IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 6, 2020, accepted May 13, 2020, date of publication May 15, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994969

Semantic Segmentation With Low Light Images by Modified CycleGAN-Based Image Enhancement

SE WOON CHO, NA RAE BAEK, JA HYUNG KOO, MUHAMMAD ARSALAN, AND KANG RYOUNG PARK
Division of Electronics and Electrical Engineering, Dongguk University, Seoul 04620, South Korea

Corresponding author: Kang Ryoung Park (parkgr@dongguk.edu)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Education through the Basic Science Research Program under Grant NRF-2018R1D1A1B07041921, in part by NRF funded by the Ministry of Science and ICT (MSIT) through the Basic Science Research Program under Grant NRF-2020R1A2C1006179, and in part by NRF funded by MSIT through the Basic Science Research Program under Grant NRF-2019R1A2C1083813.

ABSTRACT In recent years, the importance of semantic segmentation has been widely recognized and the field has been actively studied. The existing state-of-the-art segmentation methods show high performance for bright and clear images. However, in low light or nighttime environments, images are blurred and noise increases due to the nature of the camera sensor, which makes it very difficult to perform segmentation for various objects. For this reason, there are few previous studies on multi-class segmentation in low light or nighttime environments. To address this challenge, we propose a modified cycle generative adversarial network (CycleGAN)-based multi-class segmentation method that improves multi-class segmentation performance for low light images. In this study, we used low light databases generated by two road scene open databases that provide segmentation labels, which are the Cambridge-driving labeled video database (CamVid) and Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago (KITTI) database. Consequently, the proposed method showed superior segmentation performance compared with the other state-of-the-art methods.

INDEX TERMS Semantic segmentation, low light, nighttime, modified CycleGAN, road scene open database.

I. INTRODUCTION

The field of deep-learning-based semantic segmentation has been actively studied since the implementation of the fully convolutional networks (FCN) [1] and SegNet [2] proposed in 2015. Subsequently, numerous convolutional neural network (CNN)-based segmentation methods were developed, showing high performance for various segmentation databases. However, most semantic segmentation studies mainly handle daytime databases or bright images, and there have been few studies on semantic segmentation dealing with nighttime databases or low light images. In addition, existing methods show good performance for bright and clear images captured in daytime but, the performance drops significantly for nighttime or low light environments. Generally, in low light environments, the amount of light is insufficient, and the image is captured with the camera's exposure time set longer than daytime. This creates motion and optical blur in the captured images and noise is also increased due to the nature of the camera sensor, making it very difficult to perform segmentation for objects in the image.

To solve this problem, various low light image segmentation methods [3]–[15] have been studied. The existing methods can be divided into single class segmentation methods and multi-class segmentation methods. In single class segmentation studies [3]–[11], segmentation is performed on a single object only such as pedestrians, vehicles, and traffic lights in a low light environment. Furthermore, as the methods consider only the characteristics of a single target object that can be distinguished from the background, relatively high segmentation performance can be achieved. In multi-class segmentation studies [12]–[15], as segmentation is performed for multiple objects in a low light or nighttime image, the features that can clearly distinguish the respective object should be extracted. However, in low light or nighttime environments, the brightness is very low, and noise and blur cause the color and shape information of objects in the image

The associate editor coordinating the review of this manuscript and approving it for publication was Madhu S. Nair.

to alter or disappear, making the segmentation of various objects very difficult.

Considering these points, we propose a multi-class segmentation method based on image enhancement in low light environments. We used a modified version of the original cycle generative adversarial network (CycleGAN) to enhance the performance of the conversion of low light or nighttime images to daytime images [16]. Unlike the original CycleGAN, in the training process of our network, we used paired data and added L1 loss between the output and the target to improve the enhancement quality over the existing model. First, using our modified CycleGAN, we generate an output image similar to the one captured in daytime from a low light image. And then, the generated image is used as input to the segmentation network to perform segmentation on various objects.

The rest of this paper is organized as follows. Section 2 introduces the previous semantic segmentation methods. Section 3 describes the contributions of our study. Section 4 describes the proposed method in detail, and Section 5 describes the experimental results with analysis. Section 6 describes the conclusions of our study.

II. RELATED WORKS

In this section, we compare and analyze various existing low light image segmentation methods. Most existing deep-learning-based segmentation studies have proposed segmentation methods of various objects in a daytime environment where bright and clear images can be captured. As representative examples, the recent daytime segmentation methods [20]–[22] proposed various deep neural networks and demonstrated their high performance through experiments using road scene databases (Cambridge-driving labeled video database (CamVid) [18] and Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago (KITTI) database [19]). The dense segmentation network (DSNet) [20] selected dense convolutional networks (DenseNet) [23] that most frequently combine and extract multi-scale information as the backbone. Moreover, they improved the performance by modifying the bottleneck structure in the dense blocks according to the segmentation purposes. The dual attention network (DANet) [21] appended two additional attention modules to the residual network (ResNet)-based dilated FCN. The position attention module and channel attention module learn and extract long-range contextual information in the spatial and channel dimensions, respectively. Compared with the existing ResNet-based dilated FCN model, the network showed higher performance. The fully residual encoder-decoder network (FRED-Net) [22] improved segmentation performance and processing speed by reducing the number of convolutional layers in the existing SegNet model and applying residual skip connections using 1 × 1 convolutional layers. These studies [20]–[22], in addition to several others, have shown high segmentation performance using various techniques and networks in daytime environments. However, in low light or nighttime environments, segmentation performance is degraded due to various factors. In an environment with very low external light, motion and optical blur are generated due to the long exposure time of the camera and movement of objects during image acquisition. In addition, noise increases due to the nature of the camera sensor.

