IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received November 26, 2020, accepted December 13, 2020, date of publication December 31, 2020, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048437

Deep Learning-Based Thermal Image Reconstruction and Object Detection

GANBAYAR BATCHULUUN, JIN KYU KANG, DAT TIEN NGUYEN, TUYEN DANH PHAM, MUHAMMAD ARSALAN, 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, in part by the MSIT, South Korea, through the Information Technology Research Center (ITRC) support program, supervised by the Institute for Information and Communications Technology Promotion (IITP), under Grant IITP-2020-2020-0-01789, and in part by the NRF funded by the MSIT through the Basic Science Research Program under Grant NRF-2020R1A2C1006179.

ABSTRACT Recently, thermal cameras are being widely used in various fields, such as intelligent surveillance, biometrics, and health monitoring. However, the high cost of the thermal cameras poses a challenge in terms of purchase. Additionally, thermal images have an issue pertaining to blurring caused by object movement, camera movement, and camera focus settings. There have been very few studies on image restoration centered around thermal images to address such problems. Moreover, it is important to increase the processing speed of image restoration methods to jointly conduct with methods such as action recognition and object tracking that use temporal information from thermal videos. However, no study has been conducted on simultaneously performing super-resolution reconstruction and deblurring using thermal images. Furthermore, existing studies on object detection using thermal images have errors owing to the incapability in distinguishing reflections on the surrounding ground or wall due to the heat radiated from the object. To address such issues, this study proposes a deep learning-based thermal image restoration method that simultaneously performs super-resolution reconstruction and deblurring. According to recent development of deep learning, generative adversarial network (GAN)-based methods which have ability to preserve texture details in images, and yield sharper and more plausible textures than classical feed forward encoders show success in image-to-image translation tasks. Considering the advantages of GAN, we propose a deblur-SRRGAN for thermal image reconstruction. In addition, we propose a light-weighted Mask R-CNN for object detection in the reconstructed thermal image. For the input, we employ an image processing method that converts 1-channel thermal images (often used in the existing studies) into 3-channel images. The results of the experiments conducted using self-collected databases and an open database demonstrate that our method outperforms the state-of-the-art methods.

INDEX TERMS Thermal image, deep learning, super-resolution reconstruction, image deblurring, object and thermal reflection detection.

I. INTRODUCTION

In recent years, there has been an increase in the use of thermal cameras in various fields. Thermal cameras are being widely implemented in image processing tasks pertaining to the analysis of coronavirus 2019 (COVID-19). A thermal camera can examine the temperature of a body of an object in the form of an image. In other words, a thermal camera, also known as a long-wavelength infrared (LWIR) camera, can measure electromagnetic radiation (EMR) with a wavelength of 8−12 μm [1]. However, thermal images obtained from the camera are often blurry owing to various environmental factors. For example, image blurring occurs frequently when a hot object moves and the temperature of air or surroundings such as walls, floors, and windows is increased because of the object [2]. Furthermore, factors such as the heat induced from the surroundings of the target object, steam generated from the surroundings, and heat reflected from the surrounding walls, floors, or windows may cause image blurring. Moreover, there are two more factors that cause the thermal camera

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

to capture blurry images. 1) the hot air due to sun makes a thermal camera capture blurry images in the middle of the day in summer. 2) the temperature of a thermal camera is increased continuously when capturing images. For example, in case of collecting a huge amount of database, the camera's temperature is increased, and a body of the camera becomes hot. This also affects the camera to capture blurry images. Additionally, the high cost of high-resolution thermal cameras [3] is a challenge. To address the aforementioned problems, the existing studies have proposed various methods of thermal image deblurring and super-resolution reconstruction (SRR). SRR can convert low-resolution (LR) images into high-resolution (HR) images. However, in the existing studies, image deblurring and SRR processes for thermal images were performed separately [4], and no study has been conducted yet to perform image deblurring and SRR tasks simultaneously. In addition, the existing methods were developed using 1-channel (grayscale) thermal images in most cases. Unlike the previous methods, we generated a 3-channel thermal image from an original 1-channel thermal image to obtain more information from a thermal image to increase the performance of the image restoration and the object detection methods. The difference between a 1-channel thermal image and a 3-channel thermal image is that a 1-channel thermal image uses only one channel to describe the intensity of thermal radiation emitted from an object whereas a 3-channel thermal image uses three channels (R, G, and B) to describe the intensity of thermal radiation by colors. To extract more efficient spatial information from the generated 3-channel thermal images, we adjusted the proposed models by changing layers and parameters based on our experiments. So, the final optimal models have been proposed in this study. In addition, we conducted various experiments using original 1-channel thermal images and various 3-channel thermal images such as HSL, HSV, Lab, Luv, XYZ, YCrCb, and RGB. So, we selected thermal images in RGB color space for the proposed methods based on the experimental results. The methods compared in this study are not designed to use 3-channel thermal images for image restoration and object detection tasks. Therefore, our methods are superior to the previous methods in the experimental results in this study.

Thermal images comprise thermal reflections that are caused by the heat radiated from an object in the image on the surrounding floor or wall [5]. As shown in Figure 1 (red dashes), thermal reflections show similar characteristics as the objects in terms of brightness, shape, and pattern. In most cases, thermal reflections are connected to actual objects and are difficult to be distinguished from their objects. To detect exact region of an object from a thermal image, it is important to detect the thermal reflection as well. Based on our experiments, we confirmed that it makes us able to separate the region of the thermal reflection from the region of its object by providing information of thermal reflections to object detection model in training phase. Therefore, we detect thermal reflections and objects from thermal images in this study.

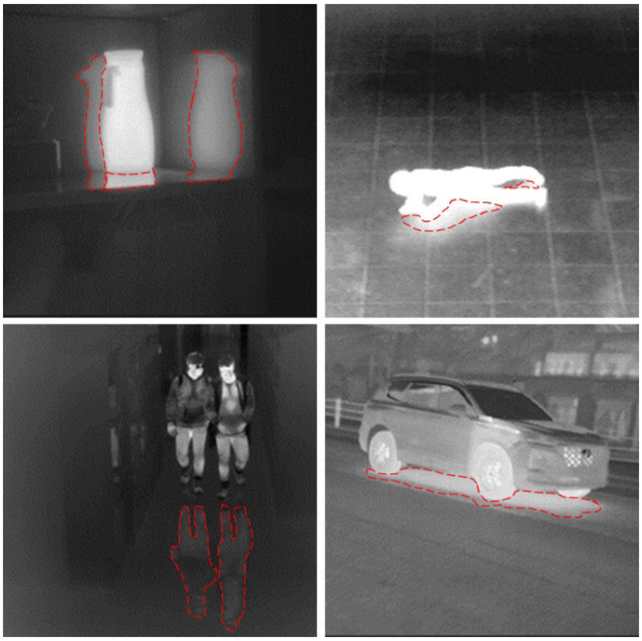


FIGURE 1. Examples of 1-channel thermal images with thermal reflections (red dashes) and objects.

Moreover, there have been previous studies such as action recognition and object tracking conducted by using thermal images. To increase the performance of such methods, the methods such as super-resolution reconstruction, image deblurring and object detection have been proposed in previous studies. However, action recognition and object tracking methods use temporal information from thermal videos which decreases the processing speed of the methods. Furthermore, the processing speed of the methods is decreased more if methods such as super-resolution reconstruction, image deblurring and object detection are conducted jointly with them. Therefore, we performed super-resolution reconstruction and image deblurring using a single generative adversarial network (GAN) to increase the processing speed of the methods.

This study is novel in the following four ways compared to the previous studies:

- For deblurring and super-resolution reconstruction, we propose a new model of deblur-SRRGAN. In addition, we propose a light-weighted Mask R-CNN whose number of layers is reduced compared to original Mask R-CNN for object detection.
- The existing studies have used original 1-channel thermal images to conduct deblurring, SRR, and object detection methods. In this study we generated and used 3-channel thermal images to extract more information to increase the performance of the proposed methods.
- Although various studies have attempted to detect an object or its thermal reflection from thermal images, studies on detecting both simultaneously have not been explored yet. This study proposes a method that simultaneously detects an object and its thermal reflection.

- The models developed in this study and self-collected databases have been published online [6, 51] for fair performance evaluation done by other researchers.

The rest of the paper is structured as follows: Section II introduces existing studies on image reconstruction and object detection. Section III discusses the details of the proposed method. Section IV presents the experimental results and comparative analyses. Section V concludes the paper.

II. RELATED WORKS

A. DEBLUR-SRR METHODS

The deblur-SRR methods developed to simultaneously implement deblurring and SRR can be categorized into the methods using handcrafted features and those using deep features.

With regard to the existing studies on deblur-SRR method using handcrafted features, Farsiu *et al.* [7] proposed a ibilateral method based on norm minimization and robust regularization. Moreover, the authors demonstrated that their proposed deblur-SRR method maintained a fast processing speed while providing robustness against images with motion, blur, and sharp edges. Matsushita *et al.* [8] proposed a method that simultaneously performs SRR and deblurring. In their proposed model, the maximum a posteriori (MAP) estimation was employed to perform deblur-SRR based on video sequencing. Linyang *et al.* [9] proposed a deblur-SRR method that addresses motion blurring using an edge-preserving gradient prior and a sparse kernel prior. Park and Lee [10] used a pioneering uni?ed framework to address four issues simultaneously dense depth reconstruction, camera pose estimation, SRR, and deblurring. Bascle *et al.* [11] proposed a focus deblurring and SRR method to enhance the motion deblurring.

However, as the previous studies still had issues of having performance limitations depending on the environmental changes of the input image and requiring accurate optimal mapping functions, a number of studies using deep features were conducted to resolve the issues as follows: Bianli *et al.* [12] proposed a deblur-SRR method based on convolutional neural networks (CNNs). Zhang *et al.* [13] proposed a deep encoder-decoder network for joint deblurring and super-resolution (ED-DSRN). Zhang *et al.* [14] proposed a gated fusion network (GFN) for the joint image deblur-SRR method. A deblurring super-resolution convolutional neural network (DBSRCNN) has been proposed for joint deblur-SRR method as well [15]. Moreover, Yun and Park [16] proposed a joint face image SRR and deblurring method based on GANs as well. Apart from the aforementioned studies, a number of studies, such as the survey papers on deblurring [17]–[21] and SRR [22]–[24] and the review papers on deblurring [25]–[27] and SRR [28], [29], have been published to summarize and compare the existing studies.

Although there have been numerous studies on deblur-SRR methods, most of them were conducted using images captured by a visible light camera. Furthermore, studies on joint deblur-SRR method that simultaneously performs image deblurring and SRR on thermal images are not conducted yet.

B. OBJECT DETECTION METHODS

In the existing studies on object detection using thermal images [30]–[37], the detection methods can be categorized into methods that detect and do not detect thermal reflections.

Davis and Sharma [30], [31] developed a method to detect humans in an image with the halo effect based on a contour-based method. The authors conducted a separate background subtraction method using visible light and thermal images, extracted contours from the derived images, and then generated silhouette images through the fusion of extracted contour features [30]. Moreover, the authors generated a contour saliency map using the background subtraction method [31]. Wong *et al.* [32] developed a human detection model by extracting a binary image based on human temperature values. Kumar *et al.* conducted a study where human detection was performed using the background subtraction method [33]. Furthermore, Gangodkar *et al.* [34] proposed a human detection model based on the block-matching algorithm.

Lee *et al.* [35] implemented a human detection model using the background subtraction method and the fusion of visible light and thermal images Jeon *et al.* [36] performed human detection for thermal images using background subtraction and background image generation. In addition, Kumar *et al.* [37] summarized and described various existing studies on human detection.

There are methods developed for thermal reflection detection as well. A method was previously proposed to detect thermal reflection based on the Mask region-based convolutional neural network (R-CNN) [5]. Furthermore, regions of thermal reflection were detected using the Mask R-CNN, and the thermal reflections were removed based on the detected regions [38].

However, no study has been conducted to simultaneously detect an object and its thermal reflection from a given thermal image. Thus, this study proposes a detection method that simultaneously detects object and its thermal reflection from 3-channel thermal images. Table 1 lists comparative summaries of the proposed and existing methods.

III. PROPOSED METHOD

A. OVERALL PROCEDURE OF PROPOSED METHODS

This section details the proposed methods. This study proposes a GAN-based SRR and deblurring method that uses 3-channel thermal images as input for image restoration tasks. Furthermore, an object and thermal reflection detection method that uses 3-channel thermal images as input is developed based on the Mask R-CNN method. Figure 2 depicts the overall procedures of the proposed methods and the details of the image restoration method and the object and thermal reflection detection method proposed in Sections III. *B* and IV. *C*. This study uses a thermal

TABLE 1. Comparative summaries of the proposed methods and previous methods.

