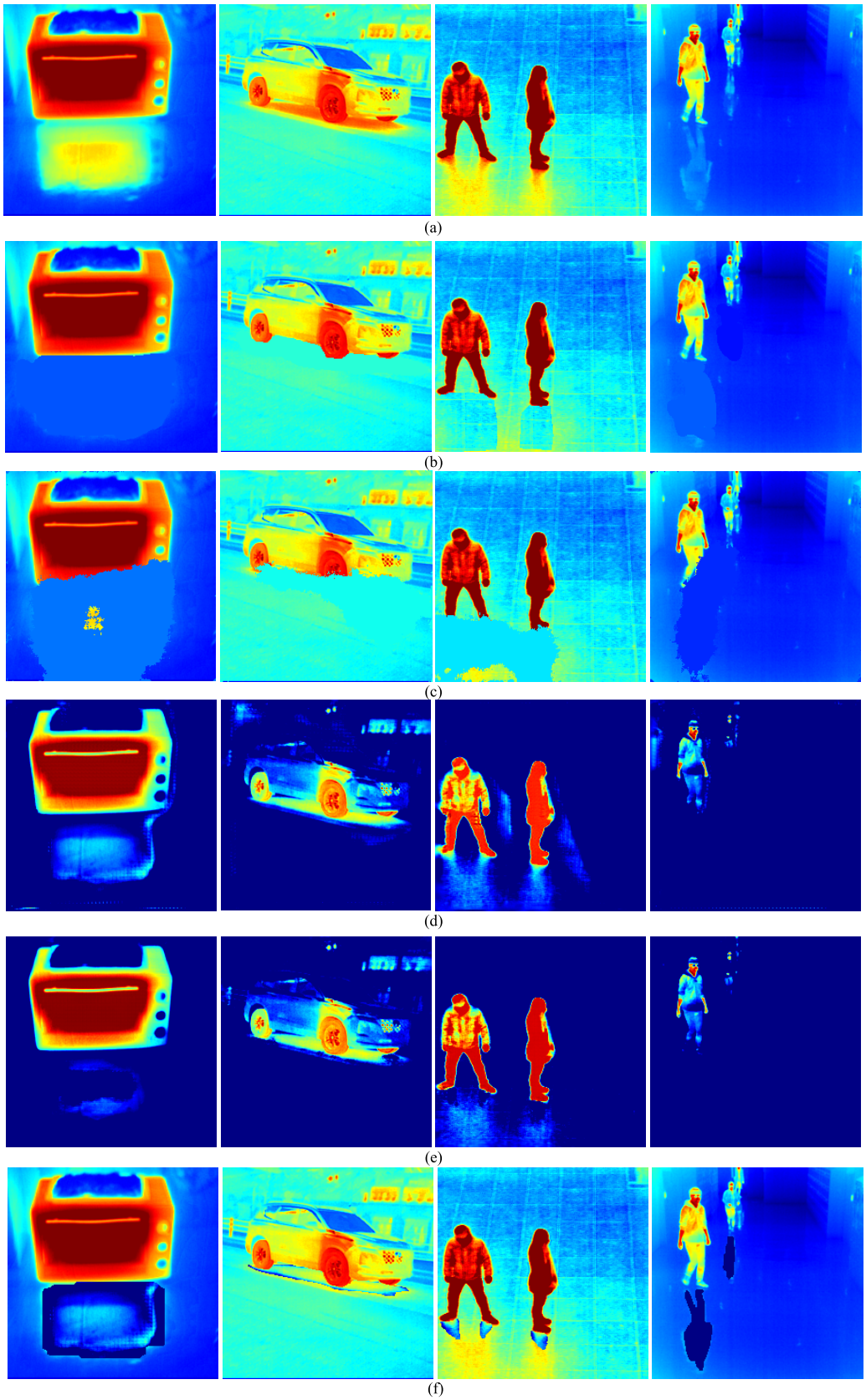

Figure 11. Examples of results of thermal reflection removal. (a) Original images; (b) ground-truth images; (c) SegNet-based removal method; (d) CycleGAN; (e) PLN; (f) Mask R-CNN + CycleGAN; (g) Mask R-CNN-based removal method; (h) FCN_V1; (i) FCN_V2; (j) the proposed method (PFCN_V1); (k) the proposed method (PFCN_V2).

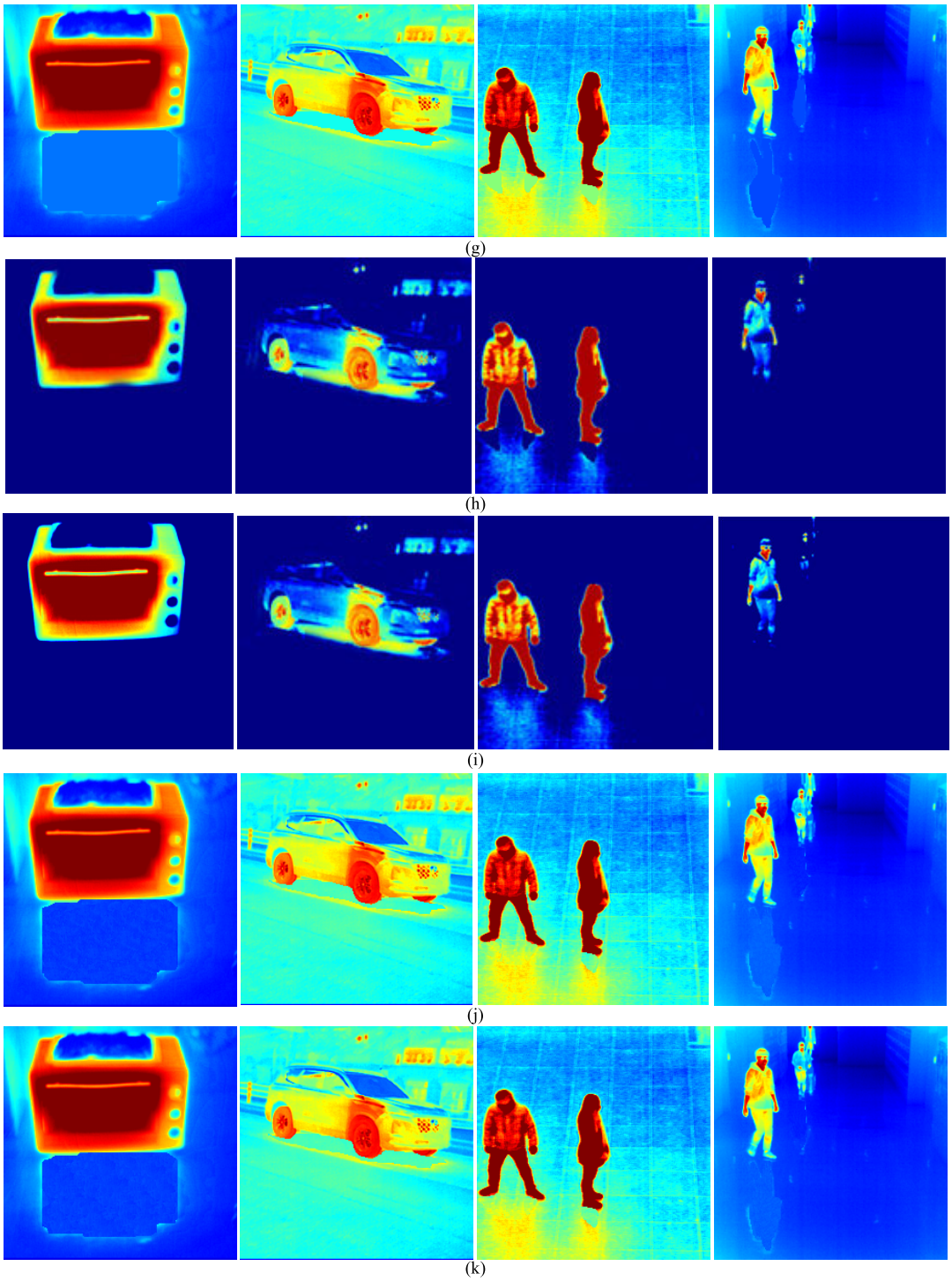

Figure 11. *(Continued.)* **Examples of results of thermal reflection removal. (a) Original images; (b) ground-truth images; (c) SegNet-based removal method; (d) CycleGAN; (e) PLN; (f) Mask R-CNN + CycleGAN; (g) Mask R-CNN-based removal method; (h) FCN_V1; (i) FCN_V2; (j) the proposed method (PFCN_V1); (k) the proposed method (PFCN_V2).**

