IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 15, 2020, accepted February 27, 2020, date of publication March 9, 2020, date of current version March 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2979239

Nighttime Data Augmentation Using GAN for Improving Blind-Spot Detection

HONGJUN LEE[1], MOONSOO RA[2], AND WHOI-YUL KIM[1]
[1]Department of Electronics and Computer Engineering, Hanyang University, Seoul 04763, South Korea
[2]LightVision Inc., Seoul 04793, South Korea

Corresponding author: Whoi-Yul Kim (wykim@hanyang.ac.kr)

This work was supported in part by the Korea Evaluation Institute of Industrial Technology (KEIT) funded by the Korea Government (MOTIE) (Road Surface Condition Detection using Environmental and In-vehicle Sensors) under Grant 20000293, and in part by the Human Resources Program in Energy Technology Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of South Korea under Grant 20184030201970.

ABSTRACT Camera-based blind-spot detection systems improve the shortcomings of radar-based systems for accurately detecting the position of a vehicle. However, as with many camera-based applications, the detection performance is insufficient in a low-illumination environment such as at night. This problem can be solved with augmented nighttime images in the training data but acquiring them and annotating the additional images are cumbersome tasks. Therefore, we propose a framework that converts daytime images into synthetic nighttime images using a generative adversarial network and that augments the synthetic images for the training process of the vehicle detector. A public dataset comprising different viewpoints of target images was used to easily obtain the images required for training the generative adversarial network. Experiments on a real nighttime dataset demonstrate that the proposed framework improved the detection performance considerably in comparison with using daytime images only.

INDEX TERMS Data augmentation, domain adaptation, generative adversarial networks, blind-spot detection.

I. INTRODUCTION

Advanced Driver Assistance Systems (ADASs) using various sensors have shown considerable success in preventing traffic accidents. Many systems have been applied to commercialized vehicles to prevent traffic accidents and save many lives. Among them, a blind-spot detection (BSD) system locates obstacles toward the rear of the host vehicle on the passenger side, the so-called ''blind-spot'', and informs the driver about the existence of any obstacles. BSD systems are widely used in many ADASs because they reduce the risk when change lanes that can sometimes lead to traffic accidents. A study analyzing the effectiveness of a BSD system found a 14% reduction in the likelihood of lane change accidents for vehicles equipped one [1].

Most commercially available BSD systems utilize active sensors such as radars and ultrasound [2]–[4]. However, active sensor-based systems have the following disadvantages: false alarms when passing through guardrails or tunnels, location inaccuracy due to the low lateral accuracy of radar, and difficulty in detecting small obstacles; camera-based systems have been studied to solve these shortcomings. A recent study was conducted by Ra *et al.* [5] in which they introduced the *side-rectilinear image* to detect objects in the blind-spot of a vehicle. As shown in Fig. 1, a side-rectilinear image is captured by a virtual camera whose optical axis is perpendicular to the rear fisheye camera. A side-rectilinear image has the advantage that the side parts of the vehicle appear as the same size in the image regardless of the position of the vehicle. Therefore, vehicle location in the BSD region is detected with high accuracy based on the hand-crafted features. Although the most recent ADASs use neural nets to detect objects, it is not always possible to install a graphics processing unit (GPU) in a vehicle to run the network in real-time. Thus, accurately detecting vehicles in the BSD region based on hand-crafted features rather than neural nets is significant.

Camera-based ADAS technology has a crucial problem in that its detection performance is severely degraded in low-light environments (e.g., at night or in a tunnel),