Category		Method	Advantage	Disadvantage
Deblur-SRR	Using handcrafted features	Based on visible light images [7–11]	Does not require large data and more time for processing	- Performance limitations depend on the environment - Requires accurate derivation of optimal mapping functions
	Using deep features	Based on visible light images [12–16]	Robust performance is maintained considering various environmental changes	Requires large data and more time for training
		Based on 3-channel thermal images (**proposed method**)	- Robust performance is maintained considering various environmental changes - Performance is improved using detailed information of 3-channel images	Requires large data and more time for training
Object detection	Detecting either objects or thermal reflections	Object detection [30–37] and thermal reflection detection [5, 38] methods based on 1-channel thermal images	Performance is not affected by shadows, illumination variations, and human clothing of various colors	- Performance gets degraded owing to the low-level information of 1-channel images and the adverse effect of thermal reflections - Requires large data and more time for training
	Detecting both objects and thermal reflections	Based on 3-channel thermal images (**proposed method**)	- Performance is improved using detailed information of 3-channel images - Performance is not affected by shadows, illumination variations, and human clothing of various colors	Requires large data and more time for training

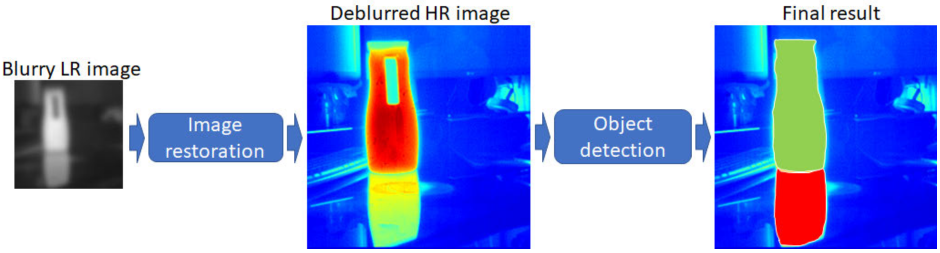


FIGURE 2. Example of an overall procedure of the proposed method. (LR and HR mean low resolution and high resolution, respectively).

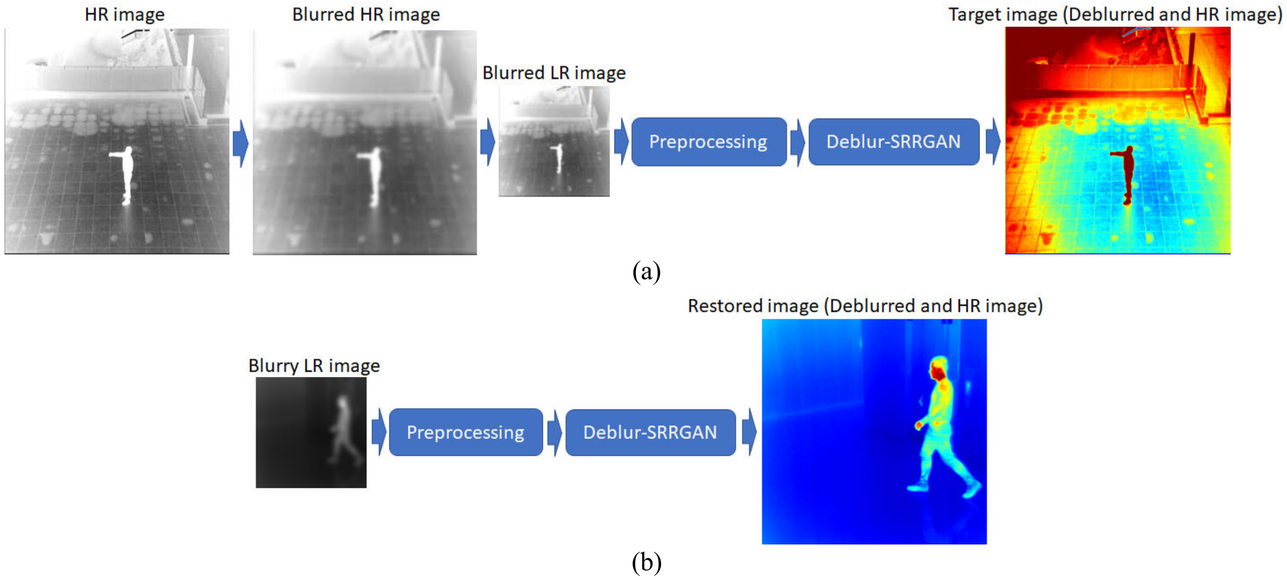
camera that can capture images at 30 frames per second (fps) [39]–[41]. In addition, the camera can measure temperatures at −40 °C to +80 °C to make the object visible in dark and light environments. In this study, various experiments were conducted using the database acquired using the thermal camera (each captured image had a depth of 14 bits and a size of 640 × 480 pixels [4], [5]) and open databases. The details of the databases and thermal camera settings used in this study are presented in the previous studies [42], [43].

B. IMAGE RESTORATION

This section details the image restoration method. Figure 3 depicts the procedures of the proposed image restoration method, and Figure 3(a) displays the training phase of the method. In the training phase, the HR image is first blurred using the Gaussian blur kernel, and then, the image is resized into an LR image.

In the preprocessing stage, the LR images are converted into 3-channel thermal images based on colormaps prior to using the images as an input in the deblur-SRRGAN structure. The jet colormap array is used to perform the image color conversion [44]. A jet colormap array enables the expression of heat in the most appropriate color in the image compared with other colormaps, and maps 256 pixels (0–255) to convert a 1-channel image into a 3-channel image. Figure 4 shows the color conversion operation.

Tables 2–5 and Figure 5 show the structure of the proposed deblur-SRRGAN. The input image and output image sizes are not defined because an input image of any size can be used in the generator. The stride and padding of the filters used in Table 2 are (1 × 1) and (1 × 1), respectively. The filter sizes of both conv2d_1 and conv2d_4 are (9 × 9), and the filter sizes of other layers are (3 × 3). We used filters with different size in Table 2. The filters with size of (9 × 9) and (3 × 3) are referred from the previous study [4]. However,



(a)

(b)

FIGURE 3. Example of procedures of the proposed deblur-SRRGAN method. (a) Training phase of deblur-SRRGAN method; (b) testing phase of deblur-SRRGAN method.

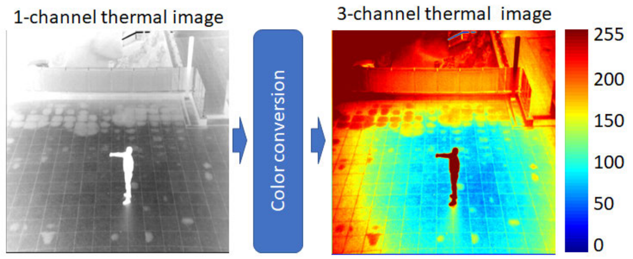


FIGURE 4. Example of the color conversion operation in the preprocessing stage.

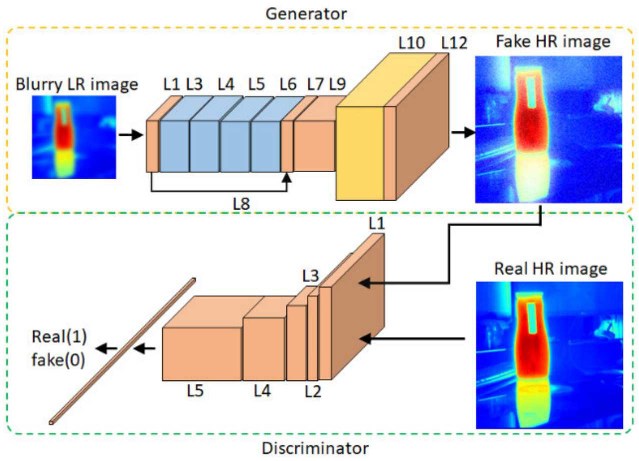


FIGURE 5. Architecture of the proposed deblur-SRRGAN method.

TABLE 2. Description of a structure of generator of the proposed deblur-SRRGAN network (upscaling factor = 2).

Layer number	Layer type	Number of filters	Number of parameters	Layer connection (connected to)
0	input layer	0	0	n/a
1	conv2d_1	64	15,616	input layer
2	prelu_1		0	conv2d_1
3	res_block_1	64	73,856	prelu_1
4	res_block_2	64	73,856	res_block_1
5	res_block_3	64	73,856	res_block_2
6	res_block_4	64	73,856	res_block_3
7	conv2d_2	64	36,928	res_block_4
8	add		0	conv2d_2 & prelu_1
9	conv2d_3	256	147,712	add
10	up2d		0	conv2d_3
11	prelu_2		0	up2d
12	conv2d_4	3	62,211	prelu_2

we do not reuse the network architecture and design from previous work [4], but propose a new deblur-SRRGAN in our research. Furthermore, up2d represents the up-sampling operation. As listed in Table 2, the generator has a structure that upscales the input image by two times (upscaling factor of 2), and the number of layers increases if an upscaling factor of 3 or 4 is used.

In such case, the number of parameters also changes. As listed in Tables 3–5, the filter size, stride, and padding are (3 × 3), (1 × 1), and (1 × 1), respectively. Prelu, res_block, conv2d, add, conv_block, dense, and sigmoid indicate parametric rectified linear unit, residual block, 2-dimensional convolution layer, addition operation, convolution block, fully connected layer, and sigmoid activation function, respectively. In Table 4, the stride of conv_block_1 and conv_block_3 is (1 × 1), and the padding of conv_block_1 is (1 × 1). The stride and padding of other convolution blocks are (2 × 2) and (0 × 0), respectively. The filter size of all

TABLE 3. Description of a structure of residual block.

Layer number	Layer type	Layer connection (connected to)
1	conv2d_1	input
2	prelu	conv2d_1
3	conv2d_2	prelu
4	add	conv2d_2 & input

TABLE 4. Description of a structure of discriminator of the proposed deblur-SRRGAN network.

Layer number	Layer type	Number of filters	Number of parameters	Layer connection (connected to)
0	input layer	0	0	n/a
1	conv_block_1	32	896	input layer
2	conv_block_2	32	9,248	conv_block_1
3	conv_block_3	64	18,496	conv_block_2
4	conv_block_4	128	73,856	conv_block_3
5	conv_block_5	256	295,168	conv_block_4
6	prelu		0	conv_block_5
7	dense		173,057	prelu
8	sigmoid		0	dense

TABLE 5. Description of a convolution block.

Layer number	Layer type	Layer connection (connected to)
1	conv2d_1	input
2	prelu	conv2d_1

the convolution blocks, input image, and output are (3 × 3), (224 × 224 × 3), and (1 × 1), respectively.

The RGB (red, green, and blue) output image obtained using the deblur-SRRGAN is not converted back into a grayscale image but is used as an input in the object and thermal reflection detection method described in the subsequent section.

TABLE 6. Description of a structure of light-weighted Resnet-50.

Occurrence	Layer type	Filter size (stride)	# of filters
1 time	Input layer		
1 time	Conv1	7×7 (2)	64
1 time	Maxpool	3×3 (2)	
2 times	Conv2 /C2	1×1 (4)	64
		3×3 (4)	64
		1×1 (4)	128
2 times	Conv3 /C3	1×1 (8)	64
		3×3 (8)	64
		1×1 (8)	128
2 times	Conv4 /C4	1×1 (16)	64
		3×3 (16)	64
		1×1 (16)	128
1 time	Conv5 /C5	1×1 (32)	64
		3×3 (32)	64
		1×1 (32)	128

TABLE 7. Description of a structure of FPN.