C. TESTING

1) TESTING RESULTS OF THERMAL REFLECTION REMOVAL

In this section, the comparison results of the proposed method and the state-of-the-art methods are provided. For the comparison, the accuracies of seven types of methods i.e., CycleGAN [44], PLN [10], Mask R-CNN + PLN [12], SegNet [43]-based removal method [12], Mask R-CNN [18]-based removal method [12], FCN_V1 [46], and FCN_V2 [46] are compared with the accuracy of the method proposed in this study. Based on the original parameters provided by

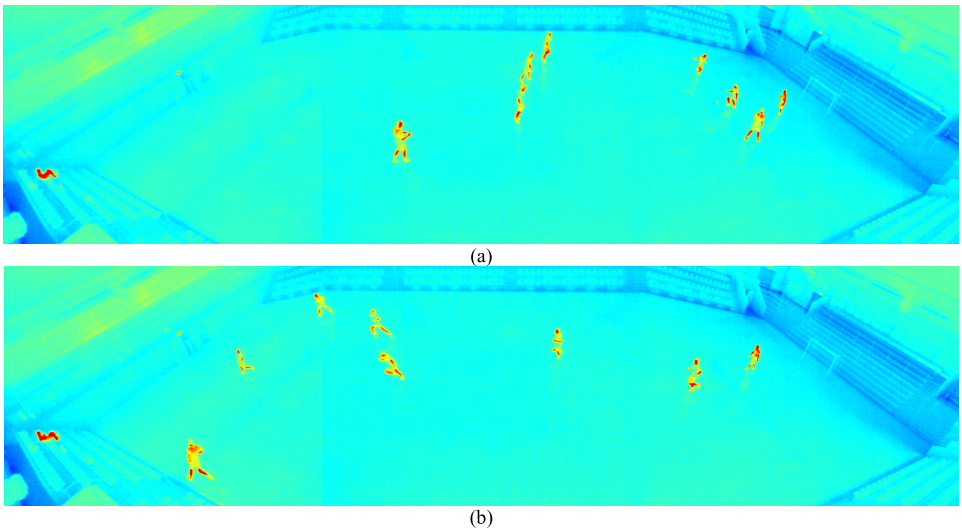

Figure 12. Example images of the open database.

authors, the optimal parameters of these seven types of methods were obtained by the further procedure of fine-tuning with the training dataset of our experimental data.

For fair comparisons, same training and testing data were used for both the previous methods and our method. For measuring the accuracy, the similarities between the ground-truth image (GT(i, j)) in which thermal reflection was manually removed and the image (Out(i, j)) in which thermal reflection was automatically removed by the algorithm were compared. Three kinds of metrics as in Equations (2)–(4) and the structural similarity index (SSIM) [56] were used to measure the accuracy.

$$\mathrm{MSE} = \frac{\left(\sqrt{\sum_{j=1}^{M}\sum_{i=1}^{N}(\mathrm{GT}\,(i,j) - \mathrm{Out}\,(i,\mathrm{j}))^2}\right)^2}{MN} \tag{2}$$

$$\mathrm{SNR} = 10log_{10}\left(\frac{\left(\frac{\sum_{j=1}^{M}\sum_{i=1}^{N}(\mathrm{GT}(i,j))^2}{MN}\right)}{MSE}\right) \tag{3}$$

$$\mathrm{PSNR} = 10log_{10}\left(\frac{255^2}{MSE}\right) \tag{4}$$

where M and N represent the image width and height, respectively. SNR and PSNR are the signal-to-noise ratio [57] and the peak-signal-to-noise ratio [58], respectively. Equation (5) expresses the mathematical formula of SSIM.

$$SSIM = \frac{(2\mu_r\mu_o + S1)(2\sigma_{ro} + S2)}{(\mu_r^2+\mu_o^2 + S1)(\sigma_r^2+\sigma_o^2 + S2)} \tag{5}$$

μ_o and σ_o represent the mean and standard deviation of the pixel values of a ground-truth image, respectively, μ_r and σ_r represent the mean and standard deviation of the pixel values of the restored image, respectively, and σ_{ro} is the covariance of the two images. $S1$ and $S2$ are positive constants set so that the denominator does not become zero.

Table 2 shows the comparison of the measured accuracies. A greater value in Table 2 indicates higher accuracy. As shown in Table 2, the accuracy of removing thermal reflection was the highest for all the methods proposed in this study.

Figure 11 shows the results of removal of the thermal reflection by the proposed method and by the state-of-the-art methods. The ground-truth image with thermal reflection removed manually is shown in Figure 11 (b). The results of the proposed method are shown in Figures 11 (j) and (k), whereas those of the SegNet-based removal [43], CycleGAN [44], PLN [10], Mask R-CNN + CycleGAN [46], Mask R-CNN-based removal [12], FCN_V1 [46], and FCN_V2 [46] are shown in Figures 11 (c)–(i). As shown in Figure 11, the accuracy of the removal of thermal reflection by the PFCN method proposed in this study is the highest for all the cases.

2) TESTING RESULTS USING OPEN DATABASE

Additional experiments were conducted using the existing thermal image open database to check whether the proposed method can be applied in other types of environments. There are several existing open thermal image databases [59]–[69]. However, there are few open databases having thermal images with thermal reflection. Therefore, in this study, we conducted additional experiments using an open database (thermal soccer dataset [59]) having thermal reflection as shown in Figure 12. The comparison results of the accuracy of all the methods are shown in Table 3. In the experiment conducted with the open database, the accuracy of the proposed method was higher than that of the state-of-the-art methods.

Figure 13 (a)–(k) show the source input image having thermal reflection, the ground-truth image with