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

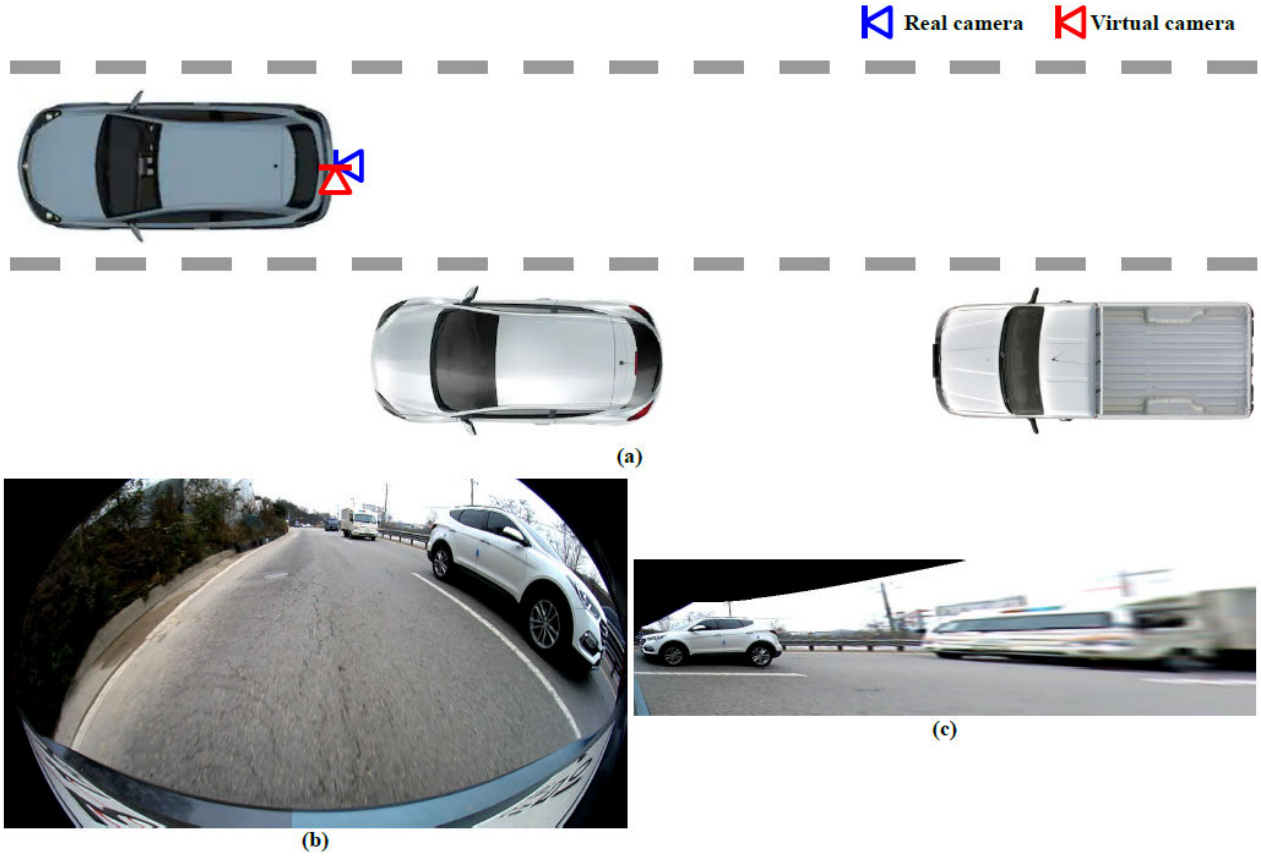

FIGURE 1. The side-rectilinear image concept introduced in [5]. (a) A top-view schematic of a situation with the host vehicle and two other vehicles, (b) the image captured by the rear fisheye camera, and (c) the side-rectilinear image captured by the virtual camera.

a problem that also afflicts camera-based BSD systems. For example, the manual for Volvo's camera-based BSD system notes possible malfunctioning in low light [6]. The performance of the camera-based system using side-rectilinear images in [5] is limited in low-illumination environments, and thus our focus in this study was on improving its performance under such conditions.

Most camera-based ADAS technologies utilize learning-based perception algorithms. The performance degradation of the perception algorithm in low-light environments is owing to the lack of nighttime image samples in the training dataset. Datasets used in many recent studies only contain images taken during the daytime, and acquiring the same number of nighttime images as can be obtained in the daytime is a labor-intensive task. Moreover, not only image acquisition but also ground truth annotation is necessary to train the perception algorithm in a supervised manner. To solve this image acquisition and annotation problem, researchers have recently attempted to use generative adversarial networks (GANs) as the core of the solution. GANs are deep learning structures that generate a synthetic image close to the real image. The structure consists of a network for generating the image and a network for determining the authenticity of the image. In particular, using a conditional GAN makes it possible to convert an image in one domain to one in another. Many researchers have improved the performance in a specific domain by adding the GAN-generated synthetic images to the training dataset [7]–[9].

Typically, to improve the performance through conditional GAN, large numbers of images in both the input and output domains (in this case, daytime and nighttime side-rectilinear images) are required to train the GAN. Additional image acquisition should be eliminated if possible since our objective is to reduce the burden of acquiring and annotating images. Therefore, we utilized publicly available datasets with different viewpoints to train the conditional GAN. Several front-view databases are available for public access to develop various ADAS algorithms (e.g., vehicle detection, pedestrian detection, semantic segmentation, and so on) [10], [11]. Furthermore, some databases, including all-day data, have been released [12], [13]. Since the mapping between input and output domains (i.e., the style change between day and night images in this study) is similar (even in images from different viewpoints), the labor-intensive tasks of acquiring images for training the GAN can be eliminated.

In this paper, we propose an any-time-of day camera-based BSD system with GAN-based synthetic augmentation applied to an existing BSD system. To improve the

nighttime performance of the existing hand-crafted feature-based BSD system, synthetic nighttime side-rectilinear images were generated by using a conditional GAN trained on a large and public front-view dataset. Even though training and testing images have different viewpoints (front for training and side-rectilinear for testing), we improved the nighttime performance of the BSD system with augmented synthetically generated nighttime images.

In summary, the main contributions of this paper are:

- We propose a new framework with a conditional GAN for data augmentation. The framework improves the performance of the perception algorithm in ADAS without any additional data acquisition.
- We use the publicly available front-view databases that have a different viewpoint from the images required for training the target perception algorithm to eliminate the burden of acquiring data to train the GAN.
- The proposed framework drastically improves the nighttime detection performance of the existing BSD system based on the hand-crafted features.

II. RELATED WORK

A. CAMERA-BASED BSD

Camera-based BSD systems can be grouped into two types depending on the position of the camera. Side-camera-based systems detect vehicles in the blind-spot region by using cameras installed at the bottom of the side mirrors. Tseng *et al.* [14] calculated the optical flow and detected vehicles based on clustering, while Singh *et al.* [15] proposed appearance-based methods that detect vehicles by appearance features like the car body, tire, and shadow of the vehicle. Side-camera-based systems have the advantage that vehicles are captured at high resolution in the image because of the dedicated cameras for blind-spot systems. However, there is a disadvantage in that the appearance of the detected vehicle is distorted depending on the distance between it and the host vehicle. Chang *et al.* [16] and Wu *et al.* [17] proposed side-camera-based systems for detecting vehicles at night. Both systems detect vehicles accurately via information on the headlights. However, it is difficult to use different appearance features in the daytime and nighttime.

Rear-camera-based systems use a rear fisheye camera, that causes large radial distortion of the images. To solve this, Tsuchiya *et al.* [18] generated a birds-eye view of the rear fisheye camera image and detected vehicles using Census transform. Dooley *et al.* [19] divided images into three regions and applied appropriate detection methods for each region. Ra *et al.* [5] simplified existing frameworks by using side-rectilinear images. Cheng and Chen [20] and Kim *et al.* [21] proposed rear-camera-based systems for use at night that are similar to the side-camera-based systems using the headlights of vehicles for appearance features at night.

As previously stated, many BSD systems either address only the daytime environment or add a separate detection algorithm for nighttime use to reduce degradation. In this study, we add synthetic nighttime data to the framework of Ra *et al.* [5] to solve the problem of radial distortion in rear-camera-based systems.

B. GANS FOR IMAGE-TO-IMAGE TRANSLATION

GANs [22] have competitively trained generative and discriminative models and have shown good results in many applications [23]–[27]. Among them, one of the most studied applications is image-to-image translation. This refers to tasks that translate the representation of an image in one scene into another in a way similar to language translation. GAN-based image-to-image translation methods can be grouped into two categories: those requiring supervision with paired images during training and those that do not.

