

# Applying Depthwise Separated Neural Network with Color Space Adjustment to Auto-colorization of Thermal Infrared Images

Ming-Tsung Yeh,<sup>1\*</sup> Wei-Yin Lo,<sup>2</sup> Yi-Nung Chung,<sup>2</sup> and Pei-Syuan Lu<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, National Chin-Yi University of Technology,  
57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan

<sup>2</sup>Department of Electrical Engineering, National Changhua University of Education,  
No. 1, Jinde Rd., Changhua City, Changhua County 50007, Taiwan

(Received December 30, 2022; accepted June 6, 2023)

**Keywords:** auto-colorization, color space adjustment, CAE, GAN, thermal infrared image

A general surveillance camera with a near-infrared illuminator provides a night vision function, but it is difficult to take a picture under foggy or smoky conditions, in a heavy rainfall environment, or under direct exposure to the sun because of poor object temperature reflection. A thermal infrared (TIR) camera can have better imaging to reflect objects in bad environments, and they have many applications in safe driving and military and scientific fields for all-weather surveillance. However, TIR images are mainly presented in grayscale, which causes the applications of TIR images to be limited and used only for rough object recognition. In previous studies, auto-colorization by predicting luminance and chrominance from grayscale images at the same time was typically performed, but the results were always blurry and abnormally colorized images. This study proposes the Depthwise Separated Colorization Generative Adversarial Network (DSCGAN) to colorize TIR images and overcome these drawbacks. Initially, the preprocessing light channel convolutional autoencoder (PLCAE) is proposed to generate the predicted L channel of the International Commission on Illumination LAB color space (CIELAB) that is used to restore some lost luminance information. Then, this predicted L channel is used as input to the proposed Colorization Generative Adversarial Network (CGAN) to create the AB channel. Finally, the data from L, A, and B channels are converted to the RGB visible light image. The experimental results indicate that our proposed PLCAE can efficiently enhance luminance details and achieve an accuracy rate of 0.9773. The proposed CGAN advances colorization accuracy and improves the peak signal-to-noise ratio (*PSNR*) to more than 26 dB. The colorized TIR images have almost the same color as the visible light images and clearly maintain object textures and details.

## 1. Introduction

The IR camera uses infrared radiation to take images. The infrared spectrum is between the microwave and visible light ranges, and is invisible to the human eye. In general, infrared cameras are sensitive to wavelengths from 0.75 to 15  $\mu\text{m}$ , which can be distinguished into four

\*Corresponding author: e-mail: [mtyeh@ncut.edu.tw](mailto:mtyeh@ncut.edu.tw)  
<https://doi.org/10.18494/SAM4296>

classes: near-infrared (NIR), short-wave infrared (SWIR), medium-wave infrared (MWIR), and long-wave infrared (LWIR) in the infrared spectral field, as shown in Fig. 1. Images are taken at different wavelengths in a variety of scenes, as shown in Fig. 2. Most night-vision cameras using active NIR illumination take pictures in poorly lit environments. However, these cameras have difficulty taking photos under foggy or smoky conditions, in a heavy rainfall environment, or under direct exposure to the sun. Thermal infrared (TIR) cameras that use LWIR are ideal for solving imaging problems in all-weather and complex environments. The SWIR camera can capture reflected light under peak sun illumination and can be used for daytime starlight imaging. The MWIR camera can detect invisible gas leaks to the human eye, and its imagery provides precise details. LWIR is widely used in infrared technology. Figure 2 shows that the LWIR camera is helpful in taking images through fog and smoke, or under high-humidity conditions. LWIR cameras are very affordable today, and they also have better performance for most applications.

TIR images, unfortunately, are mainly presented in grayscale and thus have limited applications. Accordingly, colorized infrared images, wherein the original color is restored, can improve their practicality in scientific, industrial, commercial, military, and medical fields. Three methods are used to colorize a grayscale image: interaction-based, exemplar-based, and learning-based. The interaction-based method is better for pointing out the regional color on a grayscale image. This colorization method requires heavy manual work, and Yatziv and Sapiro proposed a chrominance blending method to reduce human interaction.<sup>(2)</sup> Bugeau *et al.* and Zhang *et al.* developed different exemplar-based methods to colorize gray images.<sup>(3,4)</sup> Their proposed methods need a set of reference images used to guide the colorization. Finding a higher relational reference image with a target image that matches similar objects is essential for transferring color information. Although this approach reduces human effort, its performance, which relies on the complexity of the target image, is unsatisfactory. These two methods need some manual intervention and are only partially automatic.

The learning-based method uses many training samples to find the color features and apply them to a grayscale image for colorization. Manual scribbles and selected reference images are unnecessary in this method, but many diverse training images are required. Although the

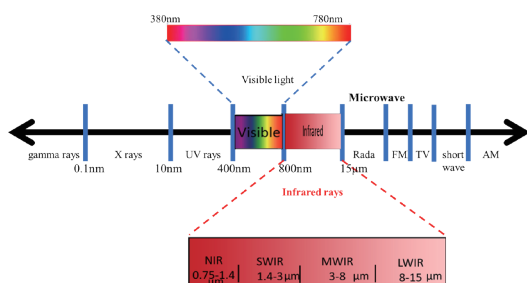


Fig. 1. (Color online) Spectrum distribution.

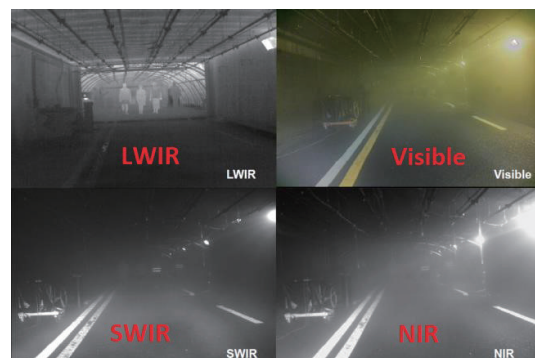


Fig. 2. (Color online) Images shot in a foggy tunnel with four different wavelengths.<sup>(1)</sup>

training cost is high, this method can automatically colorize grayscale target images after learning. Larsson *et al.* applied the Visual Geometry Group (VGG) 16 deep convolutional network to predict hue and chroma distributions combined with lightness to generate a color image.<sup>(5)</sup> Zhang *et al.* used deep CNN to train over one million images and increased the diversity of colors by class-rebalancing.<sup>(6)</sup> After the Generative Adversarial Network (GAN) was developed by Goodfellow *et al.*,<sup>(7)</sup> auto-coloring approaches based on the GAN have demonstrated impressive performance.<sup>(8–13)</sup>