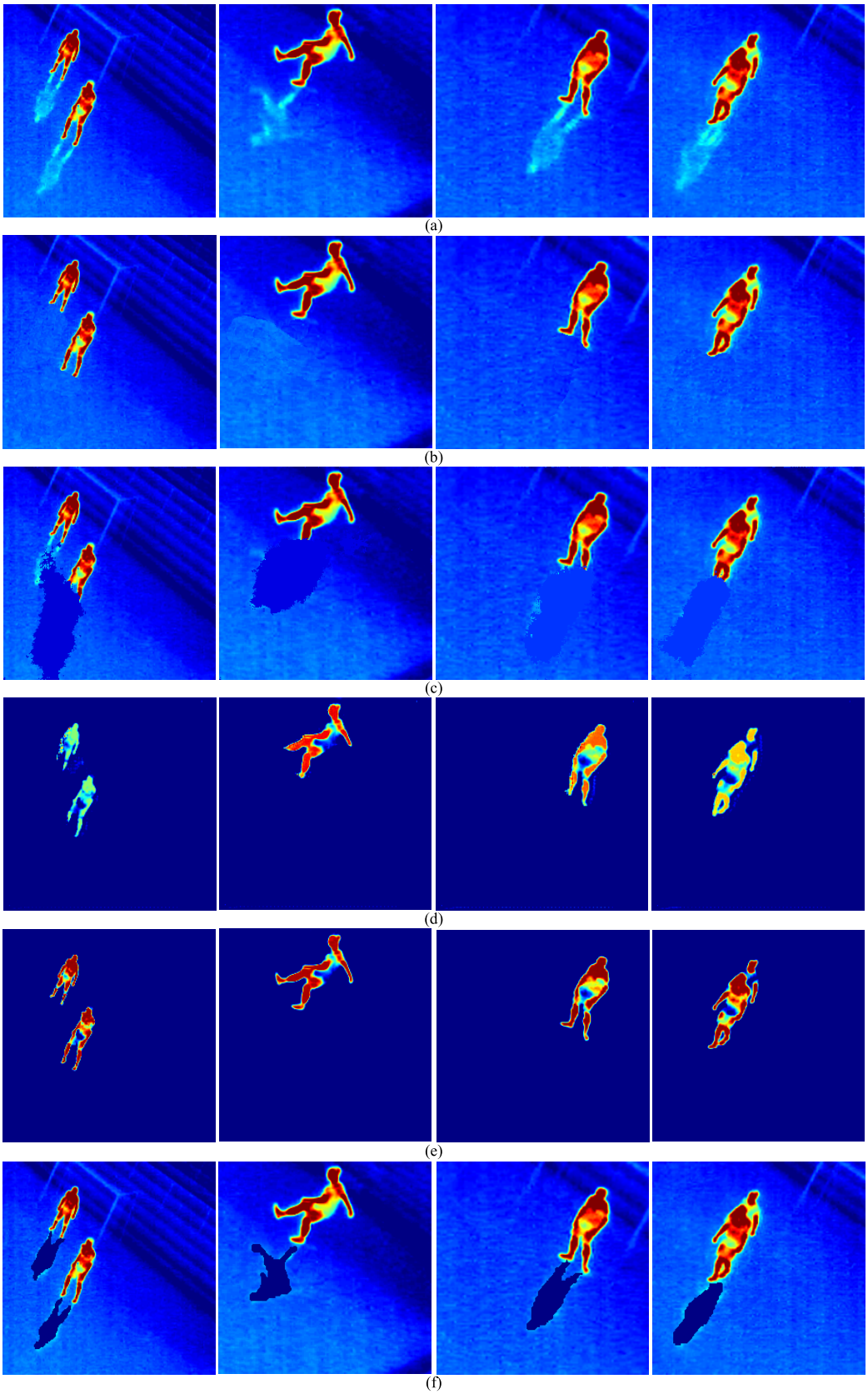

Figure 13. Examples of result images with the open database. (a) Original images; (b) ground-truth images; (c) SegNet-based removal method; (d) CycleGAN; (e) PLN; (f) Mask R-CNN + CycleGAN; (g) Mask R-CNN-based removal method; (h) FCN_V1; (i) FCN_V2; (j) the proposed method (PFCN_V1); (k) the proposed method (PFCN_V2).

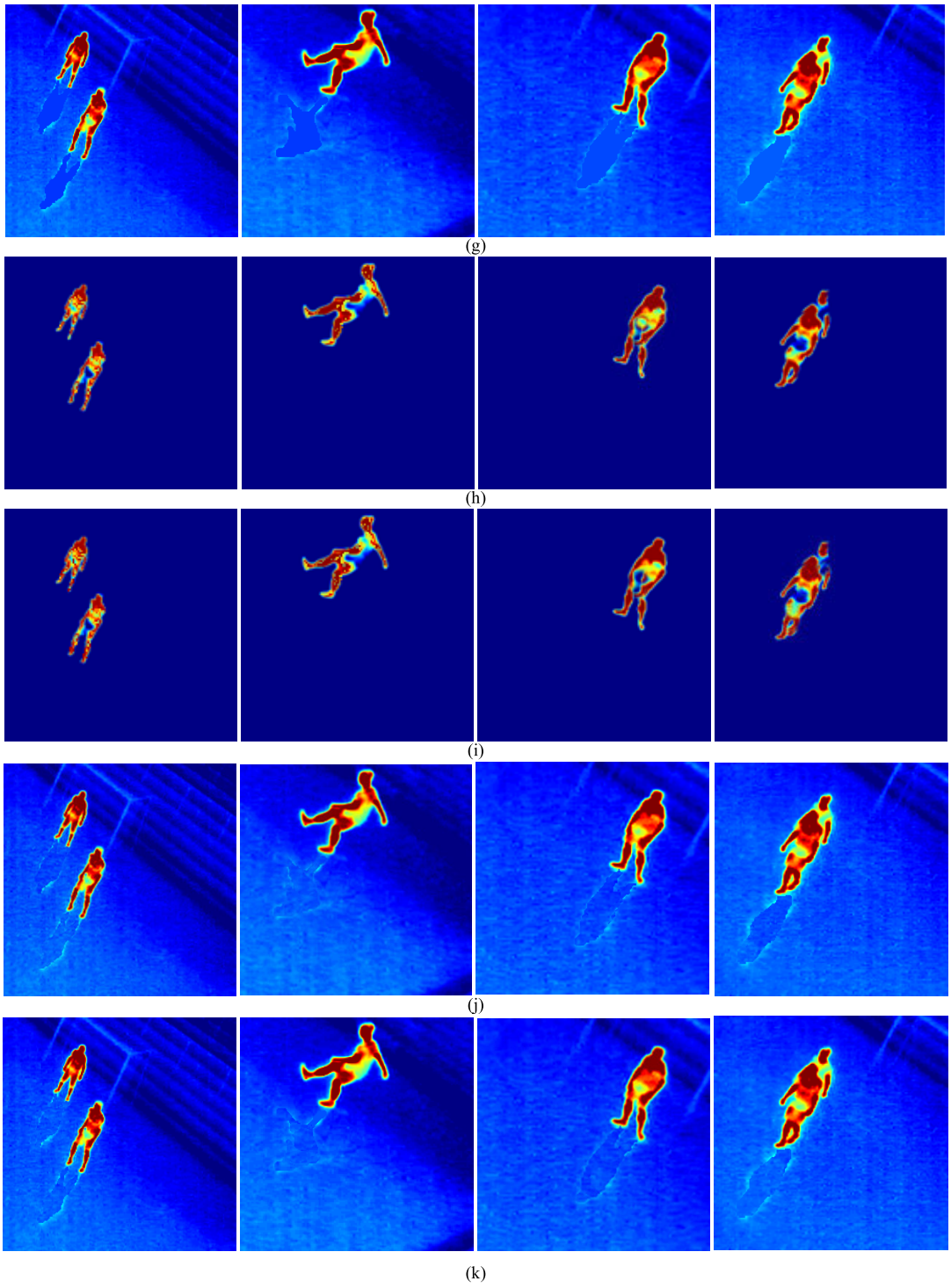

Figure 13. ***(Continued.)*** **Examples of result images with the open database. (a) Original images; (b) ground-truth images; (c) SegNet-based removal method; (d) CycleGAN; (e) PLN; (f) Mask R-CNN + CycleGAN; (g) Mask R-CNN-based removal method; (h) FCN_V1; (i) FCN_V2; (j) the proposed method (PFCN_V1); (k) the proposed method (PFCN_V2).**

thermal reflection manually removed, and the results of thermal reflection removed by all the methods. As shown in Figure 13 (j) and (k), images for which thermal reflection was removed by the proposed method were most similar to the ground-truth image.

In this research, we did not measure the accuracy of the detection, and there is no error of false acceptance and rejection. Instead, we measured the quality of restored image by our method by calculating the similarity between our restored and ground-truth images based on Equations (2) ∼ (5).