To overcome these problems, various low light image segmentation methods [3]–[15] have been studied. Low light image segmentation studies can be divided into single class segmentation methods [3]–[11] and multi-class segmentation methods [12]–[15]. Color-feature-based methods [3]–[5] used the brightness and color information of a single object. Wang and Ren [3] improved the image contrast by preprocessing with median filtering and the histogram equalization method, and segmented a single object using the brightness and color difference between the background and foreground pixels. Haltakov *et al.* [4] used texture and color information to segment the candidate regions of a single object. Alpar [5] obtained the differential image after subtracting the greyscale image from the red channel image. The object was then segmented by applying a threshold. The studies in [3]–[5] reduced the complexity of the algorithm by using brightness and color features that are distinct from the background. However, these methods have a disadvantage in that they are sensitive to changes in brightness and color in the image. Motion-based methods [6]–[8] used multiple images to remove the background and segmentation was performed on the foreground. Soumya [6] removed the background using a threshold and dynamic matrix. Lee *et al.* [7] used additional regularization terms to remove noise in differential image and reduce the false alarm rate. Li *et al.* [8] proposed a voxel surface modeling method that uses three-dimensional geometric information without using background subtraction methods [6], [7]. The studies in [6]–[8] used multiple images to remove the background and showed high segmentation accuracy for moving objects. However, these methods can only be used in environments where the camera is fixed. In addition, because the segmentation is performed based on the movement of objects, these methods cannot classify each of the segmented objects. The edge-based method [9] uses Sobel kernels to detect the edges of objects in NIR images and select candidate regions. Then, camera geometry and template matching are used to remove false positives. A threshold-based method [10] calculated the optimal threshold to distinguish the background and foreground pixels by applying iterative thresholding, and segmented the object pixels. A superpixel-based method [11] applied a simple linear iterative clustering technique as the pre-processing step and segmented a single object using K-means clustering. All these studies [3]–[11] extracted features that can distinguish between background and foreground using various methods in low light or nighttime environments, and segmented a single object. As evident from these examples, in nighttime environments, single class segmentation studies only perform segmentation for a single object and consider the features of a single target object that can be distinguished from the background, resulting in a relatively high segmentation performance. However, multi-class segmentation studies in low light environments consist of various classes and differentiation is required for all the different classes, not just for the background. In addition, segmentation is very difficult due to problems such as high similarity between classes or occlusion.

To solve these problems, multi-class segmentation methods [12]–[15] in low light or nighttime environments have been studied. Dai and Gool [12] used a dataset with five levels of illumination changes from daytime to nighttime, and proposed a gradual model adaptation method based on transfer learning. First, RefineNet [24] based on ResNet101 is trained using a daytime dataset as the first-level dataset. With the trained model, the test is performed on the second-level dataset with slightly lower illumination than daytime, and the result obtained is used as ground truth. This transfer learning process is repeated for each level of illumination, and finally, all 1 to 4 datasets are used for the training. Sakaridis *et al.* [13] proposed a guided curriculum model adaptation method using both unlabeled real datasets with three levels of illumination change and low light labeled synthetic datasets generated by CycleGAN. This method is similar to the transfer learning method of Dai *et al.* in [12]; however, the use of labeled synthetic datasets for learning improved the segmentation performance over that of previous methods. The studies in [12], [13] could train their models with nighttime datasets without manually annotated labels. However, a dataset with various illumination changes was required, the training time was long, and the training complexity was high. Sun *et al.* [14] performed night-to-day image conversion using CycleGAN, but the generated image quality was poor and the authors did not show quantitative experimental results. The second proposed method, CycleGAN-based nighttime image augmentation method, increases the number of training data by generating synthetic nighttime images, and improves segmentation accuracy by training the proposed segmentation network. Valada *et al.* [15] designed AdapNet, a ResNet-50-based FCN model, and proposed a fusion method called convoluted mixture of deep experts (CMoDE). The dataset used in the experiment has RGB and depth images in pairs. Two AdapNets trained with RGB and depth images separately are combined into a CMoDE model, and the fusion results are obtained as output. This method shows higher performance than the existing methods with the fusion of two segmentation networks but has the disadvantage of requiring RGB and depth images in pairs. The studies in [12]–[15] used various transfer learning methods and segmentation models to improve the segmentation performance in low light or nighttime environments. However, nighttime images without applying enhancement have very low visibility. In addition, it is difficult to train segmentation networks because of lack or inaccurate label information, and the performance improvement is small.

In view of these limitations of previous studies, we propose a multi-class segmentation method based on image enhancement using modified CycleGAN in low light or nighttime environments. Our modified CycleGAN has two major structural differences from the original CycleGAN. First, by modifying the structure and number of residual blocks in the original CycleGAN, we reduced the computational cost while maintaining the enhancement quality. Second, we used paired data in the training process and increased the enhancement quality by adding L1 loss between the output and the target to the loss function of the original CycleGAN. Unlike existing multi-class segmentation methods in low light or nighttime environments [12]–[15], the modified CycleGAN is used to perform a direct enhancement of low light images to improve the segmentation performance in low light environments. First, the low light images with poor visibility are converted into an enhanced image similar to the daytime image through the modified CycleGAN, an enhancement network. The generated image has increased brightness compared with low light images, and noise and blur are reduced. It is then used as input to the segmentation network. As the two networks are separated from each other, the advantages are that the training complexity of each network can be reduced, and the modified CycleGAN can be easily combined with other segmentation networks.

Table 1 summarizes the advantages and disadvantages of the existing methods and the proposed method for semantic segmentation research in low light environments.

III. CONTRIBUTIONS

Our research is novel in the following three ways compared with previous works.