In recent years, auto-coloring technologies have been applied to infrared image colorization. In some works, neural network models were trained directly using the grayscale infrared image and paired visible images as training sets. Berg *et al.* proposed to use an autoencoder structure to generate an RGB image directly from TIR input.<sup>(14)</sup> Tao *et al.* proposed using the encoder-decoder architecture with skip connections to convert infrared images to RGB images.<sup>(15)</sup> Kuang *et al.* applied the conditional GAN to TIR image colorization and obtained better results.<sup>(16)</sup> In other studies, grayscale IR images were used as the input domain and applied to a GAN-based training network to generate RGB images.<sup>(17–19)</sup> There are better options than the traditional deep neural network. However, these methods have a common weakness, that is, the grayscale image is directly applied to generate multichannel output. Some abnormal colors and artifact effects are produced in the colorized target images because some detailed information is lost during the training stage. Although Du *et al.* proposed the reference component matching module to choose proper color components as the input reference auxiliary for improving irregular colorization, the training model was more complicated.<sup>(11)</sup>

To improve the TIR image colorization, we propose the Depthwise Separated Colorization Generative Adversarial Network (DSCGAN). The main innovation is the introduction of two stages that separate the training model and the application of the International Commission on Illumination LAB color space (CIELAB) conversion. In the first stage, the preprocessing light channel convolutional autoencoder (PLCAE) is proposed to generate the predicted L channel from the TIR image that restores and compensates for the loss of some luminance details caused by the LWIR camera. Then, this predicted L channel is used as input to the proposed Colorization Generative Adversarial Network (CGAN) to create the AB channel. Finally, the L, A, and B channels can be generated and converted to RGB color images. This proposed approach can simplify training complications. Each stage only predicts one or two channels that can speed up training model convergence. The CGAN is based on the cycle GAN that can solve fewer paired dataset issues for TIR and visible paired images. The unpaired images can be used as input to train the CGAN and generate more stable and suitable AB channels.

## 2. Cameras and Proposed Methods

### 2.1 Cameras

The TIR camera used in this study is shown in Fig. 3. It is an LWIR micro-thermal camera module. The specifications of the FLIR Lepton 3.5 camera are shown in Table 1. The resolution of the LWIR images is  $160 \times 120$ . This TIR camera can take eight frames per second (FPS). The

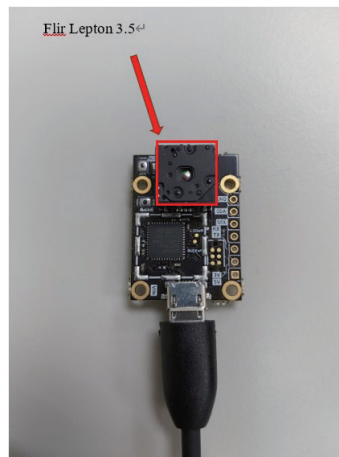


Fig. 3. (Color online) FLIR Lepton 3.5.

Table 1  
Specifications of FLIR lepton 3.5 camera.

Overview	Lepton 3.5
Sensor technology	Uncooled VOx microbolometer
Spectral range	Longwave infrared, 8 to 14 $\mu\text{m}$
Array format	160 $\times$ 120, progressive scan
Pixel size	12 $\mu\text{m}$
Effective frame rate	8.7 Hz (commercial application exportable)
Temperature compensation	Automatic. Output image independent of camera temperature.
Maximum aperture	f/1.1
Output format	User-selectable 14-bit, 8-bit (AGC applied), or 24-bit RGB (AGC and colorization applied)

visible camera used for the deep learning model is shown in Fig. 4. The resolution of the RGB images is 1280  $\times$  720. The C270 HD WEBCAM can take 30 photos per second.

## 2.2 Proposed methods

### 2.2.1 Datasets and data collection methods

The KAIST-MS dataset<sup>(20)</sup> was used in this study to compare the infrared image auto-coloring performance with the different paper experimental results. The KAIST-MS dataset is widely used to evaluate the auto-colorization effects of IR images. There are two types of IR images in this dataset, near and long-wavelength, and the dataset has an aligned pair of visible and thermal images captured from day/night traffic scenes. In addition, in this study, we also collected LWIR images day and night at the same location to find a way to reconstruct the daytime visible images from a dark environment. The graphical user interface (GUI) used to collect the visible and LWIR images in this study is shown in Fig. 5. The upper left and bottom left represent the real-time visible and LWIR images, respectively, and the refresh rate is set to 8 FPS, and the upper right and bottom right are the capture results after the user pushes the capture button. Since the TIR and visible cameras have different resolutions, the visible images are resized, cropped, and shifted using image processing to align with the TIR images in the GUI



Fig. 4. (Color online) C270 HD WEBCAM.

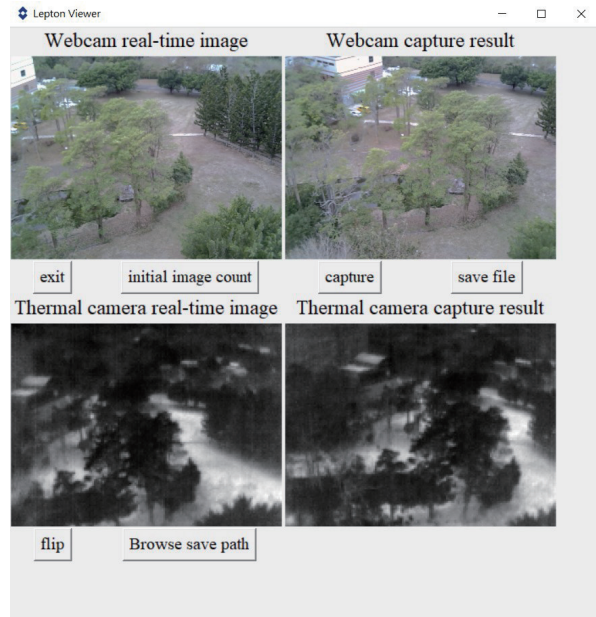


Fig. 5. (Color online) GUI for data collection.

calculation. In this study, the LWIR and visible images were separated from the luminance channel of the CIELAB color space to provide the PLCAE training samples and answers.

### 2.2.2 PLCAE methods

In deep learning, the neural network always extracts features to conduct the auto-colorization of TIR images. To obtain more features in the designed neural network, we proposed a preprocessing image approach to convert RGB into the CIELAB color space with the most features in the entire color space that people define. The RGB color space can be transformed to the XYZ color space using Eq. (1), and the XYZ color space can be converted to the LAB color space using Eqs. (2) and (3).<sup>(21)</sup> In Eq. (2), the range of the  $L^*$  channel is between 0 and 100, that of the  $a^*$  channel is between 127 and  $-128$ , and that of the  $b^*$  channel is between 127 and  $-128$ . In Eq. (3), the function domain is divided into two ranges to avoid a gradient of infinity at  $t = 0$ .