A well-known method that falls into the first category is pix2pix [28], which is based on a conditional GAN [29] that utilizes paired images as the supervision to solve image-to-image translation. This approach has shown impressive results and has been applied to various tasks [30]–[32]. Furthermore, Wang *et al.* [31] proposed pix2pixHD, which can generate high-resolution synthetic images. However, the paired images are almost impossible to obtain. For example, if the input is an image acquired from the front camera of the vehicle and the output is a semantically labeled image, all of the image samples are naturally paired. On the other hand, in the case of daytime to nighttime translation, acquiring diverse samples is difficult because it is challenging to set up a camera to take images in the same place throughout the whole day.

To solve the difficulty of acquiring paired samples, several methods in the second category tackle training without such supervision. CycleGAN [34] is a representative work that adds cycle loss to pix2pix. Even though it is easy to acquire the training database with CycleGAN, the results are largely affected by the distribution characteristics of the training database and geometric changes between the input and output domains. In other words, successful results can be obtained if color and texture changes are the only difference between the input and output domains. Liu *et al.* [35] proposed a similar approach to CycleGAN which is robust to these shortcomings, while similar studies have dealt with various tasks in an unpaired setting [36]–[38].

The image-to-image translation problem tackled in this study is to convert daytime images to nighttime ones. Therefore, it is more reasonable to use unpaired image-to-image translation methods. In addition, various daytime and nighttime unpaired images can be acquired from several public databases. Therefore, the second category that does not require paired image samples is a good option to obtain the nighttime images for data augmentation. In this work, we adopted CycleGAN to translate daytime images to nighttime ones.

For autonomous driving, the results of in-vehicle camera images have been mentioned in existing GAN-based image-to-image translation researches [34], [35], [39]. The article by Zhu *et al.* [34] includes the results of translation from semantic labels to front-view images using the Cityscapes

dataset. Moreover, the example results of day-to-night translation trained on the BDD100K dataset [12] are presented on the project page of CycleGAN [39]. This dataset includes 100K of diverse front-view video clips collected from various cities and environments. Liu *et al.* [35] presented qualitative results of the street scene image translation tasks (e.g., from sunny conditions in summer to snowy conditions in winter, from daytime to nighttime). Driving videos recorded on different days and in different cities were used to obtain these results. Although the data configuration was not presented exactly, they assumed that the training and test data were front-view images. Moreover, the results in previous studies were obtained using the same viewpoint images in the training and test data for image-to-image translation.

The perception algorithms used in autonomous driving are based on various viewpoint images as well as front-view ones [5], [40]–[42]. In previous image-to-image translation researches, acquiring the target viewpoint images is required to improve the performance in the other domain. For this reason, we show that domain adaptation is possible via a GAN trained with heterogeneous viewpoint datasets. In this work, side-rectilinear images are input into the generator, but the training data for the GAN were obtained from a publicly available front-view image dataset.

III. METHODOLOGY

A. OVERVIEW

In this study, we augment the synthetic nighttime images over the entire sequence of steps in the side-rectilinear image-based BSD system, as illustrated in Figure 2.

In the first phase, CycleGAN was trained with daytime and nighttime images from the public front-view database. Synthetic nighttime side-rectilinear images were generated by using daytime side-rectilinear images with the trained CycleGAN generator. The annotation data of the daytime images can be used for the generated nighttime images. In the second phase, the vehicle detector, which is robust in low-light environments, was trained at all vehicle detection stages using both the daytime and synthetic nighttime images.