Layer type & layer connection	Filter size (stride)	# of filters
Stage 1		
C5 → Conv1 → P5	1×1 (1)	64
C5 → Conv2 → P6	3×3 (2)	64
P6 → Relu → Conv3 → P7	3×3 (2)	64
(C4 → Conv4) + (P5 → 2×Up) → P4	1×1 (1)	64
(C3 → Conv5) + (P4 → 2×Up) → P3	1×1 (1)	64
(C2 → Conv6) + (P3 → 2×Up) → P2	1×1 (1)	128
Stage 2		
P5 → Conv7 →*P5*	3×3 (1)	64
P6 → Conv8 →*P6*	3×3 (1)	64
P7 → Conv9 →*P7*	3×3 (1)	64
P4 → Conv10 →*P4*	3×3 (1)	64
P3 → Conv11 →*P3*	3×3 (1)	64
P2 → Conv12 →*P2*	3×3 (1)	128

C. OBJECT AND THERMAL REFLECTION DETECTION

This section details the proposed object and thermal reflection detection method. The Mask R-CNN employs in the proposed method used RetinaNet [45], unlike the traditional Mask R-CNN [46] method. The RetinaNet model uses Resnet-50 [47] to extract features. Furthermore, it uses the feature pyramid network (FPN) [48] and small fully convolutional network (FCN) [49] instead of the region proposal network (RPN) [50] when detecting the region of interest (ROI) and candidate object box. In addition, FCN subnets were used to simultaneously perform box classification and box regression, and the final detected box was considered as an input to another FCN to perform mask segmentation. Based on the previous study using the Mask R-CNN, it was inferred that the use of the Resnet-FPN backbone to extract features increases accuracy and processing speed [5]. Thus, we used Resnet-FPN/RetinaNet to detect objects and their thermal reflections in this study. Here, parameters and the number of layers of all the structures of the existing Mask R-CNN [46] model were reduced to increase the processing speed, and we proposed a light-weighted Mask R-CNN. Tables 6 and 7 summarize the structures of Resnet-50 and FPN included in the Mask R-CNN model. The three FCN structures classification subnet, box regression subnet, and mask segmentation network included in the Mask R-CNN are described (Tables 8–10). Tables 6 and 7 list the convolution sets conv2, conv3, conv4, and conv5 as C2, C3, C4, and C5, respectively. As listed in Table 7, arrows, 2×Up, P2 to P7, and *P2* to *P7* denote the next step, upsampling, first feature map of FPN, and the final feature map of FPN, respectively. Here, the first feature map of FPN (extracted in stage 1 of Table 7) is obtained from the feature map extracted using Resnet-50. The final feature map of FPN (extracted in stage 2 of table 7) represents a feature map designed to minimize the aliasing effect caused by the upsampling process. More in-depth descriptions on P1–P7 can be found in a study by

TABLE 8. Description of a structure of classification subnet.

Layer type	Filter size (stride)	# of filters
Conv1	3×3 (1)	64
Relu1		
Conv2	3×3 (1)	64
Relu2		
Conv3	3×3 (1)	64
Relu3		
Conv4	3×3 (1)	128
Relu4		
Conv5	3×3 (1)	A
Sigmoid1		

TABLE 9. Description of a structure of box regression subnet.

Layer type	Filter size (stride)	# of filters
Conv1	3×3 (1)	64
Relu1		
Conv2	3×3 (1)	64
Relu2		
Conv3	3×3 (1)	64
Relu3		
Conv4	3×3 (1)	128
Relu4		
Conv5	3×3 (1)	4×A
Sigmoid1		

TABLE 10. Description of a structure of FCN at the last stage.

Layer type	Filter size (stride)	# of filters
Conv1	3×3 (1)	64
Relu1		
Conv2	3×3 (1)	64
Relu2		
Conv3	3×3 (1)	64
Relu3		
Conv4	3×3 (1)	64
Relu4		
Trans-Conv1	2×2 (1)	64
Relu5		
Conv6	1×1 (1)	1
Relu6		

Lin [45]. In Tables 8 and 9, A denotes anchors, and it was set as A = 9 in the proposed method.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP AND DATABASES

In addition to our self-collected databases (DTh-DB and DI&V-DB [51]), other open databases, such as thermal soccer dataset [52], OSU database collection (OSU thermal pedestrian database [53], OSU color-thermal database [30], terravic motion IR database [54], terravic weapon IR database [55]), LITIV-VAP dataset [56], VIPeR dataset [57], CASIA C dataset [58], BU-TIV dataset [59], and ASL dataset [60], were used in the experiments of this study. Our self-collected databases comprise thermal images of close-range objects, as well as distant objects, captured in dark and bright indoor environments. Figure 7 depicts example images of our thermal image datasets and open thermal image datasets. We used 4,000 images from our self-collected databases. The experiments were evaluated using the two-fold cross-validation procedure. In other words, half of the entire data (2,000 images) were used for training, while the other half (2,000 images) were used for the testing process. Subsequently, the process was repeated by swapping the training and testing data. The average accuracy of the two tests was used as the final value.

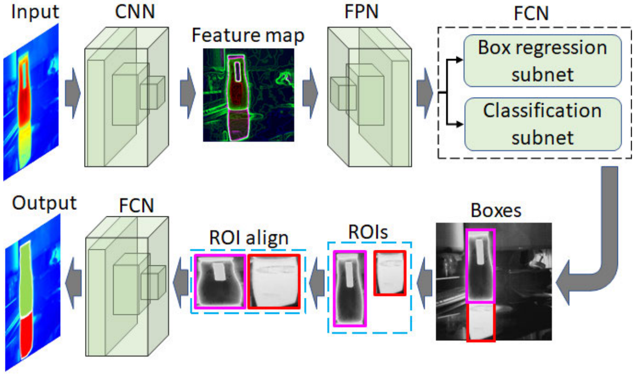


FIGURE 6. Architecture of Mask R-CNN.

The training and testing processes for the proposed algorithms were conducted on a desktop computer, and the specifications of the computer are as follows: NVIDIA GeForce GTX TITAN X graphic card [61], Intel core i7-6700 CPU @ 3.40 GHz (8 CPUs), and 32GB RAM. Furthermore, the algorithms were implemented in Python (version 3.5.4). As for the deep learning library, we used Keras application programming interface (API) (version 2.1.6-tf) with Tensorflow backend engine (version 1.9.0) [62]. The image processing part was implemented using the OpenCV library (version 4.3.0) [63].

B. TRAINING OF THE MODELS

This study compared the proposed deblur-SRRGAN and Mask R-CNN models with other state-of-the-art models, including super-resolution GAN (SRGAN) [64], super-resolution CNN (SRCNN) [65], SegNet [66], and Mask R-CNN [5, 46]. In the training process of the GAN-based models, the batch-size, training epoch, learning rate, and optimizer were configured as 1, 100, 0.0001, and adaptive moment estimation (Adam) [68], respectively. In the training process of the GAN-based models, binary cross-entropy loss was used for both discriminator loss (adversarial loss) and generator loss (reconstruction loss). In the SegNet-based model, the learning rate, learning rate drop period, learning rate drop factor, momentum, mini-batch, and optimizer were configured as 0.001, 20, 0.3, 0.9, 2, and stochastic gradient descent with momentum (SGDM), respectively.

In the Mask R-CNN-based models, the batch-size, training epoch, step-size, and optimizer were configured as 1, 100, 10,000, and Adam, respectively. The number of iteration is

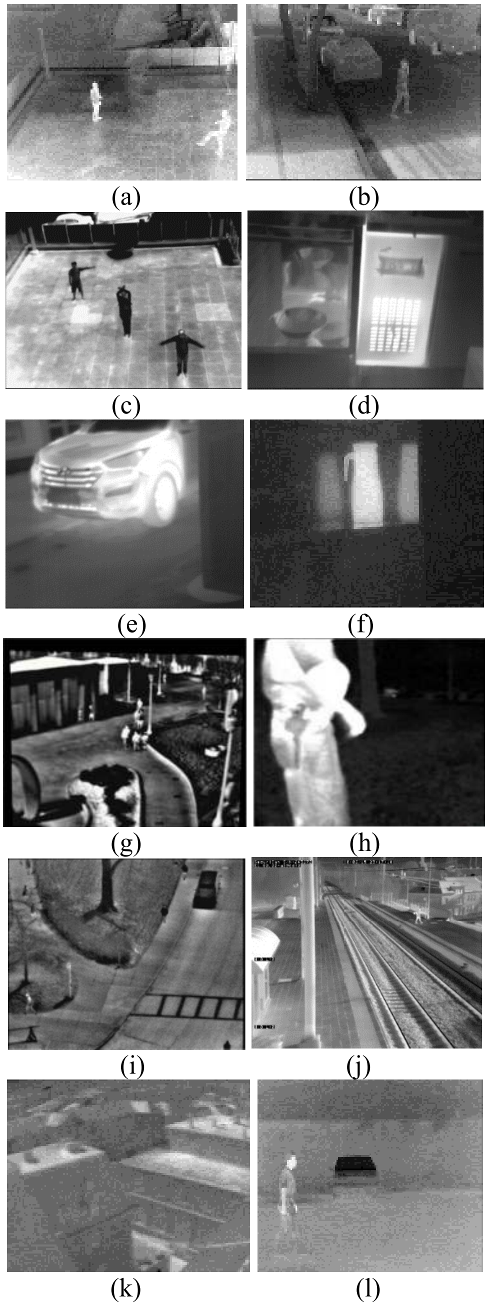


FIGURE 7. Example images from the datasets. (a–f) Images from our datasets; (g, h, i) images from the OSU database collection; (j) image from the VIPeR dataset; (k) image from the ASL dataset; (l) image from the CASIA C dataset.

defined as "the number of training images/the batch size" and it is called as 1 epoch. In our experiments, the number of training images and the batch size are respectively 2,000 and 1, and the consequent number of iteration is 2,000 (2,000/1), which corresponds to 1 epoch. However, if we set the step size of 10,000, 1 epoch is redefined as the case that the iteration of 2,000 is repeated 5 times (10,000/2,000), and the consequent number of iteration for 1 epoch becomes 10,000. Therefore, even with the same number of epochs, the larger step size means the larger number of iterations.

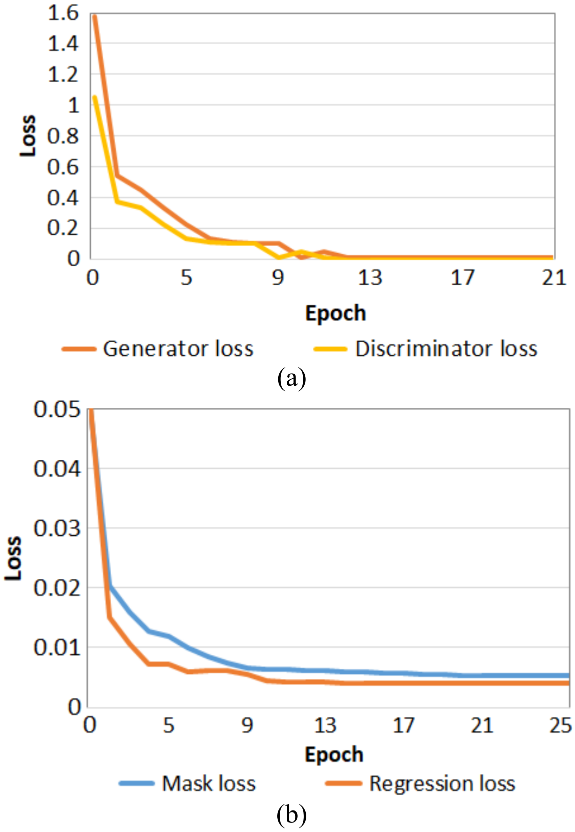


FIGURE 8. Example of training loss curves of (a) deblur-SRRGAN, and (b) proposed Mask R-CNN.

For the Mask R-CNN-based models, Resnet-50 architecture was used as the backbone, and additional fine-tuning was performed on the experimental data of this study by using weights previously trained with the ImageNet database. Finally, in the SRCNN-based model, the batch-size, training iteration, loss, learning rate, and optimizer were configured as 1, 90,000, mean squared error (MSE) [67], 0.0001, and Adam, respectively. Training loss curves are presented in Figure 8.

C. TESTING RESULTS

1) TESTING RESULTS USING THE PROPOSED DEBLUR-SRRGAN NETWORK (ABLATION STUDIES)

In this section, the proposed deblur-SRRGAN network is evaluated. In the evaluation, the MSE, SNR, peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM [69]) were calculated using Equations (1)-(4), respectively, to extract the similarity between the reconstructed image and the original HR image.