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$\begin{cases} L^* = 116 * f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* = 500 * f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \\ b^* = 200 * f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \end{cases} \quad (2)$$

Here,  $L^*$  is luminance,  $a^*$  is the chrominance that represents the red axis (positive) to the green axis (negative),  $b^*$  is the chrominance that represents the yellow axis (positive) to the blue axis (negative), and  $X_n$ ,  $Y_n$ , and  $Z_n$  represent the tristimulus values of the reference white point.

$$f(t) = \begin{cases} t^{1/3}, & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left( \frac{29^2}{6} t + \frac{4}{29} \right), & \text{otherwise} \end{cases} \quad (3)$$

The PLCAE structure and neural network model used in this study are shown in Figs. 6 and 7, respectively. The convolutional autoencoder (CAE) can make the input image for reconstructing the target image.<sup>(22)</sup> The CAE can extract the light features of the infrared image through the encoder and reassemble this information in the decoder. To transform the dark environment images into visible daytime colorized images, the L channel of the nighttime IR image needs to be reconstructed into the L channel of the same target day visible light images. This action is the same as other IR image conversion that applies PLCAE to convert the raw IR L channel into the L channel of the color visible image at the same position. The definition of the

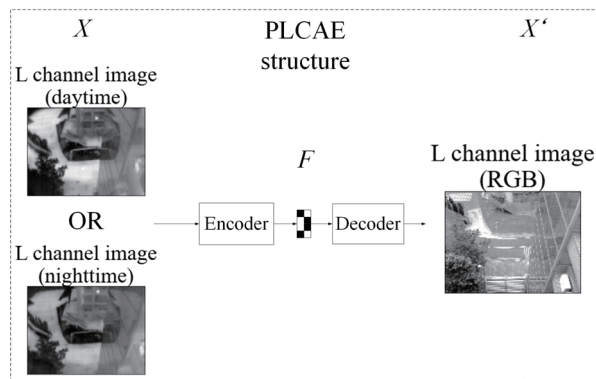


Fig. 6. (Color online) PLCAE structure used in this study.

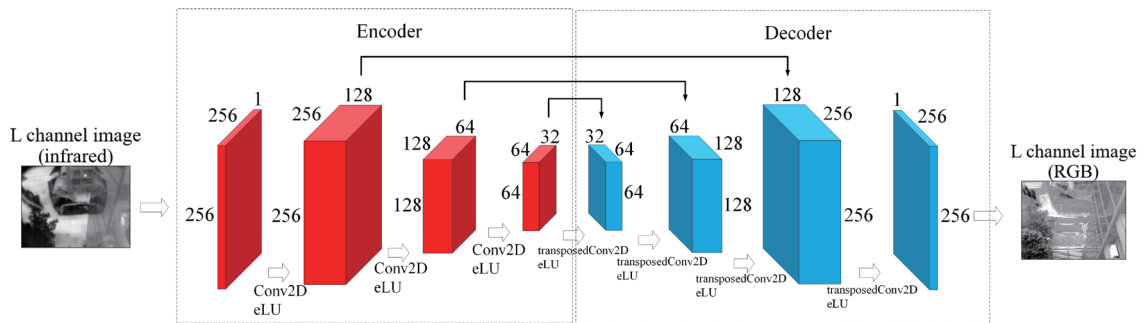


Fig. 7. (Color online) PLCAE neural network model used in this study.

PLCAE can be represented by Eqs. (4) and (5). The encoder and decoder processes are calculated using Eq. (6). The PLCAE uses the mean squared error (MSE) as the loss function for backpropagation. The calculation of the MSE loss function is shown in Eq. (7). In the PLCAE, the grayscale TIR image is copied three times to convert RGB into the L channel of CIELAB. The PLCAE input and output images are resized to  $256 \times 256 \times 1$ . Moreover, this predicted L channel is used as input for the CGAN to generate the AB channel of CIELAB to achieve the auto-colorization of TIR images.

$$\psi, \varphi = \arg \min_{\psi, \varphi} \|X' - (\varphi \circ \psi) X\|^2 \quad (4)$$

$$\psi: X \rightarrow F, \varphi: F \rightarrow X' \quad (5)$$

Here,  $X$  is the input vector,  $F$  is the feature vector,  $X'$  is the target vector,  $\psi$  represents the conversion of the input vector to feature vectors by the CAE, and  $\varphi$  represents the calculation of the feature vectors to the target images by the decoder of the CAE.

$$\text{If } \mathbf{x} \in \mathbb{R}^d = X, \text{ if } \mathbf{h} \in \mathbb{R}^p = F, \begin{cases} \mathbf{h} = \sigma(W\mathbf{x} + \mathbf{b}) \\ X' = \sigma(W'\mathbf{h} + \mathbf{b}') \end{cases} \quad (6)$$

$$L(\mathbf{x}, X') = \|\mathbf{x} - X'\|^2 = \|\mathbf{x} - \sigma'(W'(\sigma(W\mathbf{x} + \mathbf{b}))) + \mathbf{b}'\|^2 \quad (7)$$

Here,  $\sigma$  and  $\sigma'$  represent the encoder and decoder activation functions,  $W$  and  $W'$  represent the encoder and decoder weights, and  $b$  and  $b'$  represent the encoder and decoder weighting biases, respectively.

### 2.2.3 CGAN methods

Auto-colorization is a well-known deep learning application, and the effective method is to use GAN and the L channel to predict the AB channel.<sup>(23)</sup> The CGAN is a type of cycle GAN. In the CGAN, the L channel represents  $X$  domains and the AB channel represents  $Y$  domains.<sup>(23)</sup> There are two generators mapping  $G: X \rightarrow Y$  and  $F: Y \rightarrow X$ . Also, there are two discriminators  $D_X$  and  $D_Y$ , where  $D_X$  can distinguish between real and fake  $X$  domain images that the  $F$  generator predicts using  $Y$  domain data, and  $D_Y$  can distinguish between real and fake  $Y$  domain images that the  $G$  generator predicts using  $X$  domain data. The  $G$  and  $D_Y$  of the CGAN structure are shown in Fig. 8. The CGAN generator and discriminator model structure are shown in Fig. 9.