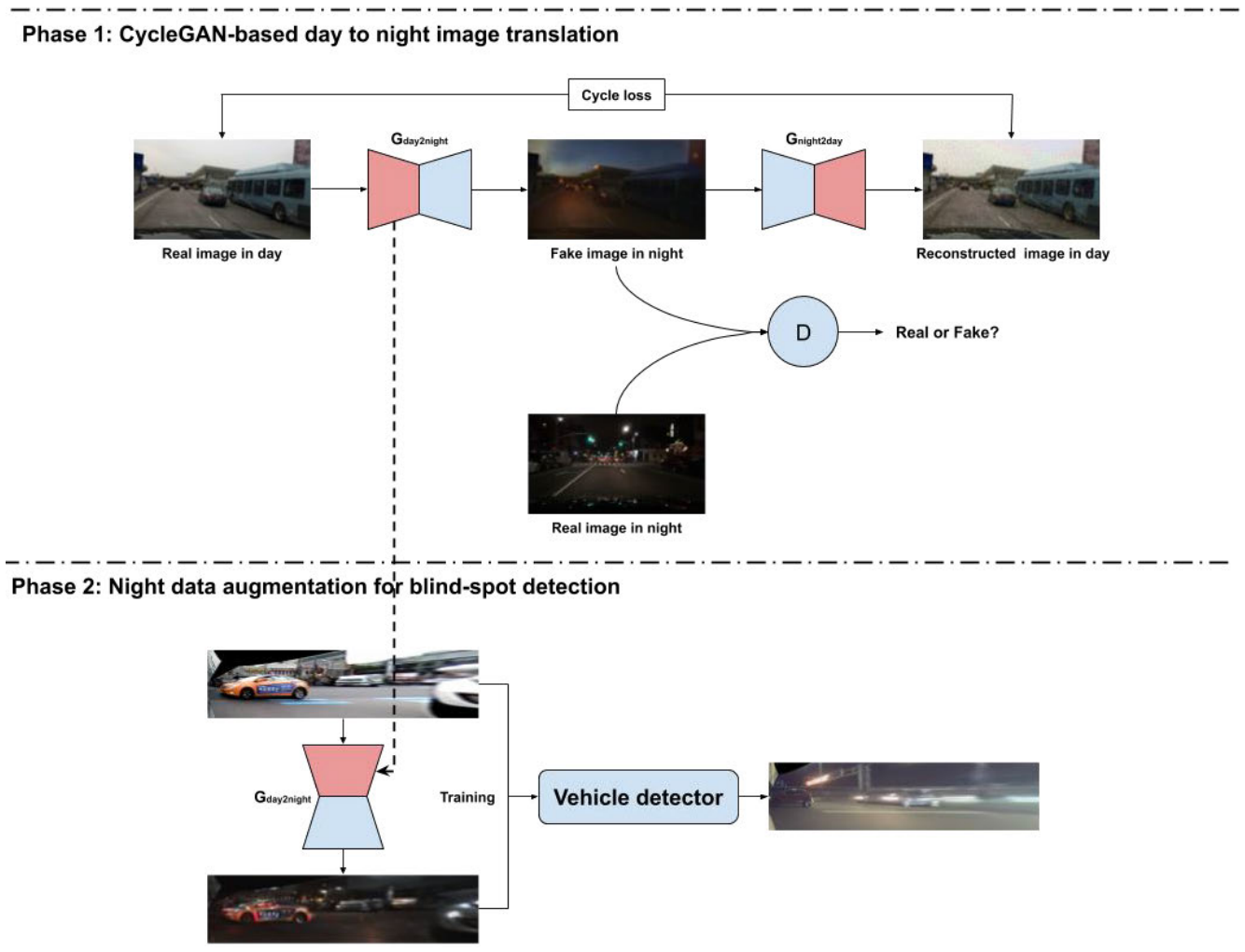

FIGURE 2. A flowchart of the proposed data augmentation framework.

B. CYCLEGAN-BASED DAY-TO-NIGHTTIME IMAGE TRANSLATION

The training process for the two networks of CycleGAN is shown in Phase 1 in Fig. 2. The aim of the training was to find mapping function $G : X_f \rightarrow Y_f, F : Y_f \rightarrow X_f$ between the daytime (X_f) and nighttime (Y_f) domains, where f denotes the front-view images. Discriminator D and generator G are then trained competitively. Our objective function consists of two losses: adversarial and cycle consistency. The adversarial loss is expressed as:

$$\begin{aligned} &L_{GAN}\left(\mathrm{G}, D_{Y_f}, X_f, Y_f\right) \\ &\quad = \mathbb{E}_{y_f \sim p_{data}(y_f)}\left[log D_{Y_f}\left(y_f\right)\right] \\ &\qquad + \mathbb{E}_{x_f \sim p_{data}(x_f)}\left[\log\left(1 - D_{Y_f}\left(G\left(x_f\right)\right)\right)\right], \end{aligned} \tag{1}$$

and the cycle consistency loss is expressed as

$$\begin{aligned} L_{cyc}\left(\mathrm{G}, \mathrm{F}\right) = &\mathbb{E}_{x_f \sim p_{data}(x_f)}\left[\left\|F\left(G\left(x_f\right)\right) - x_f\right\|_1\right] \\ &+ \mathbb{E}_{y_f \sim p_{data}(y_f)}\left[\left\|G\left(F\left(y_f\right)\right) - y_f\right\|_1\right]. \end{aligned} \tag{2}$$

Hence, the objective function can be written as

$$\begin{aligned} \mathrm{L}\left(\mathrm{G}, \mathrm{F}, D_{X_f}, D_{Y_f}\right) = &L_{GAN}\left(G, D_{Y_f}, X_f, Y_f\right) \\ &+ L_{GAN}\left(F, D_{X_f}, Y_f, X_f\right) + L_{cyc}\left(G, F\right). \end{aligned} \tag{3}$$

The daytime side-rectilinear images were fed to the trained generator to generate synthetic nighttime images augmented by the vehicle detector, as shown in Phase 2 in Fig. 2. Even though the images had been acquired from cameras installed at different locations on the vehicle, the style changes were similar to that of the daytime and nighttime images consisting of roads, vehicles, and background. In particular, as is mentioned in Section C, since only the regions in the images showing tires and vehicles are used in the training samples for the vehicle detector, they should be similar in the synthetic and actual nighttime images.

C. NIGHT DATA AUGMENTATION FOR BSD

In this section, we describe the process of data augmentation using synthetic nighttime images to improve the nighttime performance of the existing BSD system [5]. To this end, let us first describe the BSD system briefly. The BSD system detected vehicles in three stages: tire hypothesis generation and verification; front and rear tire classification; followed by vehicle hypothesis generation and verification, as described in Fig. 3 in the form of a flowchart. In the system, tire hypotheses were acquired using the Viola-Jones object detector [43] on the side-rectilinear images. Then, in the hypothesis verification stage, Histogram of Oriented Gradient (HOG) [44] features were extracted from the generated hypotheses, and the hypotheses were classified into tires and non-tires using a Support Vector Machine (SVM) [45]. Finally, the classification stage, the front and rear tires were distinguished using HOG for the features and SVM as the classifier.

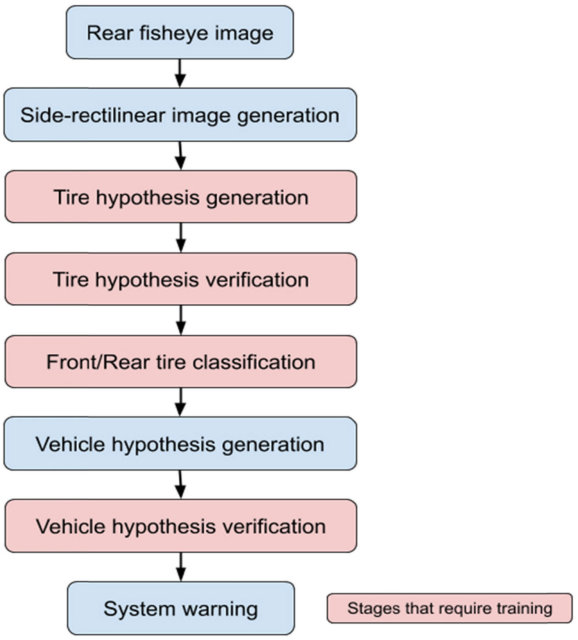

FIGURE 3. Flowchart of the existing camera-based BSD system.