- This study is the first to propose a modified CycleGAN-based low light image segmentation method that can improve semantic segmentation performance in low light or nighttime environments. To improve the performance of the conversion of a low light or nighttime image to a daytime image, this study added the L1 loss between the output and the target to the original CycleGAN to improve the image enhancement quality.
- The computational cost is reduced by modifying the residual blocks of the original CycleGAN into a bottleneck structure and reducing their number. In this study, training and tests were performed by separating the modified CycleGAN, which enhances low light images, and the segmentation network, in consideration of training complexity and efficiency.
- A trained CNN model with algorithms and low light image databases generated from open databases are available in [25] for other researchers to use.

IV. PROPOSED METHOD

A. OVERVIEW OF PROPOSED METHOD

Figure 1 is a flowchart of the proposed method. In step (1), the input is low light images. In step (2), the CamVid database is resized to 320×240 pixels and the KITTI database to 512 × 176 pixels to minimize the change in the shape

TABLE 1. Comparisons of the previous and proposed methods for semantic segmentation in low light environments.

Category			Method	Advantage	Disadvantage
Semantic segmentation methods in daytime			DSNet [20] DANet [21] FRED-Net [22]	Improved performance by modifying existing segmentation models or adding layers.	For low light or nighttime images, semantic segmentation performance is reduced.
Semantic segmentation methods in nighttime	Single class segmentation		Color-feature-based methods [3-5] Motion-based methods [6-8] Edge-based method [9] Threshold-based method [10] Superpixel-based method [11]	- Segmentation performance was improved by considering various features of a single object that can be differentiated from the background [3-11]. - The algorithm complexity is low owing to the utilization of the brightness and color features of the object [3-5]. - Multiple images were used to remove the background region [6-8].	- Difficult to apply the method for various objects [3-11]. - Sensitive to brightness and color changes [3-5]. - Can only be used in environments where the camera is fixed [6-8].
	Multi-class segmentation	Non-enhancement-based method	Dai et al. [12] Sakaridis et al. [13] Sun et al. [14] Valada et al. [15]	- No additional enhancement is used, resulting in low computational cost and short processing time [12-15].	- Poor visibility as it does not use an enhancement technique for nighttime images [12-15]. - Label information is insufficient or inaccurate, making it difficult to train the segmentation networks [12-14].
		Enhancement-based method	Proposed method	- Modified CycleGAN was used for the image enhancement to enhance visibility and improve the segmentation performance. - Training complexity is reduced by separating enhancement and segmentation models.	Processing time is increased due to the enhancement network.

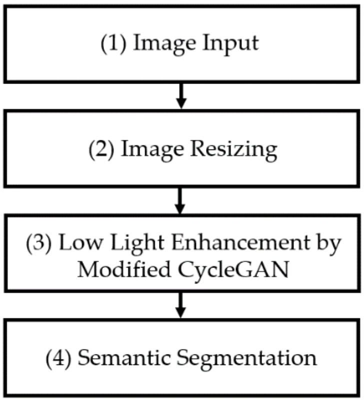

(1) Image Input
(2) Image Resizing
(3) Low Light Enhancement by Modified CycleGAN
(4) Semantic Segmentation

FIGURE 1. Flowchart of the proposed method.

of objects in the image. In step (3), the proposed modified CycleGAN is used for the enhancement of the low light images to make them similar to the original daytime image. In step (4), a segmentation network is used to segment the objects in the enhanced image and output a segmentation map.

B. IMAGE ENHANCEMENT BY MODIFIED CYCLEGAN

In general, images captured in a low light or nighttime environment have low brightness because of less external light. In addition, when acquiring an image, blur occurs because the exposure time of the camera is set long, and noise is increased due to the nature of the camera sensor. Due to these complex factors, the color and shape information of the objects in low light images is very limited or disappears, making it very difficult to segment the objects. In this paper, we propose modified CycleGAN to improve the segmentation performance for low light images, to overcome this problem. The existing CycleGAN [16] performs various image-to-image translations using unpaired data belonging to two different domains. In addition, image-to-image translation was performed using two forward and backward models for unpaired data so that only the style was similar to the target domain while maintaining the identity of the source domain image. However, in our modified CycleGAN, we aim to enhance the low light images to bright and clear images; hence, we set the low light database as the source domain and the daytime database as the target domain. In addition, unlike the existing CycleGAN, paired data are used and paired L1 loss is added to maintain the identity of the input image and to improve the enhancement quality of low light images.

Sections IV.B.1 and IV.B.2 describe the modified CycleGAN generator and discriminator, respectively, in detail, and Section IV.B.3 describes the loss function.

1) GENERATOR

In this section, we describe in detail the structures of and the differences between the generators of the original CycleGAN and our modified CycleGAN. Generally, generative adversarial networks (GANs) divided into generator models and discriminator models, and they are trained to improve each other's performance through adversarial learning techniques. CycleGAN [16] is derived from GANs [26] to perform

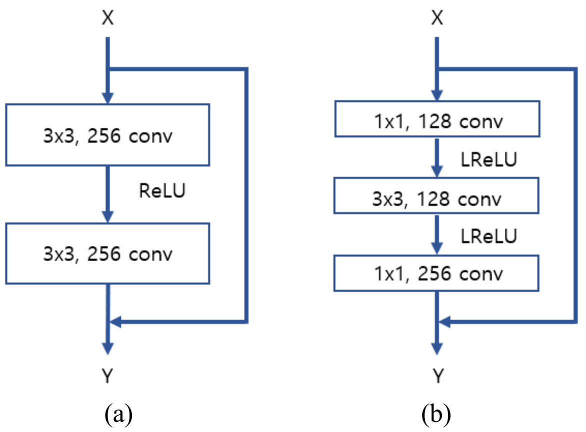


FIGURE 2. **Architectures of original and modified residual blocks: (a) an original residual block in original CycleGAN, (b) a modified residual block in modified CycleGAN.**

image-to-image translation. CycleGAN consists of two pairs of generators and discriminators and is designed to enable bi-directional image-to-image translation for different domains. The generator of the original CycleGAN can be largely divided into an encoder and a decoder. First, the encoder consists of the first convolutional layer with the size of the filters 7 × 7, and two stride-2 convolutional layers to reduce the size of the feature map. Subsequently, there are nine residual blocks as shown in Figure 2 (a) and the size of the feature map is maintained. Behind it lie the decoder part, which consists of two deconvolutional layers that use transposed convolution to increase the size of the feature map again, and finally an output layer that produces a fake image with the same size as the input image. The existing CycleGAN generator successfully performed image-to-image translation on unpaired data.