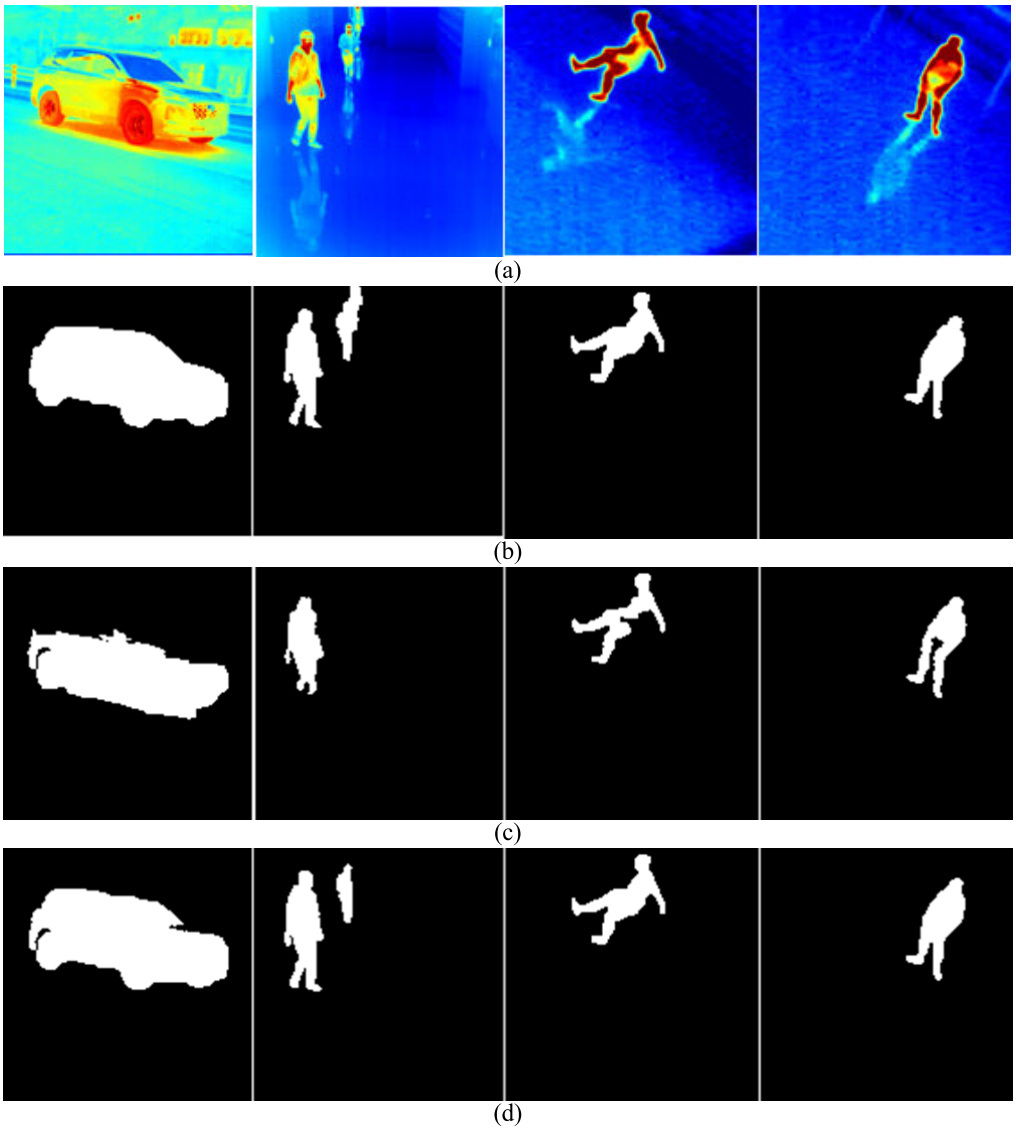

Figure 14. Example of detection results by Mask R-CNN. (a) Original images; (b) ground truth images; (c) results obtained for PLN;(d) results obtained for our method.

TABLE 4. Comparison of processing speeds of the FCN and proposed PFCN.

Method	Processing time of one image (ms)	Frames per second
FCN_V1 [46]	29	34.48
FCN_V2 [46]	393	2.54
PLN [10]	26	38.46
CycleGAN [44]	29	34.48
SegNet-based removal [43]	112	8.93
Proposed method (PFCN_V1)	22	45.45
Proposed method (PFCN_V2)	360	2.78

As shown in Tables 2, 3, and Figures 11 ∼ 13, the similarity between our restored and ground-truth images is higher than those between the restored image by previous methods [10], [12], [43], [44], [46] and the ground-truth image, which confirms the superiority of our method for thermal reflection removal.

TABLE 5. Comparison of object detection accuracy by our method with the previous method.

Metrics	PLN [10]	Our method
Recall	0.81	0.93
Precision	0.68	0.98
GlobalACC	0.84	0.99
F1	0.74	0.95
Jaccard	0.48	0.94

3) COMPARISONS OF PROCESSING SPEED OF THE FCN AND PROPOSED PFCN

In the next experiment, we compared the processing speed between FCN and the proposed PFCN for an input image. The experiment was performed on the desktop computer described in section *V.A*. As shown in Table 4, the proposed PFCN was faster than the conventional FCN.

TABLE 6. Description of the structure of the proposed FCN_V1.

Layer number	Layer type	Size of feature map (height×width×channel)	Number of filters	Filter size	Stride	Number of parameters	Layer connection (connected to)
0	input layer_1	224×224×1	0	n/a	n/a	0	n/a
1	input layer_2	224×224×1	0	n/a	n/a	0	n/a
			Downsampling				
2	conv2d_1	222×222×64	64	3×3	1×1	640	input layer_1
3	conv2d_2	222×222×64	64	3×3	1×1	640	input layer_2
4	concatenate_1	222×222×128	0	n/a	n/a	0	conv2d_1 & conv2d_2
5	conv2d_3	220×220×64	64	3×3	1×1	36,928	conv2d_1
6	conv2d_4	220×220×64	64	3×3	1×1	36,928	conv2d_2
7	conv2d_5	220×220×64	64	3×3	1×1	73,792	concatenate_1
8	concatenate_2	220×220×192	0	n/a	n/a	0	conv2d_3 & conv2d_4 & conv2d_5
9	conv2d_6	215×215×64	64	6×6	1×1	147,520	conv2d_3
10	conv2d_7	215×215×64	64	6×6	1×1	147,520	conv2d_4
11	conv2d_8	215×215×64	64	6×6	1×1	442,432	concatenate_2
12	concatenate_3	215×215×192	0	n/a	n/a	0	conv2d_6 & conv2d_7 & conv2d_8
13	conv2d_9	106×106×64	64	4×4	2×2	65,600	conv2d_6
14	conv2d_10	106×106×64	64	4×4	2×2	65,600	conv2d_7
15	conv2d_11	106×106×64	64	4×4	2×2	196,672	concatenate_3
16	concatenate_4	106×106×192	0	n/a	n/a	0	conv2d_9 & conv2d_10 & conv2d_11
17	conv2d_12	104×104×64	64	3×3	1×1	110,656	concatenate_4
			Upsampling				
18	T_conv2d_1	106×106×64	64	3×3	1×1	36,928	conv2d_12
19	T_conv2d_2	108×108×64	64	3×3	1×1	36,928	conv2d_9
20	T_conv2d_3	108×108×64	64	3×3	1×1	36,928	conv2d_10
21	T_conv2d_4	108×108×64	64	3×3	1×1	36,928	T_conv2d_1
22	concatenate_5	108×108×192	0	n/a	n/a	0	T_conv2d_2 & T_conv2d_3 & T_conv2d_4
23	T_conv2d_5	217×217×64	64	3×3	2×2	36,928	T_conv2d_2
24	T_conv2d_6	217×217×64	64	3×3	2×2	36,928	T_conv2d_3
25	T_conv2d_7	217×217×64	64	3×3	2×2	110,656	concatenate_5
26	concatenate_6	217×217×192	0	n/a	n/a	0	T_conv2d_5 & T_conv2d_6 & T_conv2d_7
27	T_conv2d_8	220×220×64	64	4×4	1×1	65,600	T_conv2d_5
28	T_conv2d_9	220×220×64	64	4×4	1×1	65,600	T_conv2d_6
29	T_conv2d_10	220×220×64	64	4×4	1×1	196,672	concatenate_6
30	concatenate_7	220×220×192	0	n/a	n/a	0	T_conv2d_8 & T_conv2d_9 & T_conv2d_10
31	T_conv2d_11	222×222×64	64	3×3	1×1	110,656	concatenate_7
32	T_conv2d_12	224×224×64	64	3×3	1×1	36,928	T_conv2d_11
33	concatenate_8	224×224×65	0	n/a	n/a	0	input layer_1 & T_conv2d_12
34	conv2d_13	224×224×1	1	3×3	1×1	586	concatenate_8
			Total number of trainable parameters: 2,133,194				