Subsequently, vehicle hypotheses were derived as combinations of front and rear tires. During this process, only geometric constraints (e.g., the minimum distance between the front and rear tires, the front tire must be located left of the rear tire, etc.) are utilized for hypotheses generation, and so no training process is required at this stage. After that, HOG features are extracted and SVM is utilized for the vehicle hypothesis in a similar way to the tire verification process.

The proposed method augments the synthetic nighttime images by doubling the training samples used in the four training stages. Fig. 4 shows a flowchart of the data augmentation process. In the tire hypothesis generation stage, only the tire region of the synthetic nighttime image was cropped using the annotation information of the daytime image and used as a training sample. In the tire hypothesis verification and tire classification stages, the region of the tire was expanded to include the car body. Finally, in the vehicle hypothesis verification stage, the vehicle training samples were cropped based on the annotation of the front and rear tire positions and pair information. The daytime and synthetic nighttime training samples used in each stage are shown in Fig. 4.

IV. EXPERIMENTS AND EVALUATION

Three different datasets were used to evaluate the performance of the proposed framework. Table 1 shows the information on three datasets. CycleGAN was trained and evaluated for day-to-night image translation with the publicly available Nexar dataset comprising 50,000 front-view images from the Nexar Challenge 2 [13]. This dataset includes both daytime and nighttime images from various cities. We used

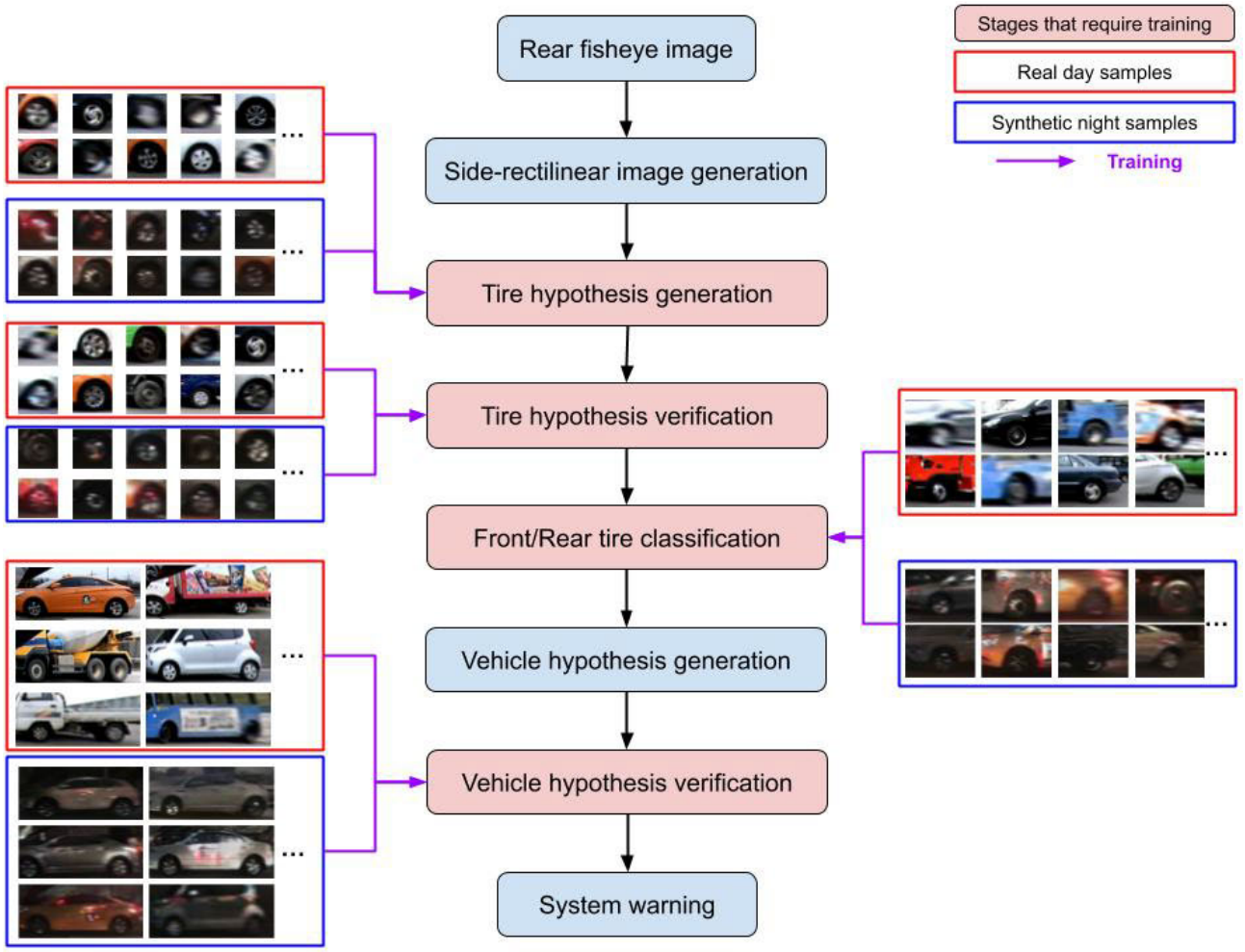

FIGURE 4. Synthetic nighttime sample augmentation at each training stage.

TABLE 1. Information on the three datasets used to evaluate the proposed framework.

	Dataset Name	Resolution	Viewpoint	Time of Acquisition	Number of Images
CycleGAN training data	Nexar dataset	1280 × 720	Front	Daytime/nighttime	4,000
Vehicle detector training data	Genesis dataset	1000 × 250	Side-rectilinear	Daytime	2,000
Vehicle detector test data	Sonata dataset	1000 × 250	Side-rectilinear	Daytime/nighttime	4,000

2,000 daytime and 2,000 nighttime images from the Nexar dataset to train CycleGAN.