In this study, we denote  $X$  domain samples as  $\{x_i\}_{i=1}^N \in X$  and  $Y$  domain samples as  $\{y_j\}_{j=1}^M \in Y$ .  $N$  is the total number of  $X$  domain samples.  $M$  is the total number of  $Y$  domain samples. In addition, the data distributions are denoted as  $x \sim p_{data}(x)$  and  $y \sim p_{data}(y)$ . The CGAN adversarial loss is shown by Eq. (8). To reduce the space of the possible mapping function, the CGAN also uses the cycle consistency loss shown as Eq. (9), e.g.,  $G(y)$  or  $G(F(x))$  should output  $Y$  domain images. The CGAN full objective loss function is shown by Eqs. (10) and (11).

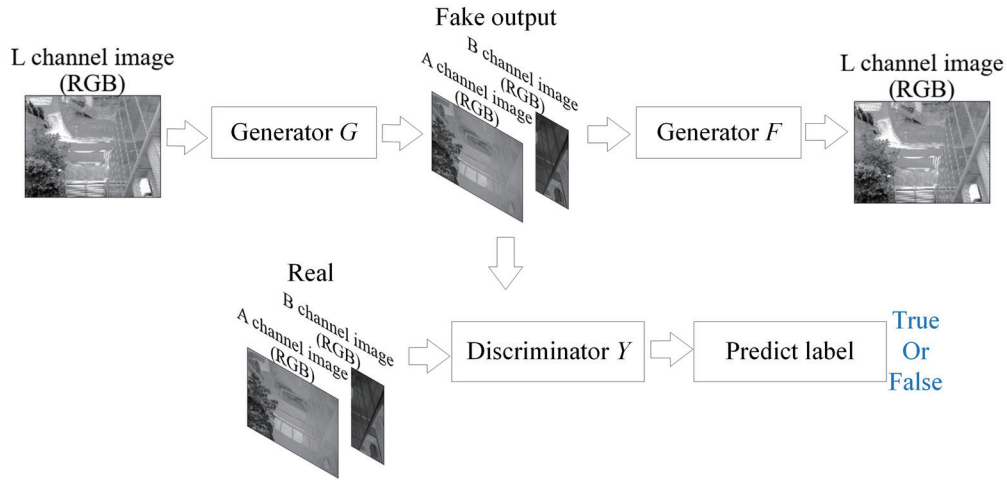


Fig. 8. (Color online) CGAN structure used in this study.

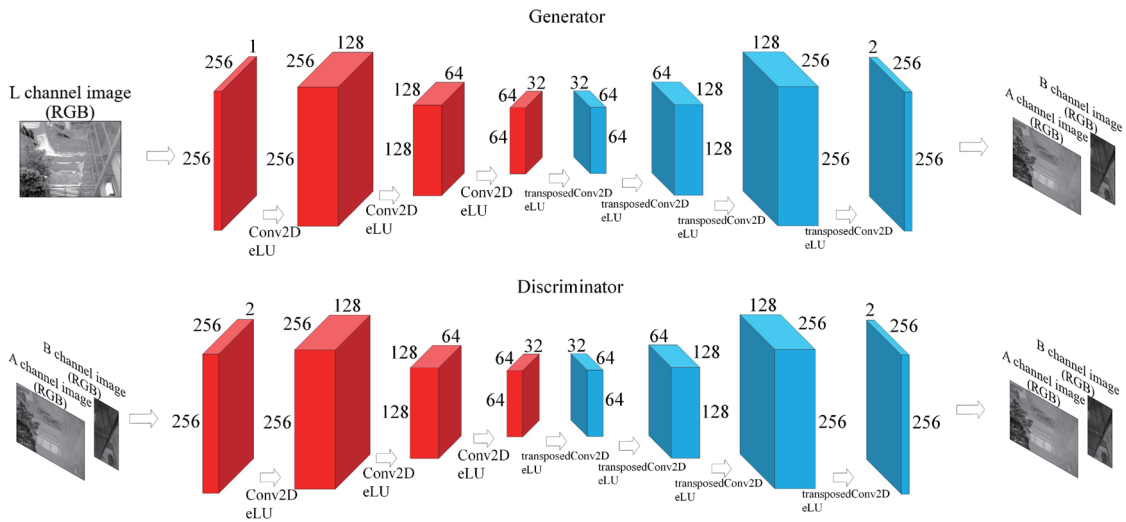


Fig. 9. (Color online) CGAN model structure used in this study.

$$L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (8)$$

Here, the purpose is to minimize the generator  $G(y)$  and the  $D_Y$  target, and the generator  $F(x)$  and the  $D_X$  target.

$$L_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|] \quad (9)$$

$$L_{GAN}(G, D_Y, X, Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (10)$$



Here,  $\lambda$  controls the weight of the two objectives.

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \quad (11)$$

In the CGAN, the  $256 \times 256 \times 1$  L channel visible images are used with the  $G$  generator to predict the  $256 \times 256 \times 2$  AB visible images. Finally, we can combine the predicted L channel of the PLCAE and the AB channels of the CGAN generator output to produce the LAB color space of the targeted image. The LAB color space can be transformed to the XYZ color space using Eqs. (12) and (13), and the XYZ color space can be transformed to the RGB color space using Eq. (14).

$$\begin{cases} X = X_n * f^{-1}\left(\frac{1}{116}(L^* + 16) + \frac{1}{500}a^*\right) \\ Y = Y_n * f^{-1}\left(\frac{1}{116}(L^* + 16)\right) \\ Z = Z_n * f^{-1}\left(\frac{1}{116}(L^* + 16) + \frac{1}{200}b^*\right) \end{cases} \quad (12)$$

$$f^{-1}(t) = \begin{cases} t^3, & \text{if } t > \frac{6}{29} \\ 3\left(\frac{6}{29}\right)^2 * \left(t - \frac{4}{29}\right), & \text{otherwise} \end{cases} \quad (13)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2404 & -1.5371 & -0.49853 \\ -0.9692 & 1.8759 & 0.04155 \\ 0.0556 & -0.2040 & 1.0573 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (14)$$

### 3. Experimental Results and Discussion

In this study, we propose deep learning neural networks in two stages to convert TIR images into RGB visible photos. This section shows the results regarding the loss, accuracy, and peak signal-to-noise ratio (*PSNR*), which are used to evaluate the training performance.

#### 3.1 PLCAE loss and accuracy graph

The PLCAE loss graph in Fig. 10 shows that the PLCAE model converges from a high point to a small value quickly, and the training process is very smooth during 2000 epochs. Our training samples were collected from random frames of video streaming, which makes these datasets have diverse scenes in each sampling. Therefore, the MSE value of the training loss function is sometimes increased at the beginning of the training process. However, various datasets improve the trained network generalization

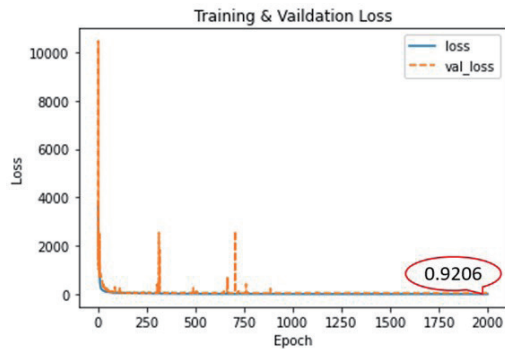


Fig. 10. (Color online) PLCAE loss graph.