However, because this study aims at enhancement of low light images, we designed a new network that modified the generator part of the original CycleGAN to fit our purposes. Figure 3 shows the architecture of our modified CycleGAN, and Tables 2 and 3 show the structures of the generator and discriminator of our network, respectively. Compared with the existing model, our modified CycleGAN generator has two major differences. The first difference is the activation function. The existing model used a rectified linear unit (ReLU) as an activation function, but this is not suitable when the input is a low light image. Most pixel values of low light images are very low and scaled to $[-1, 1]$; hence, most input pixel values are negative. In this case, when ReLU is used, the information of the region with low pixel value can be lost, and training becomes difficult. Therefore, we replace all the activation functions of the modified CycleGAN generator with leaky ReLU (LReLU) [27], thereby reducing information loss and change in the shapes, and improving learning stability. Second, we modified the residual block, which is part of the generator. Figure 2 shows the structures of the residual blocks used in the original CycleGAN and our modified CycleGAN. The residual block of the original CycleGAN shown in Figure 2 (a) consists of two convolutional layers with a filter size of 3 × 3 and a dimension of 256, having a shortcut structure that adds the input and output of the block. In contrast, our modified CycleGAN uses a modified residual block consisting of two 1 × 1 convolutional layers and a 3 × 3 convolutional layer. As shown in Figure 2 (b), the modified residual block consists of three convolutional layers and has a bottleneck structure that reduces the number of channels in feature maps to 128 and then increases it to 256 again. In addition, the number of blocks is reduced from nine to six. By modifying the number and structure of residual blocks, we reduce computational cost and processing time while maintaining the enhancement quality of the output image, compared with the existing CycleGAN model. Comparisons and descriptions are detailed in Section V.F.

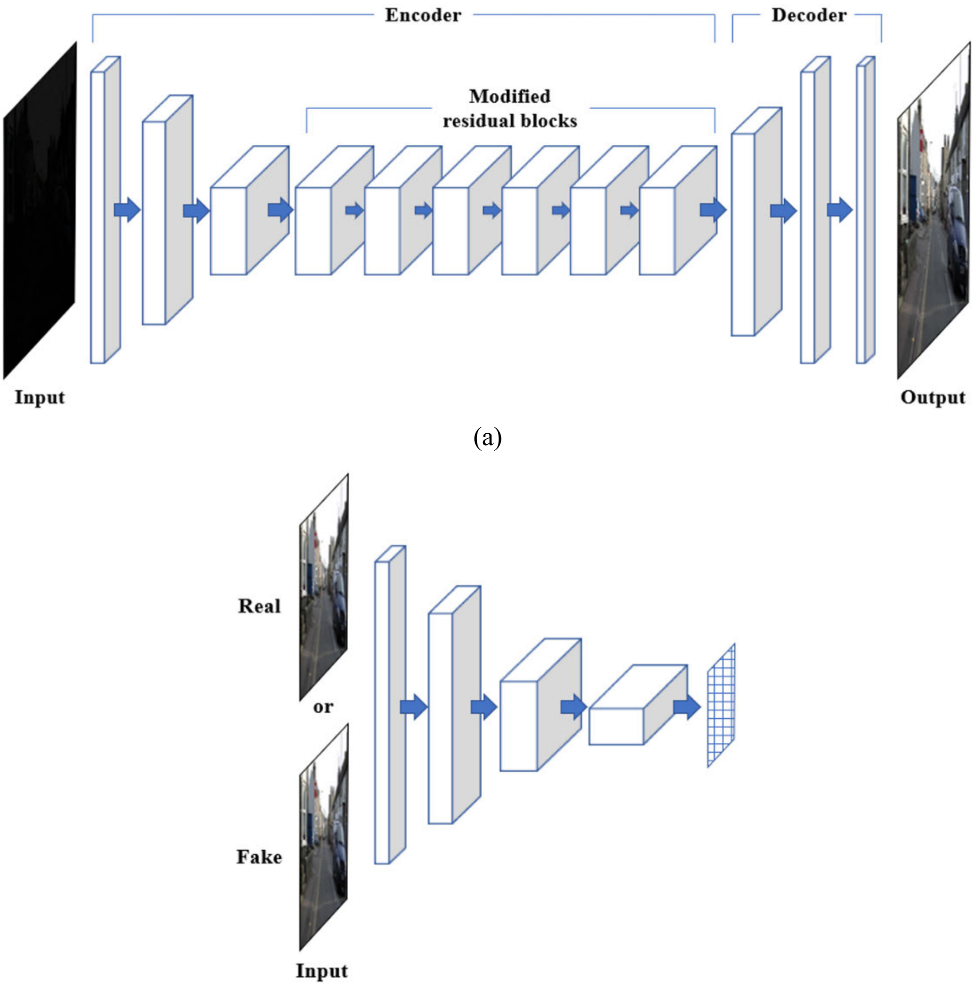
2) DISCRIMINATOR

In the existing GANs [26], discriminators exist in pairs with generators and are trained to discriminate between real and fake images as a single output from the generated fake images. In the case of CycleGAN, 70 × 70 PatchGAN [28] is used as a discriminator to distinguish between real and fake images for overlapping image patches of size 70 × 70. In this study, we used the same discriminator as the original CycleGAN. Figure 3 (b) and Table 3 show the discriminator of our modified CycleGAN. The discriminator consists of three stride-2 convolutional layers, one stride-1 convolutional layer, and an output layer. A real or fake image is reduced to 1/8 in size through convolutional layers, and the final output layer produces a one-channel prediction map with a value between 0 and 1. In Table 3, the size of the input image is $320 \times 240 \times 3$, and the output of the discriminator is a prediction map of size $40 \times 30 \times 1$. The single value of the prediction map has a 70 × 70 receptive field, which indicates that real or fake is distinguished for overlapping image patches of size 70 × 70.

