Data Augmentation with Pseudo-Infrared Night-Vision Image Conversion for Improved Nighttime Object Detection

1st Ryoma Ishizu
National Insutitute of Technology, Oshima College
Yamaguchi, Japan
d22002@oshima-k.ac.jp

2nd Hironori Kitakaze
National Insutitute of Technology, Oshima College
Yamaguchi, Japan
kitakaze@oshima-k.ac.jp

3rd Ryo Matsumura
Shunan University
Yamaguchi, Japan
matsumura@shunan-u.ac.jp

***Abstract*—This paper presents a novel data augmentation method with pseudo-infrared night-vision image conversion for RGB images captured under daylight conditions. Our proposed data augmentation method is to apply gamma correction after grayscale conversion to RGB images and composites them with a mask image that simulates infrared irradiation. Our goal is to improve the accuracy of nighttime object detection using infrared night-vision cameras. We exclusively utilized data from the "person" class in the PASCAL VOC dataset and conducted a nighttime detection performance comparison between models trained on this person dataset (original dataset) and models trained on the person dataset to which we applied our proposed data augmentation method (converted dataset). The experimental results showed that the Average Precision (AP) for the model trained on the original dataset is 74.9%. In contrast, the AP for the model trained on our converted dataset is 78.7%.**

***Index Terms*—data augmentation, infrared night-vision image, nighttime object detection, mobilenet-ssd.**

I. Introduction

Surveillance cameras play a pivotal role in ensuring the safety and security of individuals and are ubiquitously deployed in diverse settings, encompassing commercial establishments, public infrastructures, and urban thoroughfares. The imagery captured by these surveillance cameras serves as a valuable resource for detecting suspicious individuals and locating missing children. Traditionally, detecting objects of interest from surveillance camera images has relied on manual operations. However, manually detecting objects from images obtained from many cameras can be extremely challenging. Therefore, the current promotion of introducing object detection technology such as Convolutional Neural Networks (CNNs) and other methods aims to reduce the workload.

Surveillance cameras frequently employ infrared night-vision cameras with infrared illumination using infrared LEDs to facilitate nighttime surveillance. Images captured during nighttime by such cameras exhibit differences in comparison to RGB camera images captured during the daytime, as they are represented in grayscale images with diminished luminance. Given that most of object detection datasets consist of RGB images captured under daylight conditions, CNN models trained on such datasets decrease object detection accuracy when applied to images captured under nighttime by infrared night-vision cameras. Therefore, the creation of a dataset for nighttime object detection using infrared night-vision cameras is required, however, the collection of infrared night-vision images requires significant labor-intensive efforts.

In this paper, we propose a novel data augmentation method with pseudo-infrared night-vision image conversion for nighttime object detection using infrared night-vision cameras. Our proposed method is a pseudo-conversion of RGB images captured under daylight conditions into infrared night-vision images captured at nighttime. The study goal is to improve nighttime object detection using training dataset converted by the proposed method. We show the effectiveness of the proposed method through detection performance evaluation experiments using the model trained on the PASCAL VOC dataset and the model trained on the converted PASCAL VOC dataset using the proposed method.

Data augmentation methods can be classified into two categories: basic image manipulation and deep learning-based methods [1]. On the other hand, categorizing them in terms of approach, there are image processing-based and image generation-based approaches. A typical example of an image generation-based approach to data augmentation is generative adversarial networks (GANs). An example of a GANs architecture designed for image style transfer tasks is pix2pix and cycle GAN. Zhang et al. [2] proposed the use of the pix2pix and cycle GAN to convert RGB images into thermal infrared (TIR) images, thereby constructing the training dataset for TIR images. Both of these image generation-based approaches that utilize GANs necessitate the availability of training images, which is not in line with our concept of constructing a dataset with minimal labor-intensive costs.

Furthermore, the use of TIR cameras has been proposed for nighttime object detection [2]–[4], and in recent years, TIR image datasets for object detection have been published [5], [6]. However, our target is object detection using infrared night-vision cameras. Infrared night-vision cameras use near-infrared illumination with a wavelength range of 850 nm to

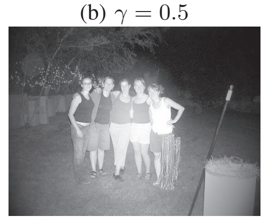
2832-1324/23/$31.00 ©2023 IEEE
DOI 10.1109/CANDARW60564.2023.00069

Fig. 1: Mask image.