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

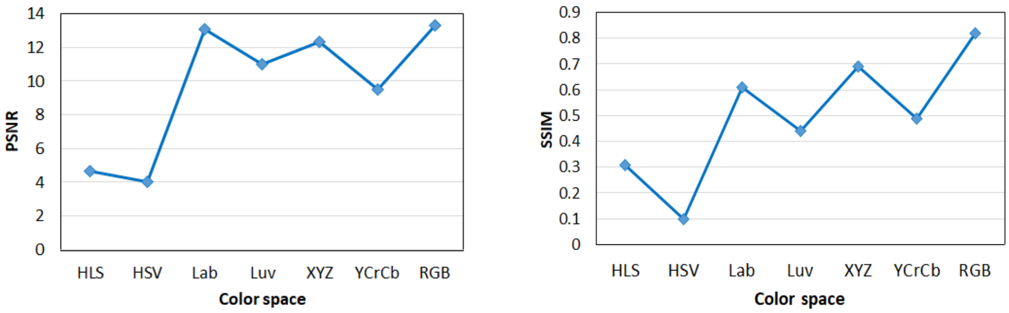
$$\text{SNR} = 10log_{10}\left(\frac{\left(\frac{\sum_{j=1}^{M}\sum_{i=1}^{N}(X(i,j))^2}{MN}\right)}{MSE}\right) \tag{2}$$

TABLE 11. Comparison of accuracies by proposed image restoration method based on image channel (upscaling factor = 2). The images were blurred by various Gaussian blur kernels (kernel size: 5 × 5, 13 × 13, and 19 × 19).

Image channel	MSE			PSNR			SNR			SSIM		
	5 × 5	13 × 13	19 × 19	5 × 5	13 × 13	19 × 19	5 × 5	13 × 13	19 × 19	5 × 5	13 × 13	19 × 19
1 (Without color conversion)	**130**	**556**	**1098**	**28.36**	**21.56**	**18.47**	**19.23**	**12.45**	**9.35**	0.73	0.58	0.52
3 (With color conversion)	715	1708	2602	20.91	16.64	14.64	17.21	10.94	8.52	**0.95**	**0.91**	**0.87**

TABLE 12. Comparison of accuracies by proposed image restoration method based on image channel (upscaling factor = 4). The images were blurred by various Gaussian blur kernels (kernel size: 5 × 5, 13 × 13, and 19 × 19).

Image channel	MSE			PSNR			SNR			SSIM		
	5 × 5	13 × 13	19 × 19	5 × 5	13 × 13	19 × 19	5 × 5	13 × 13	19 × 19	5 × 5	13 × 13	19 × 19
1 (Without color conversion)	**139**	**827**	**1426**	**28.17**	**20.04**	**17.16**	**19.05**	**10.93**	**8.05**	0.70	0.54	0.42
3 (With color conversion)	753	3619	4740	20.73	13.30	11.91	17.03	9.60	2.21	**0.90**	**0.82**	**0.77**



(a) (b)

FIGURE 9. Comparison of results based on color spaces. (a) PSNR and (b) SSIM.

$$\mathrm{PSNR} = 10log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{3}$$

$$\mathrm{SSIM} = \frac{(2\mu_Y\mu_X + C1)\,(2\sigma_{XY} + C2)}{\left({\mu_Y}^2 + \mu_X^2 + C1\right)\left({\sigma_Y}^2 + \sigma_X^2 + C2\right)} \tag{4}$$

In Equations (1)–(3), *X, Y, M,* and *N* represent original image, restored image, image width, and image height, respectively. In Equation (4), μ_X and σ_X represent the mean and standard deviation of the pixel values of a ground-truth image, respectively; μ_Y and σ_Y represent the mean and standard deviation of the pixel values of the restored image, respectively; and σ_{XY} is the covariance of the two images. *C1* and *C2* are positive constants, so that the denominator does not become zero [69].

Tables 11 and 12, which display ablation studies, compare the results derived using 1-channel images (without color conversion of Figure 4) and 3-channel images (with color conversion of Figure 4). The results are compared after providing the images that are blurred with Gaussian blur kernels (5 × 5, 13 × 13, 19 × 19) as the input. Table 11 compares the results derived based on the upscaling factor of 2, and Table 12 compares the results derived based on the upscaling factor of 4. As shown in Tables 11 and 12, the 1-channel thermal image without color conversion was found to yield higher accuracy with regard to MSE, PSNR, and SNR, whereas the 3-channel thermal image with color conversion was found to yield higher accuracy with regard to SSIM. However, the MSE, PSNR, and SNR measures are poor to evaluate the difference and similarity in the human visual image quality [70], [71]. Rather, SSIM can better evaluate the similarity in the image quality [69]. Based on the aforementioned results, the proposed method has obtained the highest accuracy.

TABLE 13. Comparison of accuracies based on various color spaces (upscaling factor = 4). The images were blurred by Gaussian blur kernel (kernel size: 13 × 13).

Color space	MSE	PSNR	SNR	SSIM
HSL	27423	4.66	1.85	0.31
HSV	28911	4.02	0.99	0.10
Lab	6523	13.11	9.52	0.61
Luv	6583	11.01	9.01	0.44
XYZ	4759	12.35	8.89	0.69
YCrCb	8322	9.51	6.42	0.49
RGB	**3619**	**13.30**	**9.60**	**0.82**

In addition, image restoration was performed using various color spaces, and the performance based on the color space was compared (Table 13 and Figure 9). As shown

TABLE 14. Comparison of the proposed deblur-SRRGAN with the state-of-the-art methods using six databases (upscaling factor = 2). The images were blurred using Gaussian blur kernel (kernel size = 13 × 13).

	Methods	Measurements	Database						
			1	2	3	4	5	6	Average
Non-learning-based	Bicubic [72]	PSNR	17.13	18.67	19.05	19.76	18.35	17.11	18.34
		SSIM	0.6653	0.7832	0.793	0.796	0.6884	0.6509	0.7267
	OKI-SR [78]	PSNR	20.12	21.05	22.95	22.61	21.41	20.24	21.39
		SSIM	0.7131	0.8113	0.8191	0.8231	0.7321	0.7008	0.7665
Learning-based	SRCNN [65]	PSNR	28.14	31.68	29.06	34.78	30.37	26.03	30.01
		SSIM	0.8953	0.9032	0.913	0.916	0.9084	0.7609	0.8828
	ScSR [73]	PSNR	28.36	31.3	28.92	35.44	30.87	25.78	30.11
		SSIM	0.9078	0.9081	0.9147	0.9183	0.9121	0.7631	0.8873
	Zhang et al.'s [74]	PSNR	29.63	32.8	29.25	35.58	31.02	26.41	**30.78**
		SSIM	0.9047	0.9085	0.9132	0.9176	0.9109	0.7643	0.8865
	SRGAN [64]	PSNR	26.86	30.73	27.64	31.38	29.59	25.01	28.53
		SSIM	0.8873	0.8983	0.8881	0.8798	0.887	0.6738	0.8523
	Bianli et al. [12]	PSNR	27.16	29.23	28.14	30.11	29.03	26.12	28.298
		SSIM	0.8772	0.8893	0.8183	0.8813	0.8281	0.7721	0.8443
	Zhang et al. [13]	PSNR	26.12	31.62	29.54	28.11	28.19	26.45	28.338
		SSIM	0.8703	0.8999	0.8941	0.8908	0.8982	0.788	0.873
	Zhang et al. [14]	PSNR	27.96	31.03	27.12	31.01	27.49	28.08	28.781
		SSIM	0.8878	0.9013	0.9011	0.8899	0.8997	0.7578	0.8729
	DBSRCNN [15]	PSNR	27.05	28.63	28.34	30.58	30.15	28.15	28.816
		SSIM	0.8793	0.9114	0.8971	0.881	0.907	0.893	0.8948
	Yun and Park [16]	PSNR	28.16	30.08	28.11	30.3	29.44	28.22	29.051
		SSIM	0.8913	0.918	0.911	0.898	0.901	0.903	0.9037
	Proposed method	PSNR	18.61	21.53	20.12	23.81	21.32	18.91	20.72
		SSIM	0.9135	0.921	0.9247	0.9281	0.9256	0.9159	**0.9214**

in Table 13, the SSIMs obtained by using RGB, XYZ, and Lab color spaces were higher than those obtained by using other color spaces. Moreover, the results obtained by using HSL and HSV color spaces were much lower compared to those obtained by using other color spaces. The reason for the difference in results is that HSL and HSV color spaces use only one channel (Hue) to describe color information whereas Lab, XYZ and RGB color spaces respectively use two channels (a and b) and three channels (X, Y, and Z) and (R, G, and B) to describe color information. In Figure 10, images by the four color spaces are compared with an original grayscale image. As shown in Figure 10, RGB image provides more spatial information compared to the images of other color space.

As a result, the performance obtained using RGB thermal images was the highest in all the cases. The proposed methods were designed based on the aforementioned experimental results. The subsequent sections compare the accuracies of the proposed and existing methods through experiments.

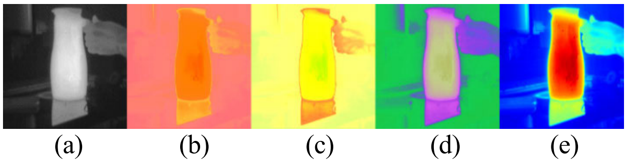


FIGURE 10. Comparison of input images in different color spaces. (a) An original grayscale image; (b) HSL image; (c) HSV image; (d) Lab image; (e) RGB image.

2) COMPARISONS OF THE PROPOSED DEBLUR-SRRGAN NETWORK WITH THE STATE-OF-THE-ART METHODS

In this section, we conducted additional experiments using various open datasets of thermal images described in Section IV.A to evaluate the applicability of the proposed

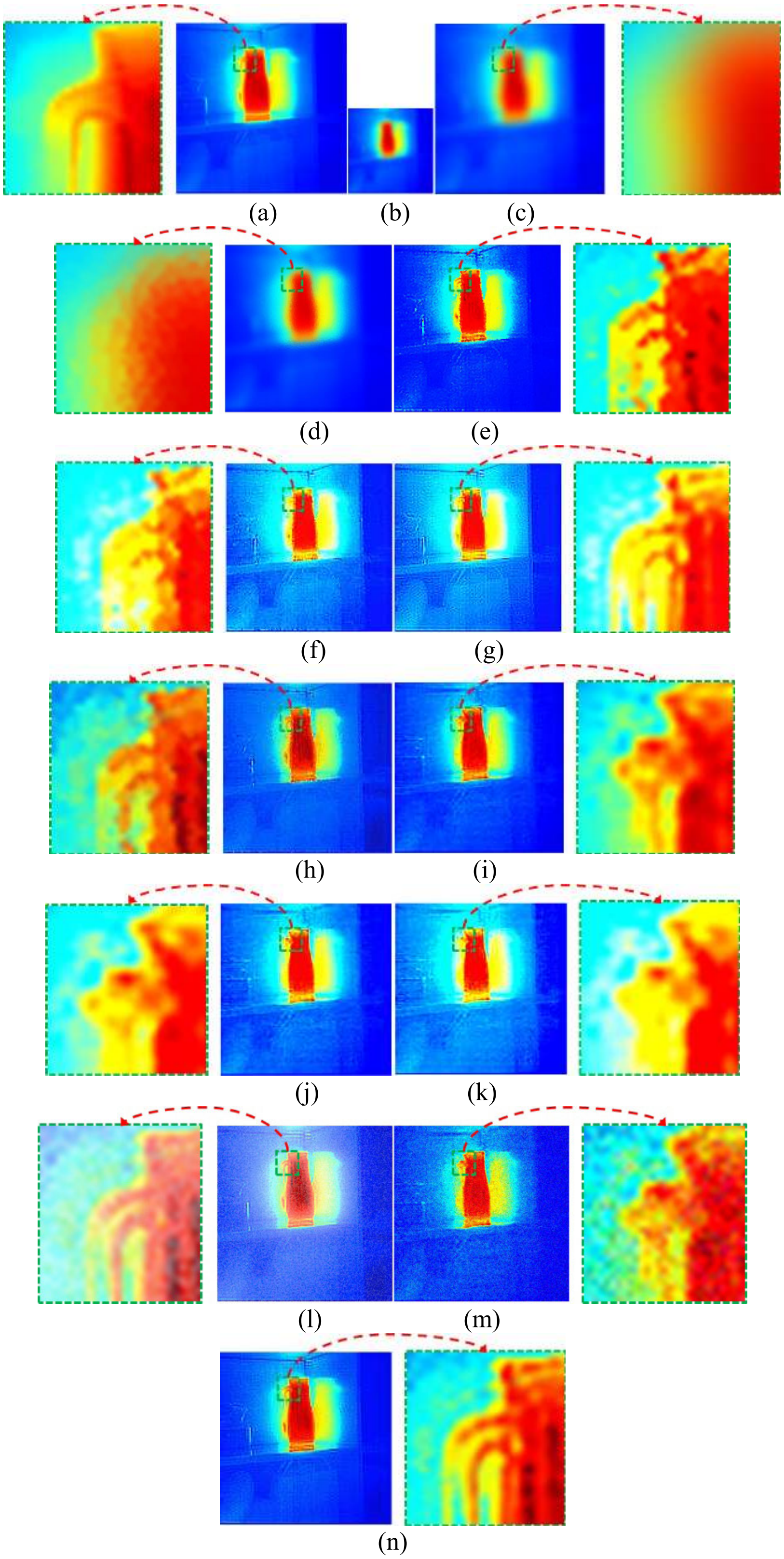


FIGURE 11. Comparison of the results based on the upscaling factor of 2 (Gaussian blur kernel size = 13 × 13). (a) An original image; (b) a downscaled image; (c) bicubic [72]; (d) OKI-SR [78]; (e) SRCNN [65]; (f) ScSR [73]; (g) Zhang *et al.*'s [74]; (h) SRGAN [64]; (i) Bianli *et al.* [12]; (j) Zhang *et al.* [13]; (k) Zhang *et al.* [14]; (l) DBSRCNN [15]; (m) Yun and Park [16]; (n) the proposed deblur-SRRGAN method.