4) COMPARISONS OF OBJECT DETECTION ACCURACY BY OUR METHOD WITH THE PREVIOUS METHOD

Although the generated background by PLN method seems to be more desired than that by our method, there exist lots of errors that the pixels inside of object are incorrectly recognized as backgrounds as shown in Figures 11 (e) and 13 (e) compared to the ground-truth images of Figures 11 (b) and 13 (b). Nevertheless, these errors are much reduced in the result image by our method as shown in Figures 11 (j), (k) and 13 (j), (k). To confirm these observations, we performed the additional experiments of object detection by Mask R-CNN [18] with the result images by PLN method and our method. Accuracy was measured based on five metrics of recall, precision, GlobalACC, F1 score, and Jaccard similarity [12]. Accuracy (ACC) is the percentage of correctly classified pixels for each class as shown in Equation (6). Here, #TP, #TN, #FP, and #FN represent the number of true positive data, true negative data, false-positive data, and false-negative data, respectively. The positive and negative data represent the pixels of the object and the background, respectively. TP represents the data that are positive and correctly classified as positive data whereas TN means data that are negative and correctly classified as negative data. FP represents data that are negative but incorrectly classified as positive data, whereas FN represents data that are positive, but incorrectly classified as negative data.

$$\text{Accuracy (ACC)} = \frac{\#\text{TP} + \#\text{TN}}{\#\text{ TP+\# TN+\# FP+\# FN}} \tag{6}$$

The global accuracy (GlobalACC) is defined as the ratio of correctly classified pixels to the total number of pixels. The F1 score is calculated based on precision and recall as shown in Equation (7). In this case, precision is calculated as #TP/(#TP+#FP), whereas recall is calculated as #TP/(#TP+#FN).

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{7}$$

For a class, the intersection over union (IoU) is the ratio of the correctly classified pixels to the total number of ground truth and predicted pixels in that class. The IoU score is also known as the Jaccard similarity, and it can be calculated with two sets X and Y as shown in Equation (8). In this case, X is the ground truth pixel of the object whereas Y is the detected

TABLE 7. Description of the structure of the proposed FCN_V2.

Layer number	Layer type	Size of feature map (height×width×channel)	Number of filters	Number of parameters	Layer connection (connected to)
0	input layer_1	224×224×1	0	0	n/a
1	input layer_2	224×224×1	0	0	n/a
		Downsampling			
2	conv2d_1	222×222×32	32	320	input layer_1
3	conv2d_2	222×222×32	32	320	input layer_2
4	concatenate_1	222×222×64	0	0	conv2d_1 & conv2d_2
5	conv2d_3	220×220×32	32	9,248	conv2d_1
6	conv2d_4	220×220×32	32	18,464	concatenate_1
7	concatenate_2	220×220×64	0	0	conv2d_3 & conv2d_4
8	conv2d_5	218×218×64	64	18,496	conv2d_3
9	conv2d_6	218×218×64	64	36,928	concatenate_2
10	concatenate_3	218×218×128	0	0	conv2d_5 & conv2d_6
11	conv2d_7	216×216×64	64	36,928	conv2d_5
12	conv2d_8	216×216×64	64	73,792	concatenate_3
13	concatenate_4	216×216×128	0	0	conv2d_7 & conv2d_8
14	conv2d_9	214×214×128	128	73,856	conv2d_7
15	conv2d_10	214×214×128	128	147,584	concatenate_4
16	concatenate_5	214×214×256	0	0	conv2d_9 & conv2d_10
17	conv2d_11	212×212×128	128	147,584	conv2d_9
18	conv2d_12	212×212×128	128	295,040	concatenate_5
19	concatenate_6	212×212×256	0	0	conv2d_11 & conv2d_12
20	conv2d_13	210×210×256	256	295,168	conv2d_11
21	conv2d_14	210×210×256	256	590,080	concatenate_6
22	concatenate_7	210×210×512	0	0	conv2d_13 & conv2d_14
23	conv2d_15	208×208×256	256	590,080	conv2d_13
24	conv2d_16	208×208×256	256	1,179,904	concatenate_7
25	concatenate_8	208×208×512	0	0	conv2d_15 & conv2d_16
26	conv2d_17	206×206×512	512	1,180,160	conv2d_15
27	conv2d_18	206×206×512	512	2,359,808	concatenate_8
28	concatenate_9	206×206×1024	0	0	conv2d_17 & conv2d_18
29	conv2d_19	204×204×512	512	2,359,808	conv2d_17
30	conv2d_20	204×204×512	512	4,719,104	concatenate_9
31	concatenate_10	204×204×1024	0	0	conv2d_19 & conv2d_20
		Upsampling			
32	T_conv2d_1	206×206×512	512	2,359,808	conv2d_19
33	T_conv2d_2	206×206×512	512	4,719,104	concatenate_10
34	concatenate_11	206×206×1024	0	0	T_conv2d_1 & T_conv2d_2
35	T_conv2d_3	208×208×512	512	2,359,808	T_conv2d_1
36	T_conv2d_4	208×208×512	512	4,719,104	concatenate_11
37	concatenate_12	208×208×1024	0	0	T_conv2d_3 & T_conv2d_4
38	T_conv2d_5	210×210×256	256	1,179,904	T_conv2d_3
39	T_conv2d_6	210×210×256	256	2,359,552	concatenate_12
40	concatenate_13	210×210×512	0	0	T_conv2d_5 & T_conv2d_6
41	T_conv2d_7	212×212×256	256	590,080	T_conv2d_5
42	T_conv2d_8	212×212×256	256	1,179,904	concatenate_13
43	concatenate_14	212×212×512	0	0	T_conv2d_7 & T_conv2d_8
44	T_conv2d_9	214×214×128	128	295,040	T_conv2d_7
45	T_conv2d_10	214×214×128	128	589,952	concatenate_14
46	concatenate_15	214×214×256	0	0	T_conv2d_9 & T_conv2d_10
47	T_conv2d_11	216×216×128	128	147,584	T_conv2d_9
48	T_conv2d_12	216×216×128	128	295,040	concatenate_15
49	concatenate_16	216×216×256	0	0	T_conv2d_11 & T_conv2d_12
50	T_conv2d_13	218×218×64	64	73,792	T_conv2d_11
51	T_conv2d_14	218×218×64	64	147,520	concatenate_16
52	concatenate_17	218×218×128	0	0	T_conv2d_13 & T_conv2d_14
53	T_conv2d_15	220×220×64	64	36,928	T_conv2d_13
54	T_conv2d_16	220×220×64	64	73,792	concatenate_17
55	concatenate_18	220×220×128	0	0	T_conv2d_15 & T_conv2d_16
56	T_conv2d_17	222×222×32	32	18,464	T_conv2d_15
57	T_conv2d_18	222×222×32	32	36,896	concatenate_18
58	concatenate_19	222×222×64	0	0	T_conv2d_17 & T_conv2d_18
59	T_conv2d_19	224×224×1	1	577	concatenate_19
		Total number of trainable parameters: 35,315,521			

TABLE 8. Description of the structure of the proposed PFCN_V1.