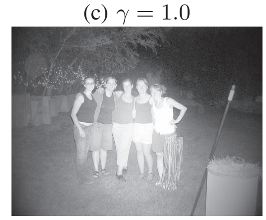
(a) original

(b) $\gamma = 0.5$

(c) $\gamma = 1.0$



(d) $\gamma = 1.5$



(e) $\gamma = 2.0$

Fig. 2: Pseudo-infrared night-vision images.

(a) indoor

(b) outdoor

Fig. 3: Test images.

940 nm, which is imaging by a normal RGB image sensor (without infrared cut filters). Due to the differing imaging characteristics between TIR cameras and infrared night-vision cameras, the above datasets do not apply to our use case.

II. Pseudo-infrared night-vision image conversion

This section describes our proposed data augmentation method, pseudo-infrared night-vision images. Our method entails the application of gamma correction subsequent to converting RGB images captured in daylight conditions to grayscale images. After that, these images are overlay-composited with a mask image, simulating infrared irradiation. The grayscale conversion is performed using the weighted average method. After grayscale conversion, gamma correction is performed as shown in eq.1. Here, $I(x, y)$ denotes the input pixel value and I_{max} denotes the maximum pixel value. This correction helps to fine-tune the contrast of the image and reduce the overall brightness.

$$I'(x, y) = I_{max} \left(\frac{I(x, y)}{I_{max}} \right)^{\frac{1}{\gamma}} \tag{1}$$

In the final step of our method, the image after grayscale conversion and gamma correction and the mask image are overlay-composited. The overlay composite is represented by eq.2. In the eq.2, $I(x, y)$ denotes the pixel value of the image after grayscale conversion and gamma correction, $Mask(x, y)$ denotes the pixel value of the mask image, and $dst(x, y)$ denotes the pixel value of the output image. Overlay composition switches between multiply and screen composition, as shown in eq.2, with the middle of the pixel value as the border. That produces a contrast-enhancing effect and allows the light/dark distribution of the mask image simulating infrared irradiation to be reflected in the composite original image. By the above process, we simulate an infrared night-vision image.

$$dst(x, y) = \begin{cases} \frac{2I(x,y)Mask(x,y)}{255} & (Mask(x, y) < 128) \\ 255 - \frac{2(255-I(x,y))(255-Mask(x,y))}{255} & (othrewise) \end{cases} \tag{2}$$

Fig.1 shows a mask image. Many of the images in the PASCAL VOC dataset are images captured from the side of the object, i.e., the optical axis of the camera and the ground are parallel. Therefore, we created a mask image assuming that the infrared is emitted horizontally. Infrared night-vision cameras are also equipped with multiple infrared LEDs arranged in a circular pattern around the lens or an array of infrared LEDs. We assumed it is a point light source and simulated the infrared irradiation in the mask image as diffuse light from this light source.

III. Experimental overview

In the experiment, to evaluate the effectiveness of the proposed data augmentation method, we compared the detection performance of models trained on the PASCAL VOC dataset with models trained on the dataset constructed by the proposed method. For this evaluation, we exclusively utilized images from the "person" class in that dataset, and we adjusted the division ratio between training and validation data to 9:1. PASCAL VOC person dataset is the "original dataset", and the proposed method applied to that dataset is the "converted dataset". For detailed verification of the proposed method, the converted datasets have four variations that each consist of gamma-corrected images with four parameters: $\gamma = 0.5, 1.0, 1.5,$ and 2.0. Fig.2(a) shows a example of

TABLE I: Detection performance (IoU≥ 0.5).

Trained dataset	Trained model's AP[%]
original dataset	74.89
converted dataset with $\gamma = 0.5$	70.58
converted dataset with $\gamma = 1.0$	78.68
converted dataset with $\gamma = 1.5$	71.15
converted dataset with $\gamma = 2.0$	72.55

PASCAL VOC person image. In addition, (b) to (e) are images obtained by pseudo-infrared night-vision image conversion for (a) using the proposed method. We utilized real infrared night-vision images of persons captured both outdoors and indoors for the test dataset which consists of 211 images.

In recent years, surveillance systems commonly employ edge AI-equipped cameras to solve the cost of data communication and privacy risks. Therefore, we use mobileNet-SSD [7] in this experiment. Concerning the training setup, the batch size is 64, and the optimization method adopted is momentum SGD with a momentum value of 0.9. Moreover, the learning rate is 0.01. Additionally, mobileNet-SSD training via transfer learning, utilizing a pre-trained model from the ImageNet.