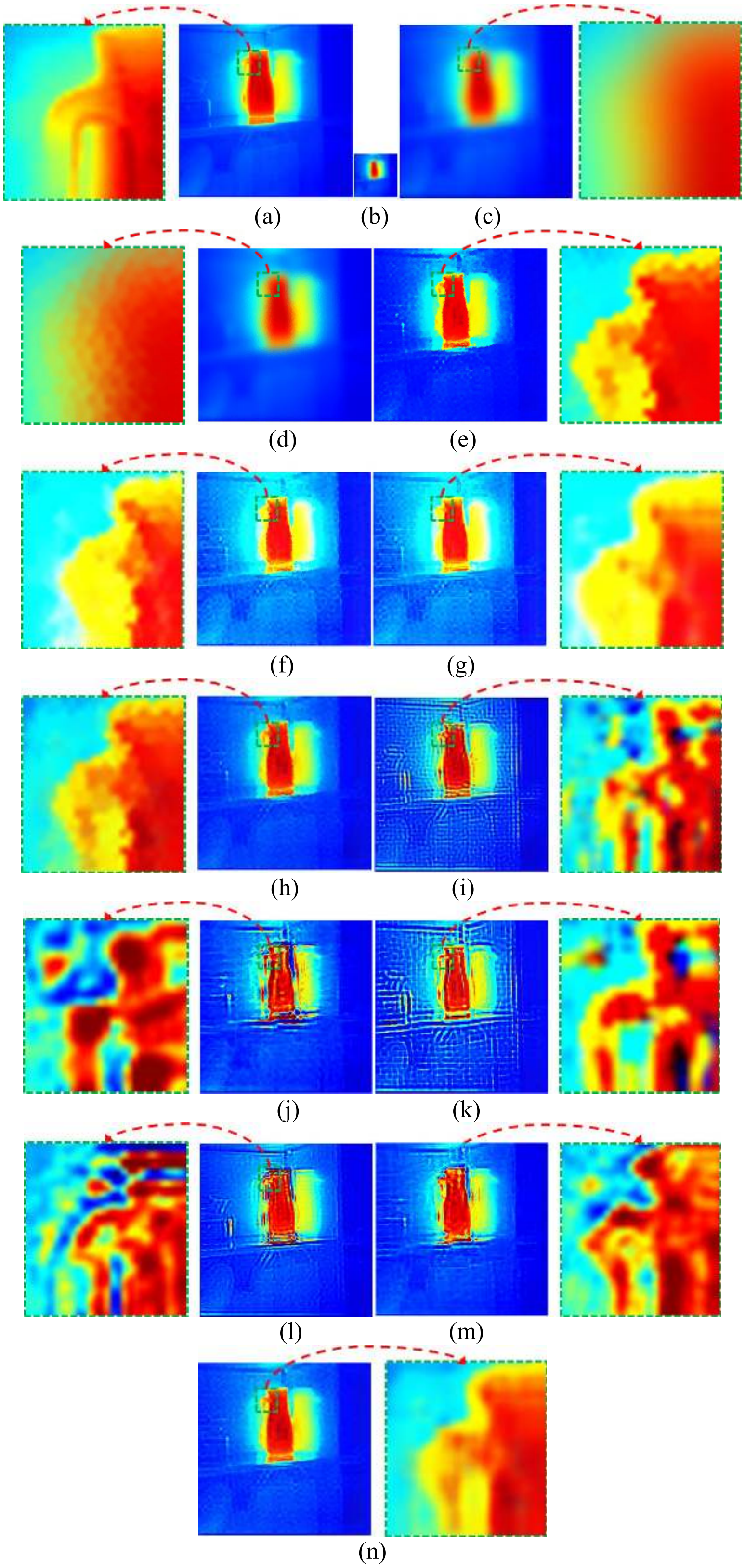


FIGURE 12. Comparison of the results based on the upscaling factor of 4 (Gaussian blur kernel size = 13 × 13). (a) An original image; (b) a downscaled image; (c) bicubic [72]; (d) OKI-SR [78]; (e) SRCNN [65]; (f) ScSR [73]; (g) Zhang *et al.*'s [74]; (h) SRGAN [64]; (i) Bianli *et al.* [12]; (j) Zhang *et al.* [13]; (k) Zhang *et al.* [14]; (l) DBSRCNN [15]; (m) Yun and Park [16]; (n) the proposed deblur-SRRGAN method.

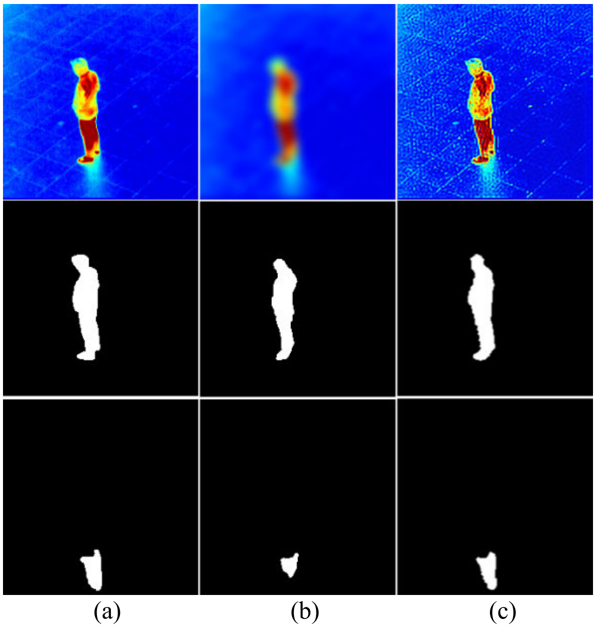


FIGURE 13. Examples of results by proposed object detection method with or without deblur-SRRGAN. The images were blurred by Gaussian blur kernel (kernel size = 13 × 13). (a) Ground truth image, and results (b) without deblur-SRRGAN, and (c) with deblur-SRRGAN. In (a) ~ (c), the upper, middle, and lower images show the input, object area, and reflection region, respectively.

image restoration method based on various environments. Moreover, the proposed method was compared with the state-of-the-art models based on six types of OSU dataset collections [30, 53–55]. For the comparison, Bicubic [72] and ordinary kriging interpolation-based super-resolution method (OKI-SR) [78] were selected as traditional methods, and SRCNN [65], sparse coding super-resolution (ScSR) [73], Zhang *et al.*'s method [74] and SRGAN [64] were selected as the state-of-the-art methods. Bicubic and OKI-SR belong to non-learning-based method whereas the others belong to learning-based method. Non-learning-based approaches usually do not require the training process whereas learning-based one need the training procedure to determine the optimal parameters or methods.

In addition, because the methods [12]–[16] can simultaneously perform super-resolution reconstruction and deblurring, we conducted additional experiments and compared the methods [12]–[16] with our method in Tables 14, 15, and Figures 11 and 12.

Tables 14 and 15, and Figures 11 and 12 compares the results of the models. Table 14 compares the results derived with an upscaling factor of 2, and Table 15 compares the results derived with an upscaling factor of 4. Original image of Figure 11(a) is same to that of Figure 12(a). From these original images, two kinds of downscaled images of Figure 11(b) (1/4 downscaling) and Figure 12(b) (1/16 downscaling) were generated, respectively. With these same original images and downscaled images, our method and other methods were trained for fair comparisons. Figures 11(c) ~ (h) and Figures 12(c) ~ (h) show the reconstructed results by other methods and ours.

According to Tables 14 and 15, and Figures 11 and 12, the SSIM results obtained using the proposed restoration method were considered superior in all the cases compared with that of the state-of-the-art methods. Moreover, the results obtained by non-learning-based methods show more blurred cases compared to those by learning-based methods. This

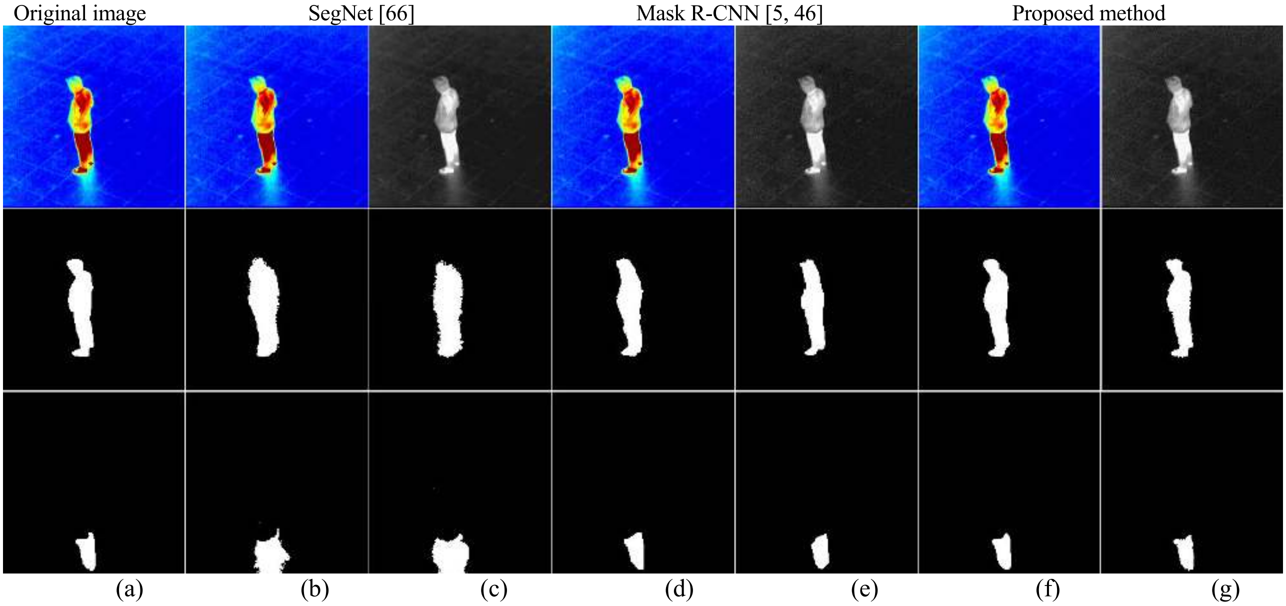


FIGURE 14. Examples of detection results using original images. From the top to the bottom, input images, detected results of an object, and a thermal reflection. (a) Ground truth masks; (b) and (c) result mask images by SegNet [66] with and without color conversion of Figure 4, respectively; (d) and (e) result mask images by Mask R-CNN [5], [46] with and without color conversion of Figure 4, respectively; (f) and (g) result mask images by proposed method with and without color conversion of Figure 4, respectively.

TABLE 15. Comparison of the proposed deblur-SRRGAN with the state-of-the-art methods using six databases (upscaling factor = 4). The images were blurred by Gaussian blur kernel (kernel size = 13 × 13).