3) LOSS FUNCTION

To maintain the identity of the input image, the original CycleGAN proposed a cycle consistency loss and combined it with least square GAN (LSGAN) loss [29]. However, to improve the performance of conversion of low light or nighttime images to daytime images, our modified CycleGAN uses paired L1 loss, which uses the L1 distance between the output and the target, in addition to the loss function of the original CycleGAN. First, we used the LSGAN loss to perform adversarial learning of our modified CycleGAN. In the existing GANs, log function-based adversarial loss was used, but least-square-based LSGAN loss was used in our modified CycleGAN to increase the learning stability and convergence speed of our network. Equation (1) expresses the LSGAN loss.

$$\mathrm{L_{LSGAN}(G, D)} = \mathrm{E}_{y\sim p_{data}(y)}\left[(\mathrm{D}(y)-1)^2\right] + \mathrm{E}_{x\sim p_{data}(x)}\left[(\mathrm{D}(\mathrm{G}(x)))^2\right] \quad (1)$$



(a)

(b)

FIGURE 3. Architecture of modified CycleGAN. The figure shows the (a) generator and (b) discriminator of our modified CycleGAN. The generator is largely divided into two parts: encoder and decoder. The encoder consists of the first convolutional layer with filter size 7 × 7, two stride-2 convolutional layers to reduce the size of the feature maps, and finally six modified residual blocks. The decoder consists of two deconvolutional layers to increase the size of the feature maps and an output layer that produces a fake image with the same size as the input image. The discriminator consists of four stride-2 convolutional layers with the filter size 4 × 4 and an output layer that distinguishes between real and fake images.

G represents the generator and D represents the discriminator. There are two types of domains, nighttime or daytime, where x is an input image of the source domain X and y is an image of the target domain Y. G(x) represents the generated output image similar to the target domain. G deceives D and maximizes the value of L_{LSGAN}. D distinguishes whether the input image is real or fake, and minimizes the value of L_{LSGAN}.

The second is cycle consistency loss, which is used in the same way as the original CycleGAN. Cycle consistency indicates that the input image can be converted to another domain through a generator and then restored to the original image using another generator. Equation (2) expresses the cycle consistency loss.

$$L_{CYCLE}(G_N, G_D) = E_{x\sim p_{data}(x)}[||G_N(G_D\,(x)) - x||_1] + E_{y\sim p_{data}(y)}\left[||G_D(G_N\,(y)) - y||_1\right] \quad (2)$$

x is a low light image of the nighttime domain, and y is an image of the daytime domain. G_N is a generator that converts the input image into the nighttime domain, and G_D is another generator that converts the input image into the daytime domain. In our modified CycleGAN, converting the input low light image x into the daytime domain and then restoring it is defined as forward cycle consistency and the reverse process as backward cycle consistency. Finally, the sum of the L1 distance between the reconstructed $G_N(G_D(x))$ and x and the L1 distance between the reconstructed $G_D(G_N(y))$ and y is the cycle consistency loss. Accordingly, the mode collapse problem was alleviated and the conversion of a low light or nighttime image to a daytime image was performed while maintaining the identity of the input image.

Third, we use the paired L1 loss to improve the enhancement quality of low light images, as shown in Equation (3). As the original CycleGAN uses unpaired data, it is not

TABLE 2. Architecture of the generator of Figure 3 (Conv, Norm, and LReLU indicate the convolutional layer, instance normalization, and leaky rectified linear unit, respectively. Tanh indicates hyperbolic tangent).

Layer type			Number of filters	Size of feature map (width×height× channel)	Size of kernel (width×height)	Number of strides	Number of paddings	Number of iterations
Input layer [image]				320×240×3				
Encoder	1st Convolutional layer (Conv + LReLU)		64	320×240×64	7×7	1×1	3×3	1
	2nd Convolutional layer (Conv + Norm + LReLU)		128	160×120×128	3×3	2×2	1×1	1
	3rd Convolutional layer (Conv + Norm + LReLU)		256	80×60×256	3×3	2×2	1×1	1
	Modified residual blocks	Conv + Norm + LReLU	128	80×60×128	1×1	1×1		6
		Conv + Norm + LReLU	128	80×60×128	3×3	1×1	1×1	
		Conv + Norm	256	80×60×256	1×1	1×1		
		Identity layer		80×60×256				
Decoder	1st Deconvolutional layer (Deconv + Norm + LReLU)		128	160×120×128	3×3	2×2	1×1	1
	2nd Deconvolutional layer (Deconv + Norm + LReLU)		64	320×240×64	3×3	2×2	1×1	1
	Output layer (Conv + Tanh)		3	320×240×3	7×7	1×1	3×3	1

possible to use the distance between the generator's output image and the target image. However, in our study, as the low light database and daytime database are in pairs, we can use the L1 distance between the output and the target as the loss value.

$$\mathrm{L_{PAIR}}(\mathrm{G_N}, \mathrm{G_D}) = \mathrm{E}_{x\sim p_{data}(x)}[||\mathrm{G_D}(x) - y||_1] + \mathrm{E}_{y\sim p_{data}(y)}\left[||\mathrm{G_N}(y) - x||_1\right] \quad (3)$$

The input image x is an image of the nighttime domain and the target image y is an image of the daytime domain. $\mathrm{G_D}$ (x) and $\mathrm{G_N}$ (x) indicate the input images converted to the daytime and nighttime domains, respectively. By adding the paired L1 loss, we obtain output images that are sharper and more similar to the target image than when using the original CycleGAN.