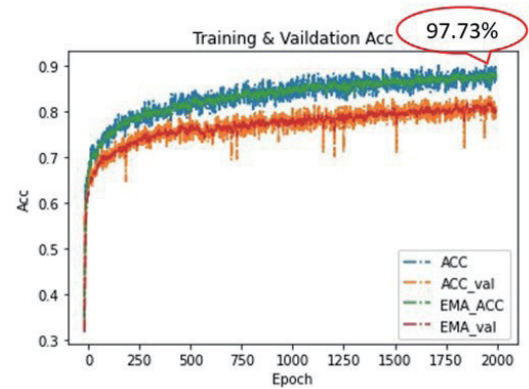


Fig. 11. (Color online) PLCAE accuracy graph.

performance. The loss value can converge to 0.9206 at the end. The PLCAE accuracy graph in Fig. 11 shows that the accuracy converges to 0.9773. In other words, the PLCAE is a very appropriate model for converting TIR luminance images into RGB luminance images before auto-coloring.

### 3.2 CGAN loss and accuracy graph

The CGAN generator  $G$  loss graph in Fig. 12 shows an unstable loss curve in the training process of generator  $G$ . This often happens in GAN because the discriminator and generator compete with each other. The generator loss curve is more stable until above 10000 epochs, and the loss value finally converges to 0.7900.

In this study, we used the  $PSNR$  to evaluate the auto-colorization result objectively. We denoted the pixel value of the ground truth image as  $Y_i$  and the pixel value of the predicted image as  $\hat{Y}_i$ . The  $PSNR$  calculation formula is shown as Eqs. (15) and (16). The numerator in the  $PSNR$  formula is simplified as one because the training images have been normalized. The CGAN generator  $G$   $PSNR$  graph is shown in Fig. 13. The  $G$   $PSNR$  value can reach 26.03 dB.

$$PSNR = 10 * \log_{10} \left( \frac{1.0^2}{MSE} \right) = 20 * \log_{10} \left( \frac{1.0^2}{\sqrt{MSE}} \right) \quad (15)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (16)$$

### 3.3 Comparison of auto-colorization results of proposed method and other methods

The experimental results obtained using the FLIR dataset<sup>(24)</sup> are shown in Figs. 14 and 15, including the PLCAE model, CGAN, and LAB-to-RGB transform. In this study, the training sample contents included most of the scenery on the road. Figure 16 shows that the deep learning

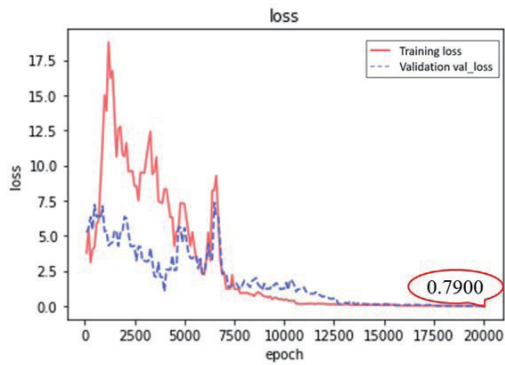


Fig. 12. (Color online) CGAN generator  $G$  loss graph.

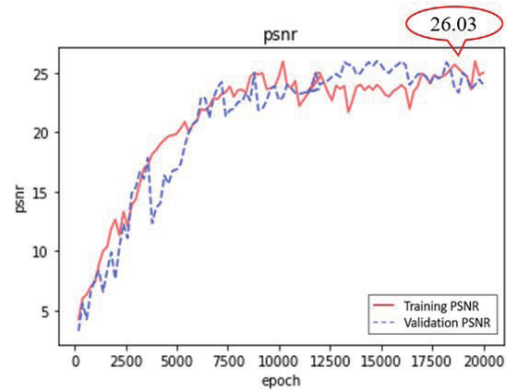


Fig. 13. (Color online) CGAN generator  $G$  PSNR graph.

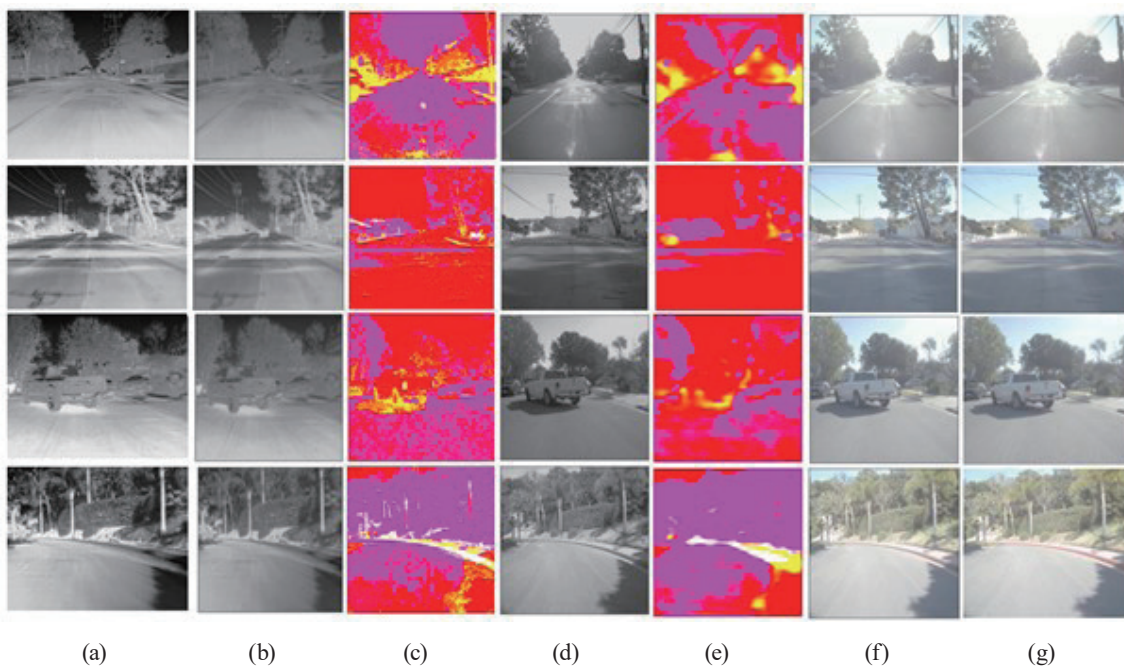


Fig. 14. (Color online) Experimental results 1 obtained using the FLIR dataset in this study. (a) TIR images, (b) TIR L channel images, (c) LAB of visible images, (d) PLCAE-predicted L channel images, (e) CGAN-predicted and combined LAB images, (f) final predicted visible images, and (g) target visible images.

model cannot predict the color correctly. For example, Fig. 16 indicates that the red color sometimes is predicted to be a gray color. Nevertheless, it should be noted that using the PLCAE model for the prediction can yield superior results in reconstructing visible image details and textures, thereby achieving a higher PSNR than other methods, as shown in Table 2. The reconstructed visible L channel images predicted using the PLCAE model are shown in Fig. 17.