IV. Result and discussion

The evaluation of the detection performance of the trained model is carried out by using Average Precision (AP) as the evaluation metric. Table I shows the APs of the models trained on each dataset. Among these trained models, the model trained on the converted dataset with $\gamma = 1.0$ showed an improvement of 3.79% in AP compared to the model trained on the original dataset. These results show that the model trained on the converted dataset that utilizes the proposed method can improve nighttime object detection, and we confirmed the effectiveness of our proposed data augmentation method.

We assumed that the light intensity and light distribution characteristics of the infrared illumination in an infrared night-vision camera would require gamma correction to adjust the luminance of the images. However, experimental results showed that gamma correction did not affect detection performance. We randomly selected 100 images each from the original, converted, and tested datasets for validation of the effect of gamma correction. PCA is performed on them to calculate the similarity between images in each dataset and real infrared night-vision images (test dataset images). The similarity is measured by calculating the median point of each dataset on the space with the first to third principal components as axes and using it to calculate the distance to the test dataset.

The distances of the converted datasets and test dataset are as follows: $74.03, 136.11, 177.68$, and 205.98 for the converted datasets with $\gamma = 0.5, 1.0, 1.5$, and 2.0, respectively. In addition, the distance between the original dataset and the test dataset is 167.09. It shows that the features of converted datasets images with $\gamma = 1.5$ and 2.0 are less similar to the features of the test dataset images than the original dataset images. Furthermore, owing to the greater distance between these datasets and the test dataset in comparison to the distance between the converted dataset with $\gamma = 1.0$ and the test dataset, the performance of the models trained on the converted datasets with $\gamma = 1.5$ and $\gamma = 2.0$ is inferior to the model trained on that dataset. Thus, gamma correction with $\gamma = 1.5$ and $\gamma = 2.0$ does not contribute to improving the detection performance. In contrast, the distance between the converted dataset with $\gamma = 0.5$ and the test dataset is shorter than the converted dataset with $\gamma = 1.0$. However, the model trained on the converted dataset with $\gamma = 1.0$ has better performance.

We implemented basic data augmentation such as luminance transformation, flipping, shifting, and cropping to further augment the training images when training the models. Applying luminance transformation to an image containing many low and high luminance pixels can lead to the emergence of areas of blocked up shadows and blown out highlights, potentially resulting in the loss of object features in the image. Analyzing the pixels across all images in the converted datasets with $\gamma = 0.5$ and $\gamma = 1.0$, we observed that the low and high luminance pixels susceptible to that transformation are approximately 20% more prevalent in the images in the converted dataset with $\gamma = 0.5$ than in the converted dataset with $\gamma = 1.0$. We have concluded that this factor affected the training process, resulting in the model trained on the converted dataset with $\gamma = 0.5$ showing lower performance compared to the other models, and the performance of that model can be improved by using another augmentation instead of luminance transformation.

V. Conclusion

In this paper, we proposed a novel data augmentation method with pseudo-infrared night-vision image conversion. Our experimental results show that the model trained on the converted dataset that utilizes the proposed method can improve nighttime object detection using infrared night-vision cameras to compare the model trained on the original PASCAL VOC dataset. As a result, we confirmed the effectiveness of our proposed method. Future tasks validate the object detection performance of models trained on the proposed method in various scenes using various models of infrared night-vision cameras. In addition, another task is to extend the object classes to be detected beyond the person class.

References

[1] Shorten, C., Khoshgoftaar, T.M., "A survey on image data augmentation for deep learning," *Journal of big data*, 6.1, pp. 1-48, 2019.

[2] Zhang, L., et al., "Synthetic data generation for end-to-end thermal infrared tracking," *IEEE Transactions on Image Processing*, 28.4, pp. 1837-1850, 2018.

[3] González, Alejandro, et al., "Pedestrian detection at day/night time with visible and FIR cameras: A comparison," *Sensors*, 16.6, 820, 2016.

[4] Manssor, S. A., Sun, S., and Elhassan, M. A., "Real-time human recognition at night via integrated face and gait recognition technologies," *Sensors*, 21.13, 4323, 2021.

[5] Teledyne FLIR, "FREE Teledyne FLIR Thermal Dataset for Algorithm Training," Retrieved September 4, 2023, https://www.flir.com/oem/adas/adas-dataset-form/

[6] Roboflow, "Top Infrared Datasets," Retrieved September 4, 2023, https://universe.roboflow.com/browse/infrared

[7] Howard, A. G., et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.