In the training of our modified CycleGAN, the three loss functions described above were used in combination. Equation (4) is our final loss function.

$$\begin{aligned}\mathrm{L_{TOTAL}}(\mathrm{G_N}, \mathrm{G_D}, \mathrm{D_N}, \mathrm{D_D}) = {} & \mathrm{L_{LSGAN}}(\mathrm{G_N}, \mathrm{D_N}) \\ & + \mathrm{L_{LSGAN}}(\mathrm{G_D}, \mathrm{D_D}) \\ & + \lambda \mathrm{L_{CYCLE}}(\mathrm{G_N}, \mathrm{G_D}) \\ & + \lambda \mathrm{L_{PAIR}}(\mathrm{G_N}, \mathrm{G_D}) \quad (4)\end{aligned}$$

The LSGAN loss is calculated as the outputs of the generator and discriminator of the forward network (night to day) and backward network (day to night) of our modified CycleGAN, respectively. Cycle consistency loss and paired L1 loss are calculated only through the outputs of the generator. λ is a balancing parameter that adjusts the size of the loss so that all the loss functions can affect the learning.

4) DIFFERENCES BETWEEN ORIGINAL CYCLEGAN AND OUR MODIFIED CYCLEGAN

In this section, we summarize the differences between the original CycleGAN and our modified CycleGAN as follows.

- The original CycleGAN performs image-to-image translation using unpaired data, but as our modified CycleGAN aims at low light image enhancement to improve semantic segmentation performance, paired data were used and the resulting enhancement quality was improved compared with that of the original model.
- The original CycleGAN uses a combination of adversarial loss and cycle consistency loss as a loss function. In our modified CycleGAN, the paired L1 loss between the output and the target is added to the loss function of the existing CycleGAN to improve the enhancement quality of low light images and to maintain the identity of the input image.

TABLE 3. Architecture of the discriminator of Figure 3 (Conv, Norm, and LReLU indicate the convolutional layer, instance normalization, and leaky rectified linear unit, respectively).

Layer type	Number of filters	Size of feature map (width×height×channel)	Size of kernel (width×height)	Number of strides	Number of paddings
Input layer [image]		320×240×3			
1st Convolutional layer (Conv + LReLU)	64	160×120×64	4×4	2×2	1×1
2nd Convolutional layer (Conv + Norm + LReLU)	128	80×60×128	4×4	2×2	1×1
3rd Convolutional layer (Conv + Norm + LReLU)	256	40×30×256	4×4	2×2	1×1
4th Convolutional layer (Conv + Norm + LReLU)	512	40×30×512	4×4	1×1	1×1
Output layer (Conv)	1	40×30×1	1×1	1×1	1×1

- As the pixel values of the low light images are very low and the network input range is scaled to $[-1, 1]$, most input pixel values are negative. In consideration of this, our modified CycleGAN reduced the loss of information on low pixel values and improved learning stability by using the activation function LReLU, instead of the ReLU used in the existing CycleGAN generator.
- In the generator of the existing CycleGAN, nine residual blocks consisting of two convolutional layers were used. However, in the generator of our modified CycleGAN, we modified the residual block in the form of a bottleneck layer using 1×1 convolution. In addition, by reducing the number of residual blocks from nine to six, the enhancement quality was maintained and the number of network learning parameters was reduced.

C. MULTI-CLASS SEGMENTATION WITH CNN

In this section, we describe the existing state-of-the-art segmentation networks used to measure the segmentation performance for low light images. In this study, we used four segmentation models. From sections IV.C.1 through sections IV.C.4 in the order of FCN [1], SegNet [2], pyramid scene parsing network (PSPNet) [17], and image cascade network (ICNet) [30], we describe the structure and characteristics of the respective model.

1) FULLY CONVOLUTIONAL NETWORK (FCN)

FCN [1] is an end-to-end network that performs segmentation using only convolutional layers. State-of-the-art classification networks used as a base model and replaced all fully connected layers with 1×1 convolutional layers to maintain spatial information. In addition, the skip layer fusion technique was used to reduce the information loss due to the pooling layer, and multi-scale feature maps were used to show high segmentation performance. In OUR experiment, we used an FCN model that has two skip layers and upsamples feature maps to eight times, and produced prediction maps as an output. We used a visual geometry group (VGG) Net-16 [31] model pretrained with imagenet as an encoder.

2) SEGNET

SegNet [2] does not use a fully connected layer and is a deep FCN composed of 26 convolutional layers using VGG Net-16 [31] as a base model. SegNet consists of an encoder that extracts the features of the input image and a decoder that upsamples the reduced feature maps back to the original size. There are five pooling layers and upsampling layers in the encoder and decoder, respectively, and each upsampling layer is connected to the corresponding pooling layer in the encoder. Unlike the existing segmentation networks, SegNet adds max pooling indices received from the corresponding pooling layers in the encoder when upsampling feature maps in each upsampling layer. The max pooling indices technique does not simply use the max value of the input feature maps, but stores the location information of the max value. With the use of the max pooling indices, the location and shape information of objects in the image is maintained and segmentation performance is improved.