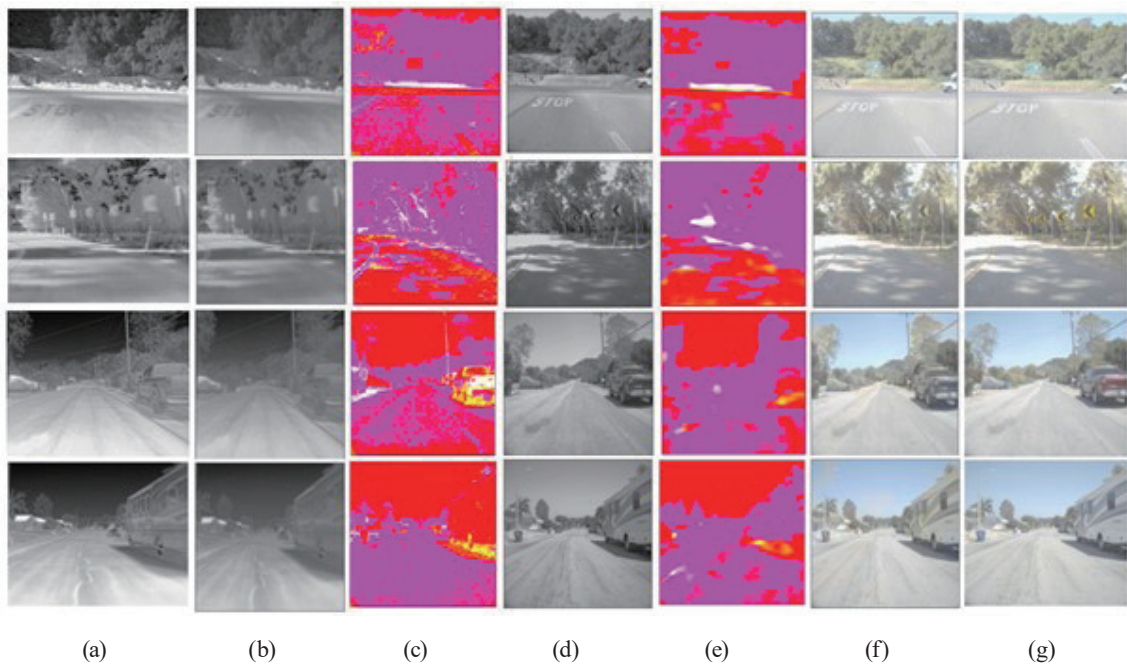


Fig. 15. (Color online) Experimental results 2 obtained using the FLIR dataset in this study. (a) TIR images, (b) TIR L channel images, (c) LAB of visible images, (d) PLCAE-predicted L channel images, (e) CGAN-predicted and combined LAB images, (f) final predicted visible images, and (g) target visible images.



Fig. 16. (Color online) Inaccurate color prediction in specific images. (a) inaccurate red color prediction in images and (b) target visible images.

We compared the auto-coloring performance of infrared images with those of other methods using the KAIST-MS dataset. The *PSNR* results of the different methods are shown in Table 2,

Table 2  
PSNR results of different methods.

Method	PSNR
Naïve	9.55
TIR2RGB	14.7
TIC-CGAN	15.85
Proposed	26.03



Fig. 17. (Color online) Comparison of PLCAE-predicted L channel and ground truth L channel. (a) and (c) PLCAE-predicted L channel images. (b) and (d) Ground truth L channel images.

which shows that the best *PSNR* is obtained by the proposed method. The prediction results obtained with the different auto-coloring approaches are demonstrated in Fig. 18, which shows that, compared with the other approaches, the image obtained by the proposed method is closest to the target visible images.

The self-collected images for testing the performance of the proposed method are shown in Fig. 19. In the daytime images, the infrared auto-coloring result can achieve colorization similar to that observed in the KAIST-MS dataset. In contrast, the nighttime colorized result has a worse performance because the PLCAE module cannot fully reconstruct the L channel details of the related visible image. Figure 20 shows the performance of the proposed method for pictures taken under different bad weather conditions in the AAU RainSnow Traffic Surveillance Dataset.<sup>(25)</sup> While the proposed method performs well under rainy and foggy conditions, its effectiveness is limited under snowy conditions.