Datasets for the vehicle detector were captured using two different vehicles, Hyundai models Genesis and Sonata, for training and testing, respectively. The training dataset was obtained with a rear fisheye camera mounted on the Hyundai Genesis. We collected the 2,000 images and converted them into side-rectilinear images. The resolution of the rear fisheye camera was 1280 × 720 pixels, with a field of view of 180 degrees. The side-rectilinear images had a resolution of 1000 × 250 pixels. The dataset for testing of the vehicle detector was obtained using a rear fisheye camera mounted on the Hyundai Sonata with a resolution and field of view the same as the one on the Genesis. However, the quality of the side-rectilinear images was far different due to the different camera poses. Fig. 5 shows a comparison of the rear fisheye images from the two vehicles. The road area taken up by the Sonata is smaller than that of the Genesis. Since the datasets are heterogeneous, the vehicle detection performance using these datasets was much lower than that obtained in [5]. We conducted experiments to show that heterogeneous datasets are not a problem when demonstrating the

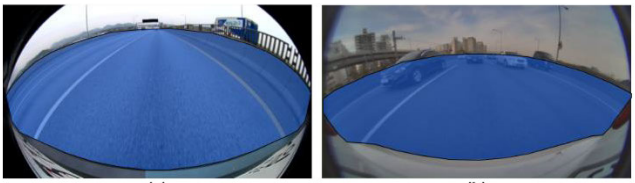

FIGURE 5. Comparison of images obtained from the fisheye cameras on the two vehicles.

performance improvement due to the data augmentation. From the results reported in Table 2, it can be seen that the recall rate was greatly reduced when the vehicle detector was trained with the daytime image dataset from one vehicle type and tested using the datasets from the other one. However, there was no problem in confirming the performance improvement when using nighttime data augmentation.

TABLE 2. Comparison of vehicle detection performance by vehicle type.

Training Dataset	Test Dataset	Precision	Recall
Genesis (Daytime)	Genesis (Daytime)	97.0%	89.0%
Genesis (Daytime)	Sonata (Daytime)	99.2%	56.9%

A. QUALITATIVE EVALUATION OF DAY-TO-NIGHT IMAGE TRANSLATION

To verify whether the CycleGAN generator trained with the front-view images could translate the daytime side-rectilinear images to the nighttime images, we compared the quality of the generated nighttime images by changing the training data for CycleGAN. Three training datasets were used: 2,000 daytime and nighttime image pairs, 500 pairs, and selected 500 pairs. Several images in the Nexar dataset were inappropriate for training because the camera installation angle was improper. Therefore, in the case of the selected 500 pairs, only the images in which the front-view was normally acquired were used and the improper images were excluded. Fig. 6 presents examples of excluded images that were considered to have been acquired incorrectly.

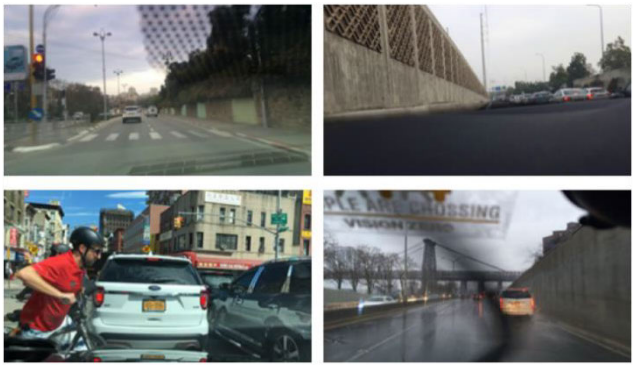

FIGURE 6. Examples of inadequate images in the Nexar dataset.

Training was performed on an NVIDIA GTX 1080Ti with 11 GB memory. The architecture of the generative network used was the neural-style transfer proposed by Cortes and Vapnic [45]. The training parameters of the network were as follows: a learning rate of 0.0002, the weight for the cycle-consistency loss λ was 10, and the number of training epochs was 200. In the comparison of the three cases, each trained generator was used to transform 2,000 side-rectilinear images of the Genesis dataset. Fig. 7 shows the synthetic nighttime images generated by the three generators. In the case of using 500 pairs, the images tended to show a lot of unnatural defects as if the light had been reflected. These defects occurred on the surface of the vehicles or near them. In the case of using the selected 500 pairs, the overall quality of the generated nighttime images was the best. Details of the roads and buildings around the region of the sky were better maintained than with the other two datasets. However, unnatural glittering occurred and there was a lot of noise on the surfaces of the vehicles. When using the 2,000 pairs dataset, there was a tendency to generate smooth images without noise on the surfaces of the vehicles and the surrounding area of the vehicles compared to the other two cases. However, the buildings around the sky and road areas were lower quality than when using the 500 selected pairs.

Our data augmentation approach added training samples by cropping the tires and body of the vehicle from the generated nighttime images. Therefore, when generating the synthetic nighttime images, the quality of the vehicle region

TABLE 3. The number of images used for learning at each stage.

Stage	# of Positives	# of Negatives
Tire hypothesis generation	10,129	3,740
Tire hypothesis verification	19,262	9,226
Front/rear tire classification	9,259	10,003
Vehicle hypothesis verification	3,820	4,225

TABLE 4. Performance at the tire hypothesis generation stage.

Tire Hypothesis Generation	Precision	Recall
Daytime data only	80.6%	82.1%
Daytime + GAN (500 pairs)	80.5%	86.1%
Daytime + GAN (Selected 500 pairs)	74.0%	90.9%
Daytime + GAN (2,000 pairs)	78.6%	92.1%

TABLE 5. Performance at the tire hypothesis verification stage.

Tire Hypothesis Verification	Precision	Recall
Daytime data only	99.7%	45.7%
Daytime + GAN (500 pairs)	87.0%	71.2%
Daytime + GAN (Selected 500 pairs)	87.6%	69.0%
Daytime + GAN (2,000 pairs)	97.6%	72.6%

FIGURE 7. Results of day-to-night image translation after training with, (a) 500 pairs, (b) the selected 500 pairs, and (c) 2,000 pairs.

TABLE 6. Performance at the tire classification stage.