3) PYRAMID SCENE PARSING NETWORK (PSPNET)

PSPNet [17] is an FCN network based on the ImageNet pre-trained ResNet-101 [32]. The structure of this network divided into feature extractor, pyramid pooling module, and prediction layer. The feature extractor uses layers up to the average pooling layer in the ResNet-101 model, and outputs 2048 dimensional feature maps reduced by 1/8 of the width and height of the input image. In the pyramid pooling module, four different average pooling layers are used to obtain reduced feature maps of sizes 1×1, 2×2, 3×3, and 6×6. The dimension of each feature map is reduced to 512 through the 1×1 convolutional layer, and then the feature map is upsampled back to its original size by applying bilinear interpolation. The four feature maps generated by the pyramid pooling module are concatenated with the output of the feature extractor, and finally, a one-channel prediction map is output through the prediction layer. Zhao *et al.* [17] applied the pyramid pooling module to a ResNet-101-based FCN model to extract global context information efficiently, and improved segmentation performance compared with that of previous segmentation networks.

4) IMAGE CASCADE NETWORK (ICNET)

ICNet [30] is a cascade network designed to perform real-time segmentation on high-resolution images. The inputs are three images, and only their resolution is different. One is the original high-resolution image, and the others are medium- and low-resolution images downsampled by factors of 2 and 4, respectively. Each multi-resolution image is the input to cascade branches, which are different sub-networks of ICNet. The output feature maps of different cascade branches are fused by two cascade feature fusion (CFF) units. The CFF unit is used to combine the output feature maps of two cascade branches and output a prediction map. ICNet uses the cascade branches, which are three sub-networks, and the CFF unit to perform real-time segmentation. In addition, to improve the optimization and segmentation performance of each cascade branch, a cascade label guidance strategy using three ground-truth labels with different scales was applied. In the study by Zhao *et al.* [30], through comparative experiments using various databases, ICNet can perform segmentation in real time, and it shows similar segmentation performance to the other existing state-of-the-art networks.

V. EXPERIMENTAL RESULTS

A. EXPERIMENTAL DATABASES

1) CAMVID AND KITTI DATABASES

In our study, two famous road scene segmentation databases were used. As shown in Figure 4 (a), the first database is CamVid [18]. This database consists of video frames that capture road scenes with a camera installed on a moving vehicle. The resolution of each image is 960 × 720 pixels (width and height, respectively), with 701 images each for RGB color images and ground-truth label images. The number of segmentation classes is 11, and each pixel value of the ground-truth label represents a class. In our experiment, for two-fold cross validation, the 701 images were divided into two subsets of 351 and 350 images. That is, the first validation was performed with 351 images for training and the remaining 350 images for testing, and the second validation was performed with 350 images for training and the remaining 351 images for testing.

As shown in Figure 4 (c), the second database is the KITTI [19] database. This database also captured images taken with a camera mounted on a moving vehicle. The image resolution is 1242 × 375 pixels (width and height, respectively), and the type and number of segmentation classes are the same as those in CamVid. However, this database does not provide a ground-truth label for the test set. In our study, a total of 445 RGB images and ground-truth labels from the KITTI database provided in a previous study [33] were used. In the same way as the above database, this database was divided into two subsets of 223 and 222 images for a two-fold cross-validation experiment.

2) SYNTHESIZED LOW LIGHT CAMVID AND KITTI DATABASES

In this experiment, we used synthesized databases that are similar to real low light environments to perform multi-class segmentation in low light environments. The road scene open databases captured in the existing nighttime environment include Berkeley deep drive 100K (BDD100K) [34], Nighttime Driving [12], and Alderley [35] databases. However, these nighttime road scene databases either do not have segmentation labels, or have very few ones, and have no daytime image pairs. In addition, images taken in real low light or nighttime environments have poor image quality and visibility due to low brightness, blur, and noise, making it difficult for humans to create segmentation labels for all the objects in the image and the labels are not accurate. Therefore, to utilize accurate segmentation labels and paired images, experiments were performed using the Syn-CamVid and Syn-KITTI databases, which are the results of converting the daytime CamVid and KITTI databases into low light images, respectively. Figures 4 (b) and (d) show example images of the Syn-CamVid and Syn-KITTI databases, respectively. To create extremely low light images similar to an actual low light environment with little external light, we have used the existing low light image generation methods in combination [36]–[38]. In a real low light environment with little external light, the brightness value does not decrease linearly. When comparing the daytime image with the nighttime image, the brightness of highly bright pixels will decrease more, whereas that of the pixels with lower brightness will decrease less. We used gamma correction [39] to produce this nonlinear brightness change. First, the clear RGB images are converted to HSV images. The three channels H, S, and V correspond to the hue, saturation, and values, respectively. The gamma correction is applied only to the V channel values to reduce the brightness value nonlinearly. Second, in a low light environment, blurry images are captured due to the amount of light and the camera's exposure time, and we used the Gaussian blur kernel to implement this effect. Finally, the noise in the low light image is generated by the camera sensor, which is added in this experiment using the Gaussian and Poisson noise functions.

Equation (5) is the formula for converting a bright and clear image into a low light image [36]–[38].

$$I_o = B_G(S \cdot (I_i)^{\gamma}) + N_G + N_P \tag{5}$$

I_i is the V channel value of the HSV image and I_o is the synthesized low light image. B_G is the Gaussian blur function, and the standard deviation σ value was set to be random between 1.5 and 2. S and γ are gamma correction parameters, and S was set as 0.06 (CamVid) or 0.08 (KITTI) and γ was set as 2.5. N_G is the Gaussian noise, and N_P is the Poisson