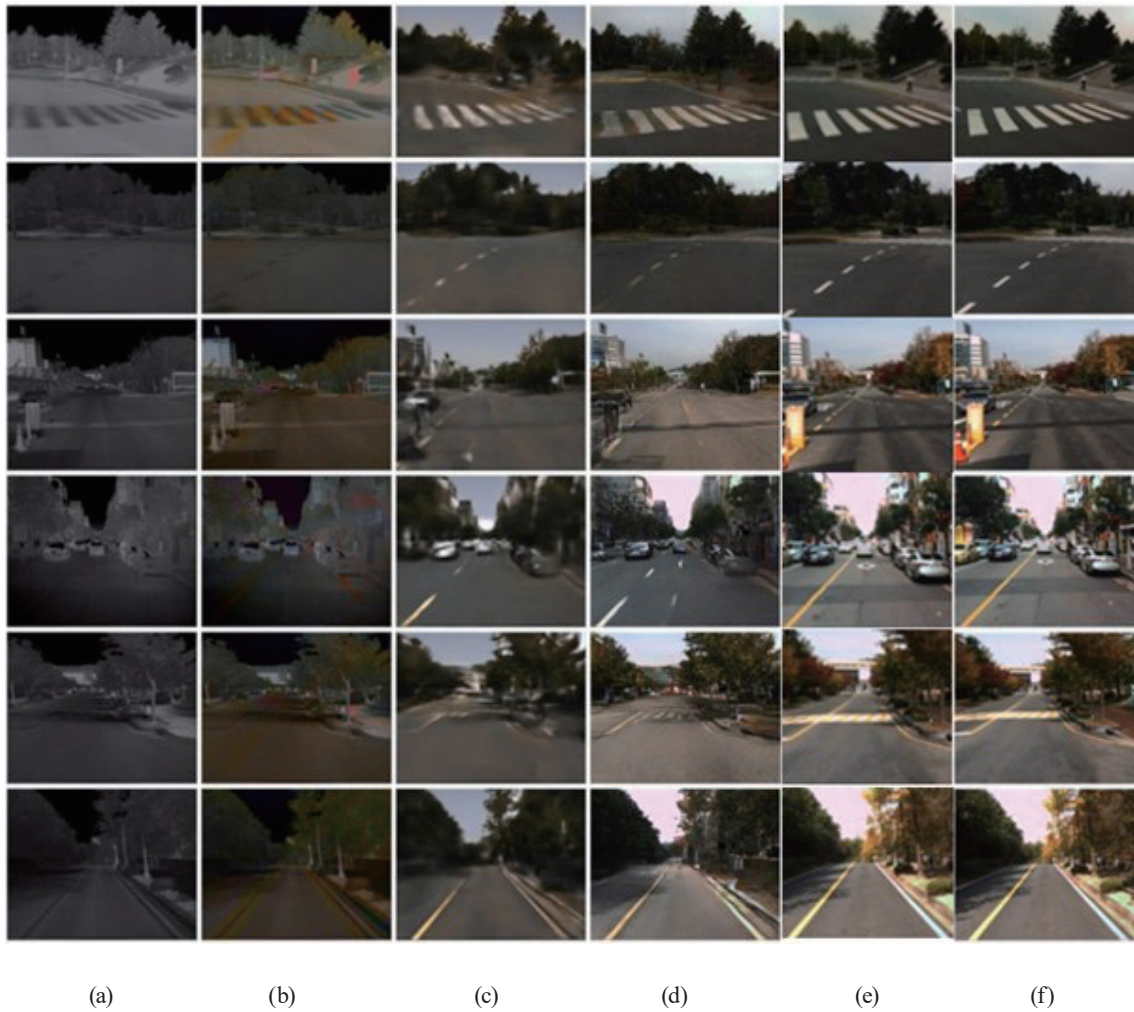


Fig. 18. (Color online) Prediction results with different auto-coloring approaches. (a) TIR image, (b) naïve images, (c) TIR2 RGB images, (d) TIR-CGAN images, (e) images obtained by the proposed method, and (f) target visible images.

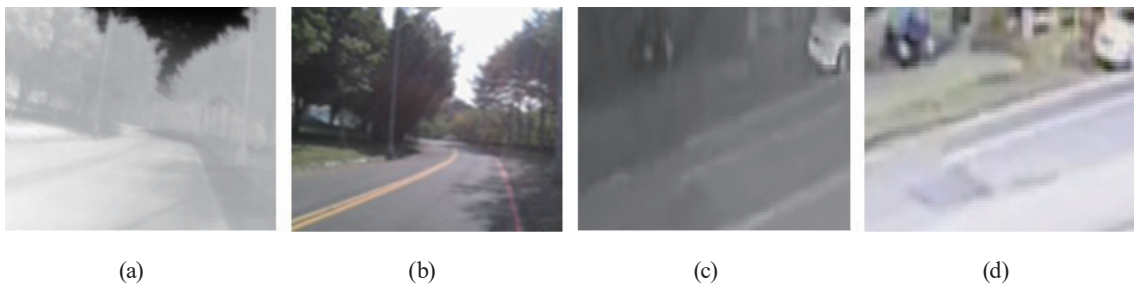


Fig. 19. (Color online) Self-collected images for testing the performance of the proposed method. (a) Daytime TIR image, (b) daytime predicted image, (c) nighttime TIR image, and (d) nighttime predicted image.



Fig. 20. (Color online) AAU RainSnow Traffic Surveillance Dataset used for evaluating the performance of the proposed method in low-temperature environments during a snowfall. (a) Original image of daytime rainfall, (b) TIR image of daytime rainfall, (c) predicted image of daytime rainfall, (d) original image of daytime fog, (e) TIR image of daytime fog, (f) predicted image of daytime fog, (g) original image of daytime snowfall, (h) TIR image of daytime snowfall, and (i) predicted image of daytime snowfall.

### 3.4 Discussion

In this study, nighttime auto-colorization did not perform as well as daytime auto-colorization. The reason can be seen in Figs. 19 and 20. Figure 19 shows that there are many different structures between nighttime and daytime. Figure 20 shows that low-temperature environments during a snowfall can also affect the ability to reconstruct TIR details accurately.

There are two solutions to address this problem. The first is to collect diverse TIR images of daytime scenes and other pictures under various weather conditions in the same location as training sets. The second is to separately train the PLCAE module for different weather conditions or times of the day. However, this would increase the complexity of network training.

#### 4. Conclusions

In this work, TIR images were auto-colored by two-stage deep learning. In the first-stage, the proposed PLCAE converted TIR luminance images into RGB luminance images to enhance luminance details before auto-coloring. The PLCAE accuracy can reach 0.9773. Then, in the second stage, the proposed CGAN used the predicted L channel to produce the AB channel. The TIR colorization result achieved a *PSNR* exceeding 26 dB. The experimental results showed that our proposed approach is better than the other methods. Moreover, the TIR auto-coloring technology can be applied to many different fields. This TIR image colorization can also help increase object detection accuracy for safe driving assistance in a poorly lit environment. Overall, the method of colorizing TIR images proposed in this paper is useful and has many potential applications.

#### Acknowledgments

This work was supported by the National Science and Technology Council under Grant NSTC 111-2222-E-167-003.