Front/Rear Tire Classification (daytime data only)	Front tire (classification)	Rear tire (classification)	Front/Rear Tire Classification (daytime + GAN (500 pairs))	Front tire (classification)	Rear tire (classification)
Front tire (ground truth)	75.75%	24.25%	Front tire (ground truth)	78.36%	21.64%
Rear tire (ground truth)	0.81%	99.19%	Rear tire (ground truth)	2.92%	97.08%
Front/Rear Tire Classification (daytime + GAN (Selected 500 pairs))	Front tire (classification)	Rear tire (classification)	Front/Rear Tire Classification (daytime + GAN (2,000 pairs))	Front tire (classification)	Rear tire (classification)
Front tire (ground truth)	71.84%	28.16%	Front tire (ground truth)	79.38%	20.62%
Rear tire (ground truth)	1.17%	98.83%	Rear tire (ground truth)	4.22%	95.78%

was better than the sky and road regions. Consequently, using the 2,000 pairs dataset was suitable for training with our data augmentation approach.

B. QUANTITATIVE EVALUATION OF NIGHTTIME DATA AUGMENTATION FOR BSD

The performance of each stage of the vehicle detector whose training dataset is augmented by the generated synthetic images was evaluated. Four experiments were conducted for each stage using only the daytime training samples and using both the daytime and synthetic nighttime training samples. The dataset for performance evaluation was the daytime and nighttime images taken with the fisheye camera on the Hyundai Sonata. Table 3 reports the number of images used for learning in each stage. The test data comprised 2,000 side-rectilinear images, including 1,582 vehicles and 3,656 tires.

The detection performance at each step was measured in the same manner as in [5]. We used the intersection over union (IOU) as the criterion for recall and precision. It was determined that the detection was successful if the IOU was greater than 0.3. Table 4 summarizes the performance of the tire hypothesis generation stage. From the results, the recall was improved by 10% by augmenting the synthetic nighttime samples, and although there was a 2% decrease in precision, it is important to test as many hypotheses as possible in the hypothesis generation stage.

TABLE 7. Performance at the vehicle verification stage.

Vehicle Hypothesis Verification	Precision	Recall
Daytime data only	100.0%	26.2%
Daytime + GAN (500 pairs)	96.6%	51.6%
Daytime + GAN (selected 500 pairs)	96.9%	47.7%
Daytime + GAN (2,000 pairs)	97.1%	55.0%

Next, the performance of the tire hypothesis verification stage is summarized in Table 5. When the samples acquired from the synthetic nighttime images were augmented in the training dataset, the precision of the tire hypothesis verification stage decreased by 2.1%, but the recall improved by 26.9%.

The performance at the tire classification stage was evaluated with a confusion matrix, as reported in Table 6. However, the result did not improve as we thought. When the

FIGURE 8. Results of vehicle detection. Training with (a) daytime samples only, (b) 500 image pairs, (c) selected 500 pairs, and (d) 2,000 image pairs.

classifier was trained with the daytime training samples only, the classification accuracy was the highest and the recall for the tire hypothesis verification stage was 45.7%. On the other hand, when augmenting synthetic nighttime samples, the recall for the tire hypothesis verification stage was 72.6%. In fact, the number of tires used to measure the performance of tire classification was 200 fewer when using only daytime samples. Moreover, the tires detected when training with the daytime samples were relatively clear and the quality was favorable, and so the accuracy was high.

Similar to the tire hypothesis verification, the performance of vehicle hypothesis verification was evaluated by measuring recall and precision. Improving the performance at this stage is the most important because the detection results are used to give the BSD warning signal. From the results in Table 7, when augmenting the nighttime samples, the precision dropped by 2.9%, but recall improved by around twice as much (from 26.2% to 55.0%).

The results of the vehicle detection applied to the test images are shown in Fig. 8. The tires were not detected

frequently using the daytime dataset, and even when detected, the tire classification was wrong. In contrast, in the case of augmenting the synthetic nighttime samples from the GAN trained with 2,000 pairs, the tires were detected and classified accurately in a low-illumination environment.

V. CONCLUSION

In this paper, we propose a new framework of data augmentation using GAN. The detection performance of an existing BSD system is improved in a low illumination environment by applying this framework. The proposed framework generates synthetic nighttime images using a conditional GAN. A public front-view image database was used to train the GAN, but since the images were from different viewpoints, the synthetic images were not a complete reproduction of reality. Nevertheless, they are sufficient for data augmentation for vehicle detection. When the augmented BSD system was applied to actual nighttime images, the detection performance was nearly doubled compared to training with daytime images only.

In future work, we will apply the proposed framework to improve the nighttime detection performance of different viewpoint image-based ADAS applications such as those based on the around view monitoring-based applications.

REFERENCES

[1] J. B. Cicchino, "Effects of blind spot monitoring systems on police-reported lane-change crashes," *Traffic Injury Prevention*, vol. 19, no. 6, pp. 615–622, Aug. 2018.

[2] BMW. (2010). *BMW 5 Series Sedan: Lane Change and Lane Departure Warning*. [Online]. Available: http://content.bmwusa.com/microsite/f10/com/en/newvehicles/5series/sedan/2010/showroom/safety/lane_change_departure_warning.html#more

[3] Daimler. (2019). *Blind Spot Assist: Radar Sensors Monitor the Areas Directly Alongside and Behind the Car*. [Online]. Available: https://media.daimler.com/marsMediaSite/en/instance/ko/Blind-Spot-Assist-Radar-sensors-monitor-the-areas-directly-alongside-and-behind-the-car.xhtml?oid=9361526

[4] Renault. (2019). *Blind Spot Warning*. [Online]. Available: http://renault.jo/cars/NewLatitude/focusbsw.html

[5] M. Ra, H. G. Jung, J. K. Suhr, and W.-Y. Kim, "Part-based vehicle detection in side-rectilinear images for blind-spot detection," *Expert Syst. Appl.*, vol. 101, pp. 116–128, Jul. 2018.

[6] Volvo Cars. (2011). *VOLVO S80 Owner's Manual*. [Online]. Available: https://az685612.vo.msecnd.net/pdfs/567b1ecd87e93a82a543f9ab2ced196ae037db84/S80_owners_manual_MY12_EN_tp14063.pdf

[7] O. Tmenova, R. Martin, and L. Duong, "CycleGAN for style transfer in X-ray angiography," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 10, pp. 1785–1794, Oct. 2019.