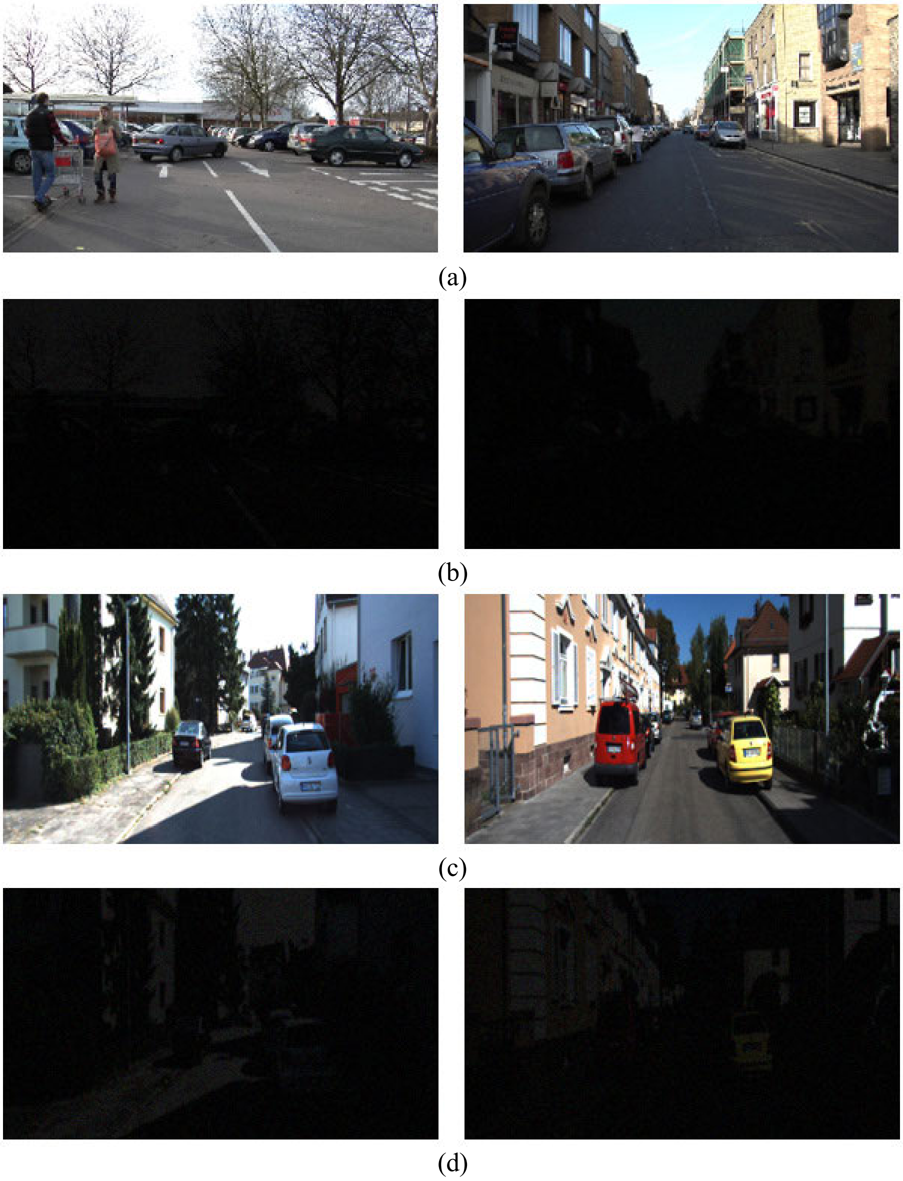


FIGURE 4. Examples of experimental databases. (a) Original CamVid database, (b) Syn-CamVid database, (c) original KITTI database, and (d) Syn-KITTI database.

noise. Table 4 summarizes the number of images, size of the images, number of classes, and pixel brightness values for the two databases used in our experiment.

B. TRAINING OF MODIFIED CYCLEGAN

Both the training code and testing code of our modified CycleGAN were implemented using the TensorFlow framework (version 1.8.0) [40]. To train our modified CycleGAN from scratch, we applied the adaptive moment estimation (ADAM) optimizer [41] with the weight parameter optimization method, and it was set that beta1 (momentum) was 0.5, beta2 was 0.999, epsilon was 1E-08, and the initial learning rate was 0.0004. The number of iterations is calculated by "the number of training images / batch size," and the number of iterations is defined as 1 epoch. For the training of our modified CycleGAN, the batch size was set to be 1 and the number of epochs to be 300. From epoch 1 to epoch 200, the learning rate remains the same at 0.0004, and for the remaining 100 epochs, It was set so that the learning rate decreases linearly to 0. To balance between the respective loss values, the $\lambda$of equation (4) was set at 10. The input image size of our network was resized to 320 × 240 pixels (width and height, respectively) for the CamVid AND Syn-CamVid databases, and 512 × 176 pixels (width and height, respectively) for the KITTI and Syn-KITTI databases, considering the aspect ratio of the two databases used in the experiment. The source domain was set to be the Syn-CamVid or Syn-KITTI database, and the target domain was set as the original daytime CamVid or KITTI database to perform the training. Figure 5 is a graph showing the changes in the training

TABLE 4. Descriptions of experimental databases.

Database	Syn-CamVid		Syn-KITTI	
Subset	Subset 1	Subset 2	Subset 1	Subset 2
Number of images	351	350	223	222
Image sizes (pixels) (width×height)	320×240		512×176	
Number of segmentation classes	11		11	
Pixel brightness value (min–max)	0–15		0–20	

loss with epochs during the training of our network. As the training loss converged to a sufficiently low value with the increase in epoch in the training, we can observe that our modified CycleGAN has been optimized.

All our experiments were performed using a desktop computer (Intel ® Core^{TM}I7-7700 CPU @ 3.6 GHz (4 cores) with 16 GB of main memory) equipped with an NVIDIA GeForce GTX 1080 (2560 compute unified device architecture (CUDA) cores) [42] with a graphics memory of 8 GB (NVIDIA, Santa Clara, CA, USA).

C. TESTING OF MODIFIED CycleGAN WITH SYNTHESIZED LOW LIGHT CAMVID AND KITTI DATABASES

1) LOW LIGHT IMAGE ENHANCEMENT WITH SYN-CAMVID DATABASE