#### References

- 1 N. Pinchon, O. Cassagnol, A. Nicolas, F. Bernardin, P. Leduc, J.-P. Tarel, R. Brémond, E. Bercier, and J. Brunet: Proc. 22nd Int. Forum on Advanced Microsystems for Automotive Applications (AMAA, Berlin, 2018) 3–15. [https://doi.org/10.1007/978-3-319-99762-9\\_1](https://doi.org/10.1007/978-3-319-99762-9_1)
- 2 L. Yatziv and G. Sapiro: IEEE Trans. Image Processing **15** (IEEE, 2006) 1120. <https://doi.org/doi:10.1109/TIP.2005.864231>
- 3 A. Bugeau, V.-T. Ta, and N. Papadakis: IEEE Trans. Image Processing **23** (IEEE, 2014) 298. <https://doi.org/10.1109/TIP.2013.2288929>
- 4 B. Zhang, M. He, J. Liao, P. V. Sander, L. Yuan, A. Bermak, and D. Chen: Proc. 2019 IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE, CA, 2019) 8044–8053. <https://doi.org/10.1109/CVPR.2019.00824>
- 5 G. Larsson, M. Maire, and G. Shakhnarovich: Proc. Computer Vision – ECCV 2016 (Springer, Cham, 2016) 577–593. [https://doi.org/10.1007/978-3-319-46493-0\\_35](https://doi.org/10.1007/978-3-319-46493-0_35)
- 6 R. Zhang, P. Isola, and A. A. Efros: Proc. Computer Vision – ECCV 2016 (Springer, Cham, 2016) 649–666. [https://doi.org/10.1007/978-3-319-46487-9\\_40](https://doi.org/10.1007/978-3-319-46487-9_40)
- 7 I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: Proc. 28th Conf. Neural Information Processing Systems (NIPS, Montreal, 2014) 2672–2680. <https://doi.org/10.48550/arXiv.1406.2661>
- 8 F. Li, L. Ma, and J. Cai: Proc. 2018 IEEE Int. Geoscience and Remote Sensing Symposium (IEEE, Valencia, 2018) 3489–3492. <https://doi.org/10.1109/IGARSS.2018.8517930>
- 9 M. Fang, Y. Song, X. Bai, Y. Ren, and S. Liu: Proc. 2019 Chinese Control and Decision Conf. (CCDC, Nanchang, 2019) 5944–5949. <https://doi.org/10.1109/CCDC.2019.8833380>
- 10 S. Yan, Y. Liu, J. Li, and H. Xiao: Proc. 2020 6th Int. Conf. Big Data and Information Analytics (BigDIA, Shenzhen, 2020) 214–219. <https://doi.org/10.1109/BigDIA51454.2020.00042>
- 11 K. Du, C. Liu, L. Cao, Y. Guo, F. Zhang, and T. Wang: Proc. IEEE Access (IEEE, 2021) 21604–21617. <https://doi.org/10.1109/ACCESS.2021.3055575>



- 12 A. K, R. R. Pasham, and S. MD: Proc. 2022 IEEE Int. Conf. Distributed Computing and Electrical Circuits and Electronics (ICDCECE, Ballari, 2022) 1–5. <https://doi.org/10.1109/ICDCECE53908.2022.9792966>
- 13 J. Huang, J. Liao, and S. Kwong: IEEE Trans. Multimedia **24** (IEEE, 2022) 1435. <https://doi.org/10.1109/TMM.2021.3065230>
- 14 A. Berg, J. Ahlberg, and M. Felsberg: Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW, Salt Lake City, 2018) 1224–122409. <https://doi.org/10.1109/CVPRW.2018.00159>
- 15 D. Tao, J. Shi, and F. Cheng: Proc. 2020 IEEE Int. Conf. Information Technology, Big Data and Artificial Intelligence (ICIBA, Chongqing, 2020) 1184–1190. <https://doi.org/10.1109/ICIBA50161.2020.9277116>
- 16 X. Kuang, J. Zhu, X. Sui, Y. Liu, C. Liu, Q. Chen, and G. Gu: Infrared Phys. Technol. **107** (2020) 103338. <https://doi.org/10.1016/j.infrared.2020.103338>
- 17 T. Sun, C. Jung, Q. Fu, and Q. Han: Proc. IEEE Access (IEEE, 2019) 112459–112469. <https://doi.org/10.1109/ACCESS.2019.2933671>
- 18 T. Le-Tien, T. H. Duy Quang, H. Y. Vy, T. Nguyen-Thanh, and H. Phan-Xuan: Proc. 2021 Int. Symp. Electrical and Electronics Engineering (ISEE, Ho Chi Minh, 2021) 90–94. <https://doi.org/10.1109/ISEE51682.2021.9418801>
- 19 S. Zhou and S.-I. Kamata: Proc. 2022 7th Int. Conf. Image, Vision, and Computing (ICIVC, Xi'an, 2022) 507–512. <https://doi.org/10.1109/ICIVC55077.2022.9887040>
- 20 S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon: Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, Boston, 2015) 1037–1045. <https://doi.org/10.1109/CVPR.2015.7298706>
- 21 A. Ciobanu, M. Luca, and V. Drug: Proc. 2020 Int. Conf. e-Health and Bioengineering (EHB, Iasi, 2020) 1–4. <https://doi.org/10.1109/EHB50910.2020.9280110>
- 22 X. Sun, A. Gossmann, Y. Wang, and B. Bischt: Proc. 2019 IEEE Symp. Ser. Computational Intelligence (SSCI, Xiamen, 2019) 1344–1353. <https://doi.org/10.1109/SSCI44817.2019.9002665>
- 23 J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros: Proc. 2017 IEEE Int. Conf. Computer Vision (ICCV, Venice, 2017) 2242–2251. <https://doi.org/10.1109/ICCV.2017.244>
- 24 FLIR thermal dataset: <https://www.flir.com/oem/adas/adas-dataset-form/> (accessed June 2022).
- 25 AAU RainSnow Traffic Surveillance Dataset: <https://www.kaggle.com/datasets/aalborguniversity/aau-rainsnow?resource=download> (accessed October 2022).

## About the Authors



**Ming-Tsung Yeh** received his M.S. and Ph.D. degrees from National Changhua University of Education, Taiwan, in 2012 and 2016, respectively. Since 2022, he has been an assistant professor at National Chin-Yi University of Technology, Taiwan. His research interests are in AI, deep learning, image processing, and intelligent control. ([mttyeh@ncut.edu.tw](mailto:mttyeh@ncut.edu.tw))



**Wei-Yin Lo** received his B.A. degree in electrical engineering from National Changhua University of Education, Taiwan, in 2021. He is currently working toward his M.S. degree in electrical engineering at National Changhua University of Education, Changhua, Taiwan. His research interests are in machine learning and image processing. ([z2815414@gmail.com](mailto:z2815414@gmail.com))



**Yi-Nung Chung** received his Ph.D. degree from Texas Tech University, Lubbock, TX, USA, in 1990. He is currently a professor at National Changhua University of Education, Changhua, Taiwan. He is also the Dean of Academic Affairs. His research interests include image processing and computer vision. ([yunchung@cc.ncue.edu.tw](mailto:yunchung@cc.ncue.edu.tw))



**Pei-Syuan Lu** received her M.S. degree in electrical engineering from National Changhua University of Education, Taiwan, in 2022. She is currently working at Delta Electronics Inc. Taiwan. Her research interests are in machine learning, image processing, and power electronics.

([angellups1999@gmail.com](mailto:angellups1999@gmail.com))