Layer number	Layer type	Size of feature map (height×width×channel)	Number of filters	Filter size	Stride	Number of parameters	Layer connection (connected to)
0	input layer_1	224×224×1	0	n/a	n/a	0	n/a
1	input layer_2	224×224×1	0	n/a	n/a	0	n/a
			Downsampling				
2	conv2d_1	222×222×64	64	3×3	1×1	640	input layer_1
3	conv2d_2	222×222×64	45	3×3	1×1	450	input layer_2
4	concatenate_1	222×222×128	0	n/a	n/a	0	conv2d_1 & conv2d_2
5	conv2d_3	220×220×64	64	3×3	1×1	36,928	conv2d_1
6	conv2d_4	220×220×64	48	3×3	1×1	19,488	conv2d_2
7	conv2d_5	220×220×64	64	3×3	1×1	62,848	concatenate_1
8	concatenate_2	220×220×192	0	n/a	n/a	0	conv2d_3 & conv2d_4 & conv2d_5
9	conv2d_6	215×215×64	64	6×6	1×1	147,520	conv2d_3
10	conv2d_7	215×215×64	47	6×6	1×1	81,263	conv2d_4
11	conv2d_8	215×215×64	64	6×6	1×1	405,568	concatenate_2
12	concatenate_3	215×215×192	0	n/a	n/a	0	conv2d_6 & conv2d_7 & conv2d_8
13	conv2d_9	106×106×64	64	4×4	2×2	65,600	conv2d_6
14	conv2d_10	106×106×64	64	4×4	2×2	48,192	conv2d_7
15	conv2d_11	106×106×64	46	4×4	2×2	128,846	concatenate_3
16	concatenate_4	106×106×192	0	n/a	n/a	0	conv2d_9 & conv2d_10 & conv2d_11
17	conv2d_12	104×104×64	64	3×3	1×1	100,288	concatenate_4
			Upsampling				
18	T_conv2d_1	106×106×64	64	3×3	1×1	36,928	conv2d_12
19	T_conv2d_2	108×108×64	64	3×3	1×1	36,928	conv2d_9
20	T_conv2d_3	108×108×64	64	3×3	1×1	36,928	conv2d_10
21	T_conv2d_4	108×108×64	64	3×3	1×1	36,928	T_conv2d_1
22	concatenate_5	108×108×192	0	n/a	n/a	0	T_conv2d_2 & T_conv2d_3 & T_conv2d_4
23	T_conv2d_5	217×217×64	64	3×3	2×2	36,928	T_conv2d_2
24	T_conv2d_6	217×217×64	64	3×3	2×2	36,928	T_conv2d_3
25	T_conv2d_7	217×217×64	64	3×3	2×2	110,656	concatenate_5
26	concatenate_6	217×217×192	0	n/a	n/a	0	T_conv2d_5 & T_conv2d_6 & T_conv2d_7
27	T_conv2d_8	220×220×64	64	4×4	1×1	65,600	T_conv2d_5
28	T_conv2d_9	220×220×64	64	4×4	1×1	65,600	T_conv2d_6
29	T_conv2d_10	220×220×64	64	4×4	1×1	196,672	concatenate_6
30	concatenate_7	220×220×192	0	n/a	n/a	0	T_conv2d_8 & T_conv2d_9 & T_conv2d_10
31	T_conv2d_11	222×222×64	64	3×3	1×1	110,656	concatenate_7
32	T_conv2d_12	224×224×64	64	3×3	1×1	36,928	T_conv2d_11
33	concatenate_8	224×224×65	0	n/a	n/a	0	input layer_1 & T_conv2d_12
34	conv2d_13	224×224×1	1	3×3	1×1	586	concatenate_8
			Total number of trainable parameters: 1,905,897				

pixel of object.

$$\text{Jaccard}\,(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{8}$$

As shown in Table 5, our restored image + Mask R-CNN showed a higher detection accuracy compared to that by PLN + Mask R-CNN for all five metrics. In Figure 14, the results of PLN + Mask R-CNN and our method + Mask R-CNN are compared. As shown in this figure, it is evident that the detection accuracy of our method + Mask R-CNN is higher than that of PLN + Mask R-CNN.

VII. CONCLUSIONS

In this study, various methods for removing thermal reflection in thermal images of diverse objects were proposed. Specifically, the new method using PFCN, which considers the heat information of nearby ground and walls, is proposed when an image is transformed only within the region where thermal reflection is detected. In the PFCN model, unnecessary channels and parameters are removed from the existing FCN structure through training, and the performance of thermal reflection removal is improved despite having fewer parameters than the FCN model. The proposed method was compared against various state-of-the-art methods (SegNet-based removal, CycleGAN, PLN, Mask R-CNN + CycleGAN, Mask R-CNN-based removal, FCN_V1, FCN_V2), and thus, the accuracy of removing thermal reflection using the proposed method was proven to be higher than that of the state-of-the-art methods when experiments were conducted using our database and additional open databases. As shown in [33], PLN and CycleGAN-based method which were compared in our experiment were used for image inpainting, and we can regard these methods as the classical image inpaining algorithm. As shown in Tables 2, 3, 5, and Figures 11 ∼ 14, our method shows the higher accuracy than these image inpaining methods.

The reason why there remains the border around the detected reflection is that the generated mask by our PFCN is a little smaller than the ground-truth reflection area. Nevertheless, it does not give much influence on the detection accuracy of object as shown in Table 5 and Figure 14.

To solve this problem of remained border, we can adjust the output threshold of PFCN, which produces the larger mask and reduces the consequent border around the detected reflections. We would research about this method in future

TABLE 9. Description of the structure of the proposed PFCN_V2.

Layer number	Layer type	Size of feature map (height×width×channel)	Number of filters	Number of parameters	Layer connection (connected to)
0	input layer_1	224×224×1	0	0	n/a
1	input layer_2	224×224×1	0	0	n/a
		Downsampling			
2	conv2d_1	222×222×32	32	320	input layer_1
3	conv2d_2	222×222×32	24	240	input layer_2
4	concatenate_1	222×222×64	0	0	conv2d_1 & conv2d_2
5	conv2d_3	220×220×32	32	9,248	conv2d_1
6	conv2d_4	220×220×32	28	14,140	concatenate_1
7	concatenate_2	220×220×64	0	0	conv2d_3 & conv2d_4
8	conv2d_5	218×218×64	64	18,496	conv2d_3
9	conv2d_6	218×218×64	54	29,214	concatenate_2
10	concatenate_3	218×218×128	0	0	conv2d_5 & conv2d_6
11	conv2d_7	216×216×64	64	36,928	conv2d_5
12	conv2d_8	216×216×64	50	53,150	concatenate_3
13	concatenate_4	216×216×128	0	0	conv2d_7 & conv2d_8
14	conv2d_9	214×214×128	128	73,856	conv2d_7
15	conv2d_10	214×214×128	100	102,700	concatenate_4
16	concatenate_5	214×214×256	0	0	conv2d_9 & conv2d_10
17	conv2d_11	212×212×128	128	147,584	conv2d_9
18	conv2d_12	212×212×128	113	231,989	concatenate_5
19	concatenate_6	212×212×256	0	0	conv2d_11 & conv2d_12
20	conv2d_13	210×210×256	256	295,168	conv2d_11
21	conv2d_14	210×210×256	221	479,570	concatenate_6
22	concatenate_7	210×210×512	0	0	conv2d_13 & conv2d_14
23	conv2d_15	208×208×256	256	590,080	conv2d_13
24	conv2d_16	208×208×256	231	991,914	concatenate_7
25	concatenate_8	208×208×512	0	0	conv2d_15 & conv2d_16
26	conv2d_17	206×206×512	512	1,180,160	conv2d_15
27	conv2d_18	206×206×512	482	2,113,088	concatenate_8
28	concatenate_9	206×206×1024	0	0	conv2d_17 & conv2d_18
29	conv2d_19	204×204×512	512	2,359,808	conv2d_17
30	conv2d_20	204×204×512	512	4,580,864	concatenate_9
31	concatenate_10	204×204×1024	0	0	conv2d_19 & conv2d_20
		Upsampling			
32	T_conv2d_1	206×206×512	512	2,359,808	conv2d_19
33	T_conv2d_2	206×206×512	512	4,719,104	concatenate_10
34	concatenate_11	206×206×1024	0	0	T_conv2d_1 & T_conv2d_2
35	T_conv2d_3	208×208×512	512	2,359,808	T_conv2d_1
36	T_conv2d_4	208×208×512	512	4,719,104	concatenate_11
37	concatenate_12	208×208×1024	0	0	T_conv2d_3 & T_conv2d_4
38	T_conv2d_5	210×210×256	256	1,179,904	T_conv2d_3
39	T_conv2d_6	210×210×256	256	2,359,552	concatenate_12
40	concatenate_13	210×210×512	0	0	T_conv2d_5 & T_conv2d_6
41	T_conv2d_7	212×212×256	256	590,080	T_conv2d_5
42	T_conv2d_8	212×212×256	256	1,179,904	concatenate_13
43	concatenate_14	212×212×512	0	0	T_conv2d_7 & T_conv2d_8
44	T_conv2d_9	214×214×128	128	295,040	T_conv2d_7
45	T_conv2d_10	214×214×128	128	589,952	concatenate_14
46	concatenate_15	214×214×256	0	0	T_conv2d_9 & T_conv2d_10
47	T_conv2d_11	216×216×128	128	147,584	T_conv2d_9
48	T_conv2d_12	216×216×128	128	295,040	concatenate_15
49	concatenate_16	216×216×256	0	0	T_conv2d_11 & T_conv2d_12
50	T_conv2d_13	218×218×64	64	73,792	T_conv2d_11
51	T_conv2d_14	218×218×64	64	147,520	concatenate_16
52	concatenate_17	218×218×128	0	0	T_conv2d_13 & T_conv2d_14
53	T_conv2d_15	220×220×64	64	36,928	T_conv2d_13
54	T_conv2d_16	220×220×64	64	73,792	concatenate_17
55	concatenate_18	220×220×128	0	0	T_conv2d_15 & T_conv2d_16
56	T_conv2d_17	222×222×32	32	18,464	T_conv2d_15
57	T_conv2d_18	222×222×32	32	36,896	concatenate_18
58	concatenate_19	222×222×64	0	0	T_conv2d_17 & T_conv2d_18
59	T_conv2d_19	224×224×1	1	577	concatenate_19
		Total number of trainable parameters: 34,491,366			