	Methods	Measurements	Database						
			1	2	3	4	5	6	Average
Non-learning-based	Bicubic [72]	PSNR	13.74	11.62	14.12	14.38	13.98	12.51	13.39
		SSIM	0.522	0.6762	0.6465	0.7345	0.5483	0.4619	0.5982
	OKI-SR [78]	PSNR	14.98	12.55	15.45	15.78	14.44	13.29	14.41
		SSIM	0.6122	0.6992	0.6512	0.7511	0.5989	0.5201	0.6387
Learning-based	SRCNN [65]	PSNR	11.43	12.9	12.83	15.43	13.23	13.92	13.29
		SSIM	0.7206	0.7681	0.7745	0.8259	0.7944	0.8238	0.7845
	ScSR [73]	PSNR	25	27.49	26.05	31.01	26.81	23.03	26.56
		SSIM	0.8444	0.8359	0.8418	0.8393	0.8395	0.7001	0.8168
	Zhang et al.'s [74]	PSNR	26.12	28.81	26.35	31.13	26.94	23.59	**27.15**
		SSIM	0.8415	0.8362	0.8405	0.8387	0.8384	0.7012	0.816
	SRGAN [64]	PSNR	22.79	25.39	24.21	28.28	24.54	24.61	24.97
		SSIM	0.8206	0.7924	0.7852	0.8003	0.7838	0.542	0.754
	Bianli et al. [12]	PSNR	22.33	23.31	24.01	25.21	23.44	23.25	23.59
		SSIM	0.8166	0.7654	0.7712	0.8102	0.7608	0.5315	0.7426
	Zhang et al. [13]	PSNR	22.59	24.31	24.99	25.13	23.24	23.51	23.96
		SSIM	0.8101	0.7784	0.7651	0.8201	0.7208	0.612	0.751
	Zhang et al. [14]	PSNR	21.32	25.01	22.36	23.21	23.99	23.25	23.19
		SSIM	0.8131	0.7662	0.7455	0.8303	0.7108	0.7213	0.7645
	DBSRCNN [15]	PSNR	21.21	25.11	23.51	27.19	23.32	23.21	23.92
		SSIM	0.8226	0.7811	0.7132	0.7503	0.7576	0.622	0.7411
	Yun and Park [16]	PSNR	21.72	24.23	22.11	26.24	23.61	23.31	23.536
		SSIM	0.8103	0.7721	0.7322	0.7526	0.7628	0.6512	0.7468
	Proposed method	PSNR	13.21	15.52	15.02	20.11	15.45	17.1	16.06
		SSIM	0.8245	0.8525	0.8498	0.8921	0.8558	0.8759	**0.8584**

TABLE 16. Comparison of detection results without color conversion of Figure 4 in original HR images.

Class	SegNet [66]					Mask R-CNN [5, 46]					Proposed method				
	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU
Object	0.96	0.71	0.82	0.95	0.69	0.90	0.86	0.87	0.97	0.80	0.90	0.88	0.89	0.98	0.81
Reflection	0.92	0.46	0.62	0.95	0.44	0.87	0.75	0.80	0.93	0.68	0.89	0.76	0.82	0.98	0.70
Average	**0.94**	0.585	0.72	0.95	0.565	0.885	0.805	0.835	0.95	0.74	0.895	**0.82**	**0.855**	**0.98**	**0.755**

is because the non-learning-based methods use a manually designed mapping function whereas the learning-based methods adopt the optimal mapping function which is obtained based on learning process using a large size of training data.

3) COMPARISON OF THE PROPOSED OBJECT AND THERMAL REFLECTION DETECTION METHOD WITH THE STATE-OF-THE-ART METHODS (ABLATION STUDIES)

In this section, the results obtained through the proposed object and thermal reflection detection method are presented. Furthermore, the proposed detection method was compared with the state-of-the-art models such as SegNet [66], conventional Mask R-CNN [5], [46], Mask-Refined R-CNN (MR R-CNN) [79], global-and-local network architecture (GLNet) [80] and multi-scale global contrast CNN (MGCC) [81], and traditional methods such as random sample consensus-based moving object detection (RSC-MOD) [82] and contour-based background subtraction method (Con-BS) [83]. Con-BS and RSC-MOD belong to non-learning-based method whereas the others belong to learning-based method.

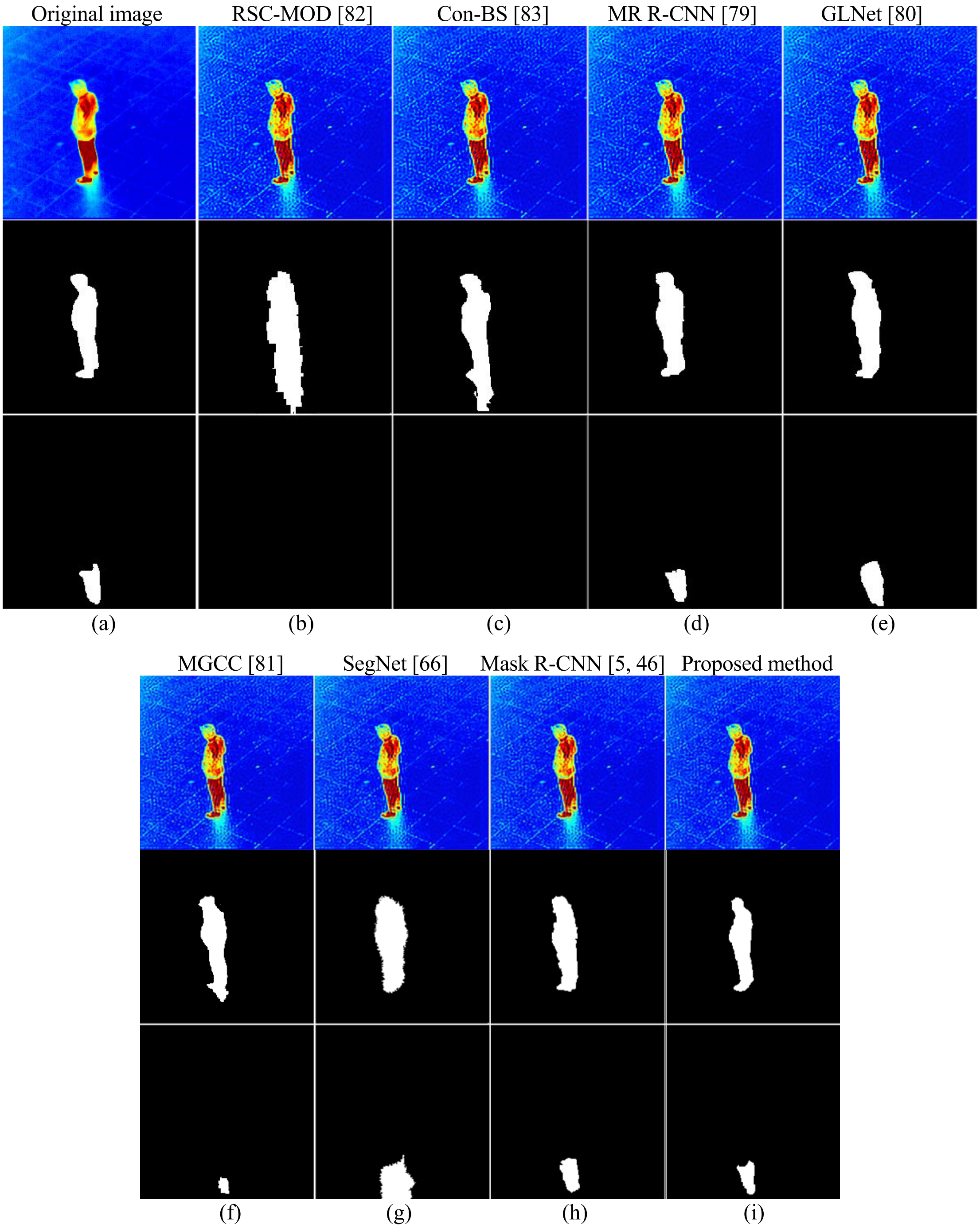


FIGURE 15. Examples of detection results using restored images (upscaling factor = 2 and kernel size = 13 × 13). From the top to the bottom, input images, detected results of an object, and a thermal reflection. (a) ground truth masks; (b) result mask images by RSC-MOD [82]; (c) result mask images by Con-BS [83]; (d) result mask images by MR R-CNN [79]; (e) result mask images by GLNet [80]; (f) result mask images by MGCC [81]; (g) result mask images by SegNet [66]; (h) result mask images by Mask R-CNN [5], [46]; (i) result mask images by proposed method.

Comparisons were made based on the similarity between the detected image and the ground truth mask image when measuring the accuracy of the detection models. In the comparison process, the object and thermal reflection pixels of the ground truth mask image are referred to as positive pixels, while the background pixels of the ground truth mask image are referred to as negative pixels. In addition, the case in which the pixels are detected as positive pixels is called the true positive case (TP), whereas the case in which the positive pixels are incorrectly detected as negative pixels is called the false negative case (FN). The case in which the negative pixels are incorrectly detected as positive pixels is called the

TABLE 17. Comparison of detection results with color conversion of Figure 4 in original HR images.

Class	SegNet [66]					Mask R-CNN [5, 46]					Proposed method				
	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU
Object	0.97	0.75	0.85	0.96	0.74	0.90	0.88	0.89	0.97	0.81	0.90	0.89	0.90	0.98	0.82
Reflection	0.94	0.49	0.64	0.95	0.47	0.87	0.77	0.82	0.95	0.70	0.89	0.79	0.83	0.98	0.72
Average	**0.955**	0.62	0.745	0.955	0.605	0.885	0.825	0.855	0.96	0.755	0.895	**0.84**	**0.865**	**0.98**	**0.77**

TABLE 18. Comparison of detection results without color conversion of Figure 4 and with Deblur-SRRGAN, and the state-of-the-art methods (upscaling factor = 2). The images were blurred by Gaussian blur kernel (kernel size = 13 × 13).

Class	SegNet [66]					Mask R-CNN [5, 46]					Proposed method				
	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU
Object	0.87	0.54	0.67	0.91	0.49	0.82	0.65	0.71	0.93	0.57	0.82	0.67	0.73	0.94	0.58
Reflection	0.65	0.40	0.49	0.94	0.30	0.61	0.66	0.63	0.92	0.46	0.63	0.67	0.65	0.97	0.48
Average	**0.76**	0.47	0.58	0.925	0.395	0.715	0.655	0.67	0.925	0.515	0.725	**0.67**	**0.69**	**0.955**	**0.53**

TABLE 19. Comparison of detection results without color conversion of Figure 4 and with Deblur-SRRGAN, and the state-of-the-art methods (upscaling factor = 4). The images were blurred by Gaussian blur kernel (kernel size = 13 × 13).

Class	SegNet [66]					Mask R-CNN [5, 46]					Proposed method				
	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU	TPR	PPV	F1	ACC	IoU
Object	0.74	0.53	0.62	0.90	0.44	0.70	0.64	0.66	0.92	0.51	0.70	0.66	0.68	0.93	0.52
Reflection	0.55	0.36	0.43	0.93	0.25	0.52	0.60	0.56	0.91	0.38	0.54	0.61	0.58	0.96	0.40
Average	**0.645**	0.445	0.525	0.915	0.345	0.61	0.62	0.61	0.915	0.445	0.62	**0.635**	**0.63**	**0.945**	**0.46**

TABLE 20. Comparison of results obtained by proposed object detection method with or without deblur-SRRGAN. The images were blurred by Gaussian blur kernel (kernel size = 13 × 13).

Method	Measurement	Upscaling factor of 2			Upscaling factor of 4		
		Object	Reflection	Average	Object	Reflection	Average
Without deblur-SRRGAN	TPR	0.69	0.50	0.595	0.62	0.44	0.53
	PPV	0.79	0.73	0.76	0.7	0.69	0.695
	F1	0.74	0.61	0.675	0.66	0.52	0.59
	ACC	0.92	0.91	0.915	0.89	0.87	0.88
	IoU	0.57	0.47	0.52	0.52	0.4	0.46
With deblur-SRRGAN	TPR	0.71	0.51	**0.61**	0.63	0.47	**0.55**
	PPV	0.78	0.77	**0.775**	0.76	0.83	**0.795**
	F1	0.74	0.61	0.675	0.69	0.6	**0.645**
	ACC	0.95	0.97	**0.96**	0.94	0.96	**0.95**
	IoU	0.59	0.49	**0.54**	0.53	0.43	**0.48**

false positive case (FP). Furthermore, #TP, #FP, and #FN are defined as the numbers of TP, FP, and FN, respectively. The accuracies of the object and thermal reflection detection models were derived based on true positive rate (TPR) (#TP/(#TP + #FN)), positive predictive value (PPV) (#TP/(#TP + #FP)), accuracy (ACC) [75], F1 score (F1) [76], and intersection over union (IoU) [77] as follows.

$$\text{Accuracy (ACC)} = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN} \quad (5)$$

$$F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} \quad (6)$$

$$IoU(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \quad (7)$$

In the object detection experiment, six databases were combined into one database to conduct accuracy tests. Tables 16 and 17, which are ablation studies, list the comparison among detection results obtained in the proposed and state-of-the-art methods using 1-channel (without color conversion of Figure 4) and 3-channel images (with color conversion of Figure 4) of the original HR images. Tables 18–21, which are also ablation studies, list the comparison of object and thermal reflection detection results obtained in the proposed and the state-of-the-art methods

TABLE 21. Comparison of detection results with color conversion of Figure 4 and Deblur-SRRGAN, and the state-of-the-art methods. The images were blurred by Gaussian blur kernel (kernel size = 13 × 13).