[8] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, "GAN augmentation: Augmenting training data using generative adversarial networks," 2018, *arXiv:1810.10863*. [Online]. Available: http://arxiv.org/abs/1810.10863

[9] S. W. Huang, C. T. Lin, S. P. Chen, Y. Y. Wu, P. H. Hsu, and S. H. Lai, "AugGAN: Cross domain adaptation with GAN-based data augmentation," in *Proc. 15th Eur. Conf. Comput. Vis.*, 2018, pp. 731–744.

[10] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.

[11] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[12] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving video database with scalable annotation tooling," 2018, *arXiv:1805.04687*. [Online]. Available: http://arxiv.org/abs/1805.04687

[13] Nexar. (2017). *Nexar Challenge #2—Nexar*. [Online]. Available: https://www-staging.getnexar.com/challenge-2

[14] D.-C. Tseng, C.-T. Hsu, and W.-S. Chen, "Blind-spot vehicle detection using motion and static features," *Int. J. Mach. Learn. Comput.*, vol. 4, no. 6, pp. 516–521, 2014.

[15] S. Singh, R. Meng, S. Nelakuditi, Y. Tong, and S. Wang, "SideEye: Mobile assistant for blind spot monitoring," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2014, pp. 408–412.

[16] S.-M. Chang, C.-C. Tsai, and J.-I. Guo, "A blind spot detection warning system based on Gabor filtering and optical flow for E-mirror applications," presented at the IEEE Int. Symp. Circuits Syst. (ISCAS), Florence, Italy, May 2018.

[17] B.-F. Wu, H.-Y. Huang, C.-J. Chen, Y.-H. Chen, C.-W. Chang, and Y.-L. Chen, "A vision-based blind spot warning system for daytime and nighttime driver assistance," *Comput. Electr. Eng.*, vol. 39, no. 3, pp. 846–862, Apr. 2013.

[18] C. Tsuchiya, S. Tanaka, H. Furusho, K. Nishida, and T. Kurita, "Real-time vehicle detection using a single rear camera for a blind spot warning system," *SAE Int. J. Passenger Cars-Electron. Electr. Syst.*, vol. 5, no. 1, pp. 146–153, 2012.

[19] D. Dooley, B. McGinley, C. Hughes, L. Kilmartin, E. Jones, and M. Glavin, "A blind-zone detection method using a rear-mounted fisheye camera with combination of vehicle detection methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 264–278, Jan. 2016.

[20] G. Cheng and X. Chen, "A vehicle detection approach based on multi-features fusion in the fisheye images," in *Proc. 3rd IEEE Int. Conf. Comput. Res. Develop.*, vol. 4, Mar. 2011, pp. 1–5.

[21] S. G. Kim, K. H. Jung, and K. Yi, "Blind spot monitoring at night-time using rear-view camera," *Int. J. Internet Technol. Secured Trans.*, vol. 8, no. 1, pp. 15–24, Jun. 2018.

[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Benzio, "Generative adversarial nets," in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 1–9.

[23] S. Reed, Z. Akata, X. Yan, L. Logeswaran, H. Lee, and B. Schiele, "Generative adversarial text to image synthesis," in *Proc. 29th Int. Conf. Mach. Learn.*, 2016, pp. 1060–1069.

[24] 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," 2016, *arXiv:1609.04802*. [Online]. Available: http://arxiv.org/abs/1609.04802

[25] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary equilibrium generative adversarial networks," 2017, *arXiv:1703.10717*. [Online]. Available: http://arxiv.org/abs/1703.10717

[26] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2089–2093.

[27] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," 2017, *arXiv:1701.05957*. [Online]. Available: http://arxiv.org/abs/1701.05957

[28] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[29] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784

[30] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5400–5409.

[31] X. Wang, H. Yan, C. Huo, J. Yu, and C. Pant, "Enhancing Pix2Pix for remote sensing image classification," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2332–2336.

[32] Z. Wang, Z. Chen, and F. Wu, "Thermal to visible facial image translation using generative adversarial networks," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1161–1165, Aug. 2018.

[33] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[35] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 700–708.
[36] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," 2016, *arXiv:1611.02200*. [Online]. Available: http://arxiv.org/abs/1611.02200
[37] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2849–2857.
[38] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2107–2116.
[39] J.-Y. Zhu. (2017). *CycleGAN Project Page*. [Online]. Available: https://junyanz.github.io/CycleGAN/
[40] D. H. Kang, C. M. Kang, J.-S. Kim, S. Kim, W.-Y. Kim, S.-H. Lee, and C. C. Chung, "Vision-based autonomous indoor valet parking system," in *Proc. 17th Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2017, pp. 41–46.
[41] J. K. Suhr and H. G. Jung, "Rearview camera-based Stixel generation for backing crash prevention," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 117–134, Jan. 2020.
[42] C. Wang, H. Zhang, M. Yang, X. Wang, L. Ye, and C. Guo, "Automatic parking based on a Bird's eye view vision system," *Adv. Mech. Eng.*, vol. 6, Jan. 2014, Art. no. 847406.
[43] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
[44] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 886–893.
[45] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[46] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. 13rd Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

HONGJUN LEE received the B.S. degree in electronics and computer engineering from Hanyang University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree. His research interests include generative adversarial networks, data augmentation, perspective distortion correction, and stereo vision.

MOONSOO RA received the B.S. and Ph.D. degrees in electronics and computer engineering from Hanyang University, Seoul, South Korea, in 2011 and 2019, respectively. He is currently working as a Founding Member and CTO with LightVision Inc. His research interests include pattern recognition, machine learning, autonomous vehicles, and video surveillance.

WHOI-YUL KIM received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989. From 1989 to 1994, he was with the Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas. He joined Hanyang University, in 1994, where he is currently a Professor with the Department of Electronic Engineering. His research interests include health monitoring using mobile devices, visual surveillance, virtual devices, machine vision systems, advanced driver assistance systems, and 3D vision systems for sport and in-home appliances.