work. A method for transforming low-resolution thermal images to high-resolution images will be examined in future research. Furthermore, a method for detecting the object, thermal reflection, and halo effect in thermal images and removing thermal reflection and halo effect will be studied as well.

APPENDIX

See Table 6–9.

REFERENCES

[1] L. St-Laurent, D. Prévost, and X. Maldague, "Thermal imaging for enhanced foreground—Background segmentation," in *Proc. Int. Conf. Quant. Infr. Thermography*, Padua, Italy, Jun. 2006, pp. 1–10.

[2] H. Dong, P. Neekhara, C. Wu, and Y. Guo, "Unsupervised image-to-image translation with generative adversarial networks," 2017, *arXiv:1701.02676*. [Online]. Available: http://arxiv.org/abs/1701.02676

[3] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," 2017, *arXiv:1703.00848*. [Online]. Available: http://arxiv.org/abs/1703.00848

[4] Z. Gan, L. Chen, W. Wang, Y. Pu, Y. Zhang, H. Liu, C. Li, and L. Carin, "Triangle generative adversarial networks," 2017, *arXiv:1709.06548*. [Online]. Available: http://arxiv.org/abs/1709.06548

[5] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," 2017, *arXiv:1711.09020*. [Online]. Available: http://arxiv.org/abs/1711.09020

[6] T. Kim, M. Cha, H. Kim, J. Kwon Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," 2017, *arXiv:1703.05192*. [Online]. Available: http://arxiv.org/abs/1703.05192

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv:1611.07004*. [Online]. Available: http://arxiv.org/abs/1611.07004

[8] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," 2016, *arXiv:1606.07536*. [Online]. Available: http://arxiv.org/abs/1606.07536

[9] S. Mo, M. Cho, and J. Shin, "InstaGAN: Instance-aware image-to-image translation," 2018, *arXiv:1812.10889*. [Online]. Available: http://arxiv.org/abs/1812.10889

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," 2016, *arXiv:1603.08155*. [Online]. Available: http://arxiv.org/abs/1603.08155

[11] S. Henke, D. Karstädt, K.-P. Möllmann, F. Pinno, and M. Vollmer, "Identification and suppression of thermal reflections in infrared thermal imaging," in *Proc. Inframation*, Las Vegas, CA, USA, Oct. 2004, pp. 287–298.

[12] G. Batchuluun, H. S. Yoon, D. T. Nguyen, T. D. Pham, and K. R. Park, "A study on the elimination of thermal reflections," *IEEE Access*, vol. 7, pp. 174597–174611, Dec. 2019.

[13] B. Zeise and B. Wagner, "Temperature correction and reflection removal in thermal images using 3D temperature mapping," in *Proc. 13th Int. Conf. Informat. Control, Autom. Robot.*, Lisbon, Portugal, Jul. 2016, pp. 158–165.

[14] N. Li, Y. Zhao, Q. Pan, and S. G. Kong, "Removal of reflections in LWIR image with polarization characteristics," *Opt. Express*, vol. 26, no. 13, pp. 16488–16504, Jun. 2018.

[15] FLIR Systems. (2019). *Uncooled Detectors for Thermal Imaging Cameras*. [Online]. Available: http://www.flirmedia.com/MMC/CVS/Appl_Stories/AS_0015_EN.pdf

[16] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 162–182, May 2007.

[17] J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, Cambridge, U.K., Aug. 2004, pp. 713–716.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2017, *arXiv:1703.06870*. [Online]. Available: http://arxiv.org/abs/1703.06870

[19] W. K. Wong, H. L. Lim, C. K. Loo, and W. S. Lim, "Home alone faint detection surveillance system using thermal camera," in *Proc. 2nd Int. Conf. Comput. Res. Develop.*, Kuala Lumpur, Malaysia, May 2010, pp. 747–751.

[20] C. Beyan, "Object tracking for surveillance applications using thermal and visible band video data fusion," M.S. thesis, Dept. Info. Sys., Middle East Tech. Univ., Ankara, Turkey, Dec. 2010.

[21] P. Kumar, A. Mittal, and P. Kumar, "Fusion of thermal infrared and visible spectrum video for robust surveillance," in *Proc. Int. Conf. Comput. Vis., Graph. Image Process.*, Madurai, India, Dec. 2006, pp. 528–539.

[22] P. Kumar, A. Mittal, and P. Kumar, "Study of robust and intelligent surveillance in visible and multi-modal framework," *Informatica*, vol. 31, pp. 447–461, Dec. 2007.

[23] D. Gangodkar, P. Kumar, and A. Mittal, "Segmentation of moving objects in visible and thermal videos," in *Proc. Int. Conf. Comput. Commun. Informat.*, Coimbatore, India, Jan. 2012, pp. 1–5.

[24] J. Lee, J.-S. Choi, E. Jeon, Y. Kim, T. Le, K. Shin, H. Lee, and K. Park, "Robust pedestrian detection by combining visible and thermal infrared cameras," *Sensors*, vol. 15, no. 5, pp. 10580–10615, 2015.

[25] E. Jeon, J.-S. Choi, J. Lee, K. Shin, Y. Kim, T. Le, and K. Park, "Human detection based on the generation of a background image by using a far-infrared light camera," *Sensors*, vol. 15, no. 3, pp. 6763–6788, 2015.

[26] E. Jeon, J. Kim, H. Hong, G. Batchuluun, and K. Park, "Human detection based on the generation of a background image and fuzzy system by using a thermal camera," *Sensors*, vol. 16, no. 4, p. 453, 2016.

[27] G. Batchuluun, H. S. Yoon, J. K. Kang, and K. R. Park, "Gait-based human identification by combining shallow convolutional neural network-stacked long short-term memory and deep convolutional neural network," *IEEE Access*, vol. 6, pp. 63164–63186, 2018.

[28] G. Batchuluun, R. A. Naqvi, W. Kim, and K. R. Park, "Body-movement-based human identification using convolutional neural network," *Expert Syst. Appl.*, vol. 101, pp. 56–77, Jul. 2018.

[29] G. Batchuluun, J. H. Kim, H. G. Hong, J. K. Kang, and K. R. Park, "Fuzzy system based human behavior recognition by combining behavior prediction and recognition," *Expert Syst. Appl.*, vol. 81, pp. 108–133, Sep. 2017.