	Method	Measurement	Upscaling factor of 2			Upscaling factor of 4		
			Object	Reflection	Average	Object	Reflection	Average
Non-learning-based	RSC-MOD [82]	TPR	0.72	0.59	0.655	0.61	0.52	0.565
		PPV	0.6	0.57	0.585	0.53	0.51	0.52
		F1	0.68	0.58	0.63	0.56	0.51	0.535
		ACC	0.78	0.75	0.765	0.62	0.62	0.62
		IoU	0.41	0.4	0.405	0.33	0.3	0.315
	Con-BS [83]	TPR	0.78	0.58	0.68	0.65	0.51	0.58
		PPV	0.6	0.6	0.6	0.55	0.56	0.555
		F1	0.67	0.58	0.625	0.59	0.53	0.56
		ACC	0.88	0.83	0.855	0.79	0.75	0.77
		IoU	0.5	0.4	0.45	0.42	0.36	0.39
Learning-based	MR R-CNN [79]	TPR	0.85	0.63	0.74	0.73	0.55	0.64
		PPV	0.62	0.66	0.64	0.65	0.6	0.625
		F1	0.71	0.64	0.675	0.68	0.58	0.63
		ACC	0.92	0.92	0.92	0.9	0.89	0.895
		IoU	0.57	0.46	0.515	0.5	0.41	0.455
	GLNet [80]	TPR	0.89	0.65	0.77	0.76	0.56	0.66
		PPV	0.58	0.41	0.495	0.57	0.39	0.48
		F1	0.7	0.5	0.6	0.65	0.45	0.55
		ACC	0.91	0.92	0.915	0.91	0.9	0.905
		IoU	0.51	0.31	0.41	0.46	0.25	0.355
	MGCC [81]	TPR	0.80	0.60	0.7	0.68	0.50	0.59
		PPV	0.68	0.65	0.66	0.63	0.58	0.605
		F1	0.73	0.62	0.67	0.66	0.53	0.595
		ACC	0.92	0.87	0.89	0.88	0.81	0.845
		IoU	0.57	0.45	0.51	0.49	0.39	0.44
	SegNet [66]	TPR	0.88	0.66	**0.77**	0.75	0.57	**0.66**
		PPV	0.57	0.43	0.5	0.56	0.39	0.475
		F1	0.69	0.52	0.595	0.64	0.45	0.545
		ACC	0.92	0.94	0.93	0.91	0.93	0.92
		IoU	0.52	0.32	0.42	0.47	0.26	0.365
	Mask R-CNN [5, 46]	TPR	0.82	0.61	0.715	0.7	0.52	0.61
		PPV	0.67	0.67	0.67	0.66	0.61	0.635
		F1	0.73	0.65	**0.69**	0.68	0.58	0.63
		ACC	0.93	0.94	0.935	0.92	0.93	0.925
		IoU	0.58	0.48	0.53	0.52	0.4	0.46
	Proposed method	TPR	0.71	0.51	0.61	0.63	0.47	0.55
		PPV	0.78	0.77	**0.775**	0.76	0.83	**0.795**
		F1	0.74	0.61	0.675	0.69	0.6	**0.645**
		ACC	0.95	0.97	**0.96**	0.94	0.96	**0.95**
		IoU	0.59	0.49	**0.54**	0.53	0.43	**0.48**

using images (without or with color conversion of Figure 4) restored through deblur-SRRGAN.

To confirm that the proposed deblur-SRRGAN is helpful to improve object detection performance, we performed the additional experiments. As shown in Table 20 and Figure 13, proposed object detection method with deblur-SRRGAN shows the higher accuracies than that without deblur-SRRGAN, which confirms that the proposed

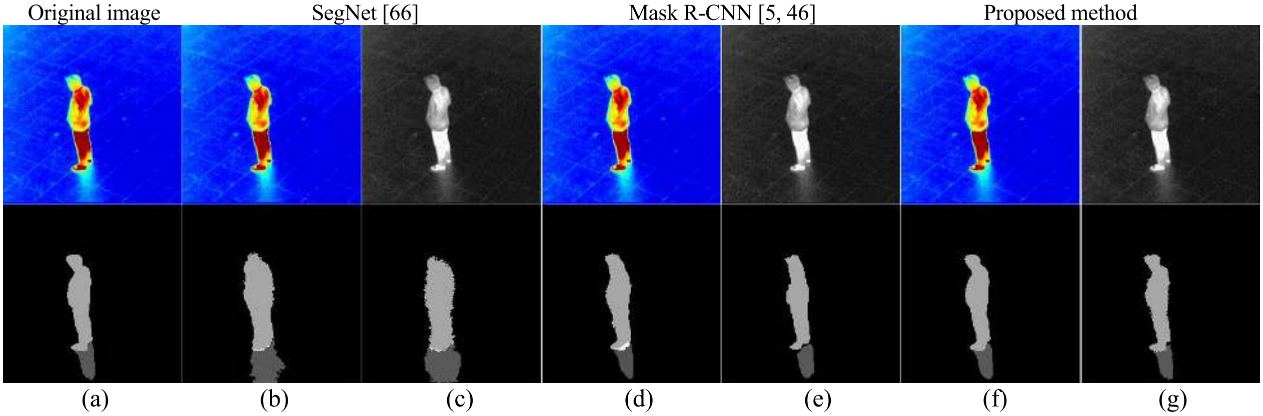


FIGURE 16. Example of pixel classification result using original images. From the top to the bottom, input images and detected results of an object and a thermal reflection. (a) Ground truth masks; (b) and (c) result mask images by SegNet [66] with and without color conversion of Figure 4, respectively; (d) and (e) result mask images by Mask R-CNN [5], [46] with and without color conversion of Figure 4, respectively; (f) and (g) result mask images by proposed method with and without color conversion of Figure 4, respectively.

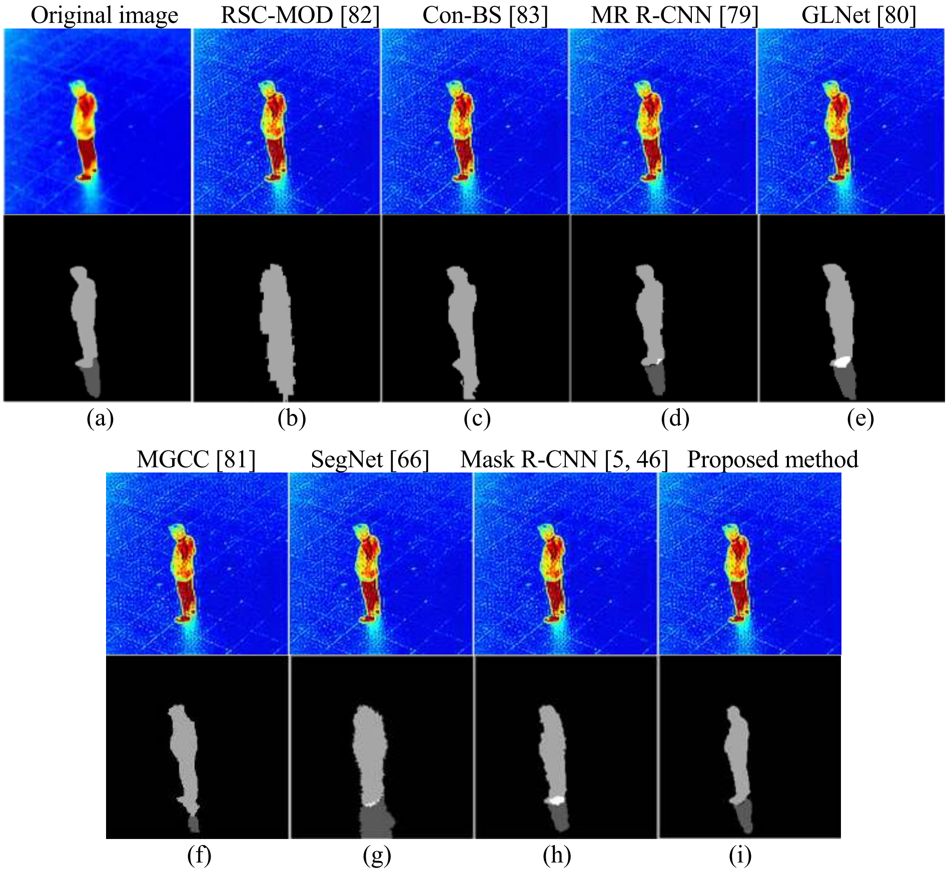


FIGURE 17. Example of pixel classification result using restored images (upscaling factor = 2 and kernel size = 13 × 13). From the top to the bottom, input images and detected results of an object and a thermal reflection. (a) ground truth masks; (b) result mask images by RSC-MOD [82]; (c) result mask images by Con-BS [83]; (d) result mask images by MR R-CNN [79]; (e) result mask images by GLNet [80]; (f) result mask images by MGCC [81]; (g) result mask images by SegNet [66]; (h) result mask images by Mask R-CNN [5], [46]; (i) result mask images by proposed method.

deblur-SRRGAN is helpful to improve object detection performance.

Furthermore, Figures 14 and 15 compare detection results derived using different methods. As shown in Tables 16–21, all state-of-the-art methods provide promising results. However, SegNet [66] provides lower localization result compared to Mask R-CNN-based methods. The reason for the difference in results is that the training loss of SegNet decreases until epoch 10 in training phase whereas training losses of Mask R-CNN-based methods decrease until

TABLE 22. Processing speed for the proposed methods (unit: ms).

Method	Processing time	Frames per second
Deblur-SRRGAN	11.42	87.56
Object detection	81.31	12.29
Total	**92.73**	**10.78**

epoch 25. This means the Mask R-CNN model learns more information from thermal datasets compared to the SegNet model. Nevertheless, the Mask R-CNN shows the lower accuracy in localization than our method because proposed light-weighted Mask R-CNN reducing the number of layers and filters show the better generalization performance in testing data than the Mask R-CNN. According to Tables 16–21 and Figures 14 and 15, the proposed method yielded better object and thermal reflection detection accuracy compared with the state-of-the-art methods.

Original image SegNet [66] Mask R-CNN [5], [46] Proposed method Figures 16 and 17 show whether the pixels corresponding to the regions of an object and a thermal reflection are well distinguished. Figure 16 shows the pixel classification comparison results using original images, and Figure 17 shows the pixel classification comparison results using restored images. According to Figures 16 and 17, the proposed method yielded the highest detection accuracy among all the models. Moreover, as shown in Figures 16 and 17, the overlapping effect between the object pixels and thermal reflection pixels detected using the proposed method is significantly smaller than that of other methods.

V. CONCLUSION

In this study, novel methods for thermal image restoration and object and thermal reflection detection were proposed. For the image restoration method, a 1-channel grayscale thermal image was converted into a 3-channel RGB thermal image; then, the image was restored using the proposed deblur-SRRGAN method, which simultaneously performed deblurring and SRR. In addition, the proposed Mask R-CNN model detected the object and thermal reflection in the 3-channel RGB thermal image that was restored using the deblur-SRRGAN method. The experiments were conducted using various databases, including our self-collected databases and numerous open databases of thermal images to compare the performances of the proposed methods with other state-of-the-art methods. That is, we trained from scratch all the methods using our self-collected databases and open databases for the comparisons. Furthermore, ablation studies were conducted to observe and compare the effects of the images without color conversion and with color conversion on the performance using various databases. Subsequently, various color spaces were used to conduct comparative experiments. The experimental results demonstrated that the proposed image restoration and object and thermal reflection detection methods outperformed the state-of-the-art methods.

We trained all methods using same databases for the fair comparisons. We conducted experiments using six databases separately, which can confirm the generalization and general performance in different environments and camera settings.