[30] G. Batchuluun, Y. Kim, J. Kim, H. Hong, and K. Park, "Robust behavior recognition in intelligent surveillance environments," *Sensors*, vol. 16, no. 7, p. 1010, 2016.

[31] H. Eum, J. Lee, C. Yoon, and M. Park, "Human action recognition for night vision using temporal templates with infrared thermal camera," in *Proc. 10th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Jeju-do, South Korea, Oct. 2013, pp. 617–621.

[32] R. Gade and T. B. Moeslund, "Thermal cameras and applications: A survey," *Mach. Vis. Appl.*, vol. 25, no. 1, pp. 245–262, Jan. 2014.

[33] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5485–5493.

[34] T. Yu, Z. Guo, X. Jin, S. Wu, Z. Chen, W. Li, Z. Zhang, and S. Liu, "Region normalization for image inpainting," 2019, *arXiv:1911.10375*. [Online]. Available: http://arxiv.org/abs/1911.10375

[35] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake, UT, USA, Jun. 2018, pp. 5505–5514.

[36] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," 2018, *arXiv:1806.03589*. [Online]. Available: http://arxiv.org/abs/1806.03589

[37] C. Xie, S. Liu, C. Li, M.-M. Cheng, W. Zuo, X. Liu, S. Wen, and E. Ding, "Image inpainting with learnable bidirectional attention maps," 2019, *arXiv:1909.00968*. [Online]. Available: http://arxiv.org/abs/1909.00968

[38] H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," 2019, *arXiv:1905.12384*. [Online]. Available: http://arxiv.org/abs/1905.12384

[39] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-net: Image inpainting via deep feature rearrangement," 2018, *arXiv:1801.09392*. [Online]. Available: http://arxiv.org/abs/1801.09392

[40] Y.-G. Shin, M.-C. Sagong, Y.-J. Yeo, S.-W. Kim, and S.-J. Ko, "PEPSI++: Fast and lightweight network for image inpainting," 2019, *arXiv:1905.09010*. [Online]. Available: http://arxiv.org/abs/1905.09010

[41] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative image inpainting with adversarial edge learning," 2019, *arXiv:1901.00212*. [Online]. Available: http://arxiv.org/abs/1901.00212

[42] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018, *arXiv:1803.07422*. [Online]. Available: http://arxiv.org/abs/1803.07422

[43] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[44] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv:1703.10593*. [Online]. Available: http://arxiv.org/abs/1703.10593

[45] Digital Media Lab. (2019). *CNN Model for Thermal Reflection Removal*. [Online]. Available: http://dm.dgu.edu/link.html.

[46] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2014, *arXiv:1411.4038*. [Online]. Available: http://arxiv.org/abs/1411.4038
[47] FLIR Systems. (2019). *FLIR TauTM 2*. [Online]. Available: https://www.flir.com/products/tau-2/
[48] Mathworks. (2020). *Colormap*. [Online]. Available: https://www.mathworks.com/help/matlab/ref/colormap.html
[49] Github. (2020). *Keras-Surgeon*. [Online]. Available: https://github.com/BenWhetton/keras-surgeon
[50] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, NY, USA: Academic, 2013.
[51] NVIDIA Corporation. (2019). *NVIDIA Titan X*. [Online]. Available: https://www.nvidia.com/en-us/geforce/products/10series/titan-x-pascal/
[52] Keras. (2019). *Keras: The Python Deep Learning Library*. [Online]. Available: https://keras.io/
[53] OpenCV. (2019). *OpenCV: Open Source Computer Vision*. [Online]. Available: http://opencv.org/
[54] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York, NY, USA: Springer-Verlag, 1998.
[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[57] T. Stathaki, *Image Fusion: Algorithms and Applications*. Cambridge, MA, USA: Academic, 2008.
[58] D. Salomon, *Data Compression: The Complete Reference*, 4th ed. New York, NY, USA: Springer-Verlag, 2006.
[59] R. Gade and T. B. Moeslund, "Constrained multi-target tracking for team sports activities," *IPSJ Trans. Comput. Vis. Appl.*, vol. 10, no. 1, pp. 1–11, Dec. 2018, doi: 10.1186/s41074-017-0038-z.
[60] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION)*, Breckenridge, CO, USA, Jan. 2005, pp. 1–6.
[61] B. Abidi, "DOE University Research Program in Robotics Under Grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program Under Grant R01-1344-18; FAA/NSSA Grant R01-1344-48/49; Office of Naval Research Under Grant #N000143010022," IEEE OTCBVS WS Series Bench. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/Data/02/download.html
[62] R. Miezianko, "TerravIC research infrared database—Terravic facial infrared database," IEEE OTCBVS WS Series Bench. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/Data/04/download.html
[63] R. Miezianko, "TerravIC research infrared database—Terravic motion infrared database," IEEE OTCBVS WS Series Bench. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/Data/05/download.html
[64] R. Miezianko, "Terravic research infrared database—Terravic weapon infrared database," IEEE OTCBVS WS Series Bench. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/Data/06/download.html
[65] A. Akula, N. Khanna, R. Ghosh, S. Kumar, A. Das, and H. K. Sardana, "Adaptive contour-based statistical background subtraction method for moving target detection in infrared video sequences," *Infr. Phys. Technol.*, vol. 63, pp. 103–109, Mar. 2014.
[66] G.-A. Bilodeau, A. Torabi, P.-L. St-Charles, and D. Riahi, "Thermal–visible registration of human silhouettes: A similarity measure performance evaluation," *Infr. Phys. Technol.*, vol. 64, pp. 79–86, May 2014.
[67] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Columbus, OH, USA, Jun. 2014, pp. 1–8.
[68] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, and C. Kanan, "VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Boston, MA, USA, Jun. 2015, pp. 10–16.
[69] S. Brahmbhatt, C. Ham, C. C. Kemp, and J. Hays, "ContactDB: Analyzing and predicting grasp contact via thermal imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 8709–8719.

GANBAYAR BATCHULUUN received the B.S. degree in electronic engineering from Huree University, Ulaanbaatar, Mongolia, in 2011, the M.S. degree in electronic engineering from Paichai University, Daejeon, South Korea, in 2014, and the Ph.D. degree in electronics and electrical engineering from Dongguk University, Seoul, South Korea, in 2019. He has been a Professor with the Division of Electronics and Electrical Engineering, Dongguk University, since March 2019. His research interests include biometrics and pattern recognition. He designed the entire system and wrote the original draft of this article.

NA RAE BAEK received the B.S. degree in electronics and electrical engineering from Dongguk University, Seoul, South Korea, in 2017, where she is currently pursuing the combined course of M.S. and Ph.D. degrees in electronics and electrical engineering. Her research interests include biometrics and pattern recognition. She helped to implement pruned fully convolutional network.

DAT TIEN NGUYEN received the B.S. degree in electronics and telecommunications from HUST, Hanoi, Vietnam, in 2009, and the Ph.D. degree in electronics and electrical engineering from Dongguk University, in 2015. He has been a Professor with the Division of Electronics and Electrical Engineering, Dongguk University, since March 2015. His research interests include image processing, biometrics, and deep learning. He supervised this research and revised the original article.

TUYEN DANH PHAM received the B.S. degree in electronics and telecommunications from HUST, Hanoi, Vietnam, in 2010, and the M.S. and Ph.D. degrees in electronics and electrical engineering from Dongguk University, in 2013 and 2017, respectively. He has been a Professor with the Division of Electronics and Electrical Engineering, Dongguk University, since March 2017. His research interests include image processing, biometrics, and deep learning. He helped experiments and analysis.

KANG RYOUNG PARK (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, South Korea, in 1994 and 1996, respectively, and the Ph.D. degree in electrical and computer engineering from Yonsei University, in 2000. He has been a Professor with the Division of Electronics and Electrical Engineering, Dongguk University, since March 2013. His research interests include image processing and biometrics. He helped experiments and analysis.