The limitation of the proposed methods in real-time task such as action recognition is processing speed when using both methods of deblur-SRRGAN and light-weighted Mask R-CNN-based object detection jointly. As shown in Table 22, our system including both methods can process 10.78 frames per one second, which cannot cope with the recognition of fast action. Therefore, we would research the method of enhancing the processing speed of our system in future work. In addition, we intend to conduct research on object skeleton generation and behavior detection methods using the obtained images (results) of this study. Moreover, studies on the application of the proposed methods to visible light cameras or near-infrared (NIR) light camera images will be conducted.

REFERENCES

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

[2] B. Oswald-Tranta, "Temperature reconstruction of infrared images with motion deblurring," *J. Sensors Sensor Syst.*, vol. 7, no. 1, pp. 13–20, Jan. 2018.

[3] FLIR Systems. (2020). *FLIR Tau 2*. [Online]. Available: https://www.flir.com/products/tau-2/

[4] G. Batchuluun, Y. W. Lee, D. T. Nguyen, T. D. Pham, and K. R. Park, "Thermal image reconstruction using deep learning," *IEEE Access*, vol. 8, pp. 126839–126858, 2020.

[5] 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, 2019.

[6] Digital Media Lab. (2020). *Dongguk Single Model Both for Thermal Image Super-Resolution Reconstruction and Deblurring, and Detection Model of Object and Thermal Reflection*. [Online]. Available: http://dm.dgu.edu/link.html

[7] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[8] Y. Matsushita, H. Kawasaki, S. Ono, and K. Ikeuchi, "Simultaneous deblur and super-resolution technique for video sequence captured by hand-held video camera," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 4562–4566.

[9] L. He, G. Li, and J. Liu, "Joint motion deblurring and superresolution from single blurry image," *Math. Problems Eng.*, vol. 2015, pp. 1–10, Jan. 2015.

[10] H. Park and K. M. Lee, "Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence," Sep. 2017, *arXiv:1709.05745*. [Online]. Available: http://arxiv.org/abs/1709.05745

[11] B. Bascle, A. Blake, and A. Zisserman, "Motion deblurring and super-resolution from an image sequence," in *Proc. Eur. Conf. Comput. Vis.*, Cambridge, U.K., Apr. 1996, pp. 571–582.

[12] B. Du, X. Ren, and J. Ren, "CNN-based image super-resolution and deblurring," in *Proc. Int. Conf. Video, Signal Image Process.* Wuhan, China: Association for Computing Machinery, Oct. 2019, pp. 70–74.

[13] X. Zhang, F. Wang, H. Dong, and Y. Guo, "A deep encoder-decoder networks for joint deblurring and super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 1448–1452.

[14] X. Zhang, H. Dong, Z. Hu, W.-S. Lai, F. Wang, and M.-H. Yang, "Gated fusion network for joint image deblurring and super-resolution," Jul. 2018, *arXiv:1807.10806*. [Online]. Available: http://arxiv.org/abs/1807.10806

[15] F. Albluwi, V. A. Krylov, and R. Dahyot, "Image deblurring and super-resolution using deep convolutional neural networks," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Aalborg, Denmark, Sep. 2018, pp. 1–6.

[16] J. U. Yun and I. K. Park, "Joint face super-resolution and deblurring using a generative adversarial network," Dec. 2019, *arXiv:1912.10427*. [Online]. Available: http://arxiv.org/abs/1912.10427

[17] F. Vankawala, A. Ganatra, and A. Patel, "A survey on different image deblurring techniques," *Int. J. Comput. Appl.*, vol. 116, no. 13, pp. 15–18, Apr. 2015.

[18] S. F. Hamood, M. S. M. Rahim, O. Farook, and D. Kasmuni, "A survey on various image deblurring methods," *J. Eng. Appl. Sci.*, vol. 11, no. 3, pp. 561–569, 2016.

[19] D. Singh and R. K. Sahu, "A survey on various image deblurring techniques," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 12, pp. 4736–4739, 2013.

[20] N. Patel and K. N. Jariwala, "A survey on image enhancement using image super-resolution and deblurring methods," *J. Emerg. Technol. Innov. Res.*, vol. 1, no. 5, pp. 359–370, 2014.

[21] S. Sahu, M. K. Lenka, and P. K. Sa, "Blind deblurring using deep learning: A survey," Jul. 2019, *arXiv:1907.10128*. [Online]. Available: http://arxiv.org/abs/1907.10128

[22] K. Nasrollahi and T. B. Moeslund, "Super-resolution: A comprehensive survey," *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1423–1468, Aug. 2014.

[23] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," Feb. 2019, *arXiv:1902.06068*. [Online]. Available: http://arxiv.org/abs/1902.06068

[24] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," Mar. 2019, *arXiv:1904.07523*. [Online]. Available: http://arxiv.org/abs/1904.07523

[25] S. H. Umale and A. M. Sahu, "A review on various techniques for image deblurring," *Int. J. Comput. Sci. Mobile Comput.*, vol. 3, no. 4, pp. 263–268, 2014.

[26] M. M. Sada and M. M. Goyani, "Image deblurring techniques—A detail review," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 4, pp. 176–188, Jan. 2018.

[27] R. Wang and D. Tao, "Recent progress in image deblurring," Sep. 2014, *arXiv:1409.6838*. [Online]. Available: http://arxiv.org/abs/1409.6838

[28] W. Yang, X. Zhang, Y. Tian, W. Wang, and J.-H. Xue, "Deep learning for single image super-resolution: A brief review," Jul. 2018, *arXiv:1808.03344*. [Online]. Available: http://arxiv.org/abs/1808.03344

[29] S. Borman and R. L. Stevenson, "Super-resolution from image sequences—A review," in *Proc. Midwest Symp. Circuits Syst.*, Notre Dame, IN, USA, Aug. 1998, pp. 374–378.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] 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, May 2015.

[36] E. S. Jeon, J. H. Kim, H. G. Hong, G. Batchuluun, and K. R. Park, "Human detection based on the generation of a background image and fuzzy system by using a thermal camera," *Sensors*, vol. 16, no. 4, pp. 1–31, 2016.

[37] P. Kumar, A. Mittal, and P. Kumar, "Study of robust and intelligent surveillance in visible and multimodal framework," *Informatica*, vol. 32, pp. 63–77, Apr. 2008.

[38] G. Batchuluun, N. R. Baek, D. T. Nguyen, T. D. Pham, and K. R. Park, "Region-based removal of thermal reflection using pruned fully convolutional network," *IEEE Access*, vol. 8, pp. 75741–75760, 2020.

[39] G. Batchuluun, Y. G. Kim, J. H. Kim, H. G. Hong, and K. R. Park, "Robust behavior recognition in intelligent surveillance environments," *Sensors*, vol. 16, no. 7, pp. 1–23, 2016.

[40] 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.

[41] 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.

[42] 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.

[43] G. Batchuluun, D. T. Nguyen, T. D. Pham, C. Park, and K. R. Park, "Action recognition from thermal videos," *IEEE Access*, vol. 7, pp. 103893–103917, Aug. 2019.

[44] Mathworks. (2020). *Colormap*. [Online]. Available: https://www.mathworks.com/help/matlab/ref/jet.html

[45] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, Oct. 2017, pp. 2999–3007.

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

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[48] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.

[49] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," Mar. 2014, *arXiv:1411.4038*. [Online]. Available: http://arxiv.org/abs/1411.4038

[50] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Jan. 2015, *arXiv:1506.01497*. [Online]. Available: http://arxiv.org/abs/1506.01497

[51] Digital Media Lab. (2020). *Dongguk Thermal Image Database (DTh-DB) and Dongguk Items & Vehicles Database (DI&V-DB)*. [Online]. Available: http://dm.dgu.edu/link.html

[52] 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.

[53] J. Davis and M. Keck, "A two-stage approach to person detection in thermal imagery," in *Proc. Workshop Appl. Comput. Vis.*, Breckenridge, CO, USA, Jan. 2005, pp. 6–111.

[54] R. Miezianko, "Terravic research infrared database-terravic motion infrared database," in *Proc. IEEE OTCBVS WS Ser. Bench*, to be published.

[55] R. Miezianko, "Terravic research infrared database-terravic weapon infrared database," in *Proc. IEEE OTCBVS WS Ser. Bench*, to be published.

[56] R. Bergevin, P.-L. St-Charles, and G.-A. Bilodeau, "Mutual foreground segmentation with multispectral stereo pairs," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Venice, Italy, Oct. 2017, pp. 375–384.

[57] K. Van Beeck, K. Van Engeland, J. Vennekens, and T. Goedeme, "Abnormal behavior detection in LWIR surveillance of railway platforms," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Lecce, Italy, Aug./Sep. 2017, pp. 1–6.

[58] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient night gait recognition based on template matching," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Hong Kong, Aug. 2006, pp. 1000–1003.

[59] 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. 201–208.

[60] (2020). *ASL Datasets*. [Online]. Available: https://projects.asl.ethz.ch/datasets/doku.php?id=ir:iricra2014.q11

[61] NVIDIA Corporation. (2019). *NVIDIA Titan X*. [Online]. Available: https://www.nvidia.com/en-us/geforce/products/10series/titan-x-pascal/

[62] Keras. (2019). *Keras: The Python Deep Learning Library*. [Online]. Available: https://keras.io/

[63] OpenCV. (2019). *OpenCV: Open Source Computer Vision*. [Online]. Available: http://opencv.org/

[64] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 105–114.

[65] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[66] 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.

[67] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York, NY, USA: Springer-Verlag, 1998.

[68] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[69] 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.

[70] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommun. Syst.*, vol. 49, no. 1, pp. 35–48, Jan. 2012.

[71] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.

[72] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[73] X. Lu, H. Yuan, P. Yan, Y. Yuan, and X. Li, "Geometry constrained sparse coding for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1648–1655.

[74] X. Zhang, C. Li, Q. Meng, S. Liu, Y. Zhang, and J. Wang, "Infrared image super resolution by combining compressive sensing and deep learning," *Sensors*, vol. 18, no. 8, pp. 1–15, 2018.

[75] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[76] L. Derczynski, "Complementarity, F-score, and NLP evaluation," in *Proc. Int. Conf. Lang. Resour. Eval.*, Portorož, Slovenia, May 2016, pp. 261–266.

[77] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Addison-Wesley, 2005.

[78] Q. Zhang and J. Wu, "Image super-resolution using windowed ordinary Kriging interpolation," *Opt. Commun.*, vol. 336, pp. 140–145, Feb. 2015.

[79] Y. Zhang, J. Chu, L. Leng, and J. Miao, "Mask-refined R-CNN: A network for refining object details in instance segmentation," *Sensors*, vol. 20, no. 4, p. 1010, Feb. 2020.

[80] C.-Y. Lin, Y.-C. Chiu, H.-F. Ng, T. K. Shih, and K.-H. Lin, "Global-and-local context network for semantic segmentation of street view images," *Sensors*, vol. 20, no. 10, p. 2907, May 2020.

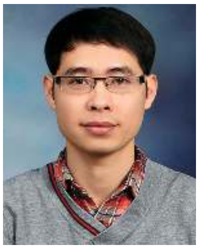
[81] W. Feng, X. Li, G. Gao, X. Chen, and Q. Liu, "Multi-scale global contrast CNN for salient object detection," *Sensors*, vol. 20, no. 9, p. 2656, May 2020.

[82] K. Lenac, I. Maurovic, and I. Petrovic, "Moving objects detection using a thermal camera and IMU on a vehicle," in *Proc. Int. Conf. Electr. Drives Power Electron. (EDPE)*, Tatranska Lomnica, Slovakia, Sep. 2015, pp. 212–219.

[83] J. W. Davis and V. Sharma, "Background-subtraction in thermal imagery using contour saliency," *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 161–181, Feb. 2007.

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 also designed the entire system and wrote the original draft of article.

JIN KYU KANG received the B.S. degree in electronics and electrical engineering from Dongguk University, Seoul, South Korea, in 2016, where he is currently pursuing the combined course of M.S. and Ph.D. degrees in electronics and electrical engineering. His research interests include biometrics and deep learning. He also helped to perform the experiments and analysis.



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 also supervised this study 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 also helped experiments and analysis.

MUHAMMAD ARSALAN received the B.S. degree in computer engineering from COMSATS University Islamabad, Pakistan, in 2012, the M.S. degree in computer science from the NCBA&E, Lahore, Pakistan, in 2016, and the Ph.D. degree in electronics and electrical engineering from Dongguk University, in 2020. He has been a Professor with the Division of Electronics and Electrical Engineering, Dongguk University, since September 2020. His research interests include computer vision and deep learning. He also 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 also helped experiments and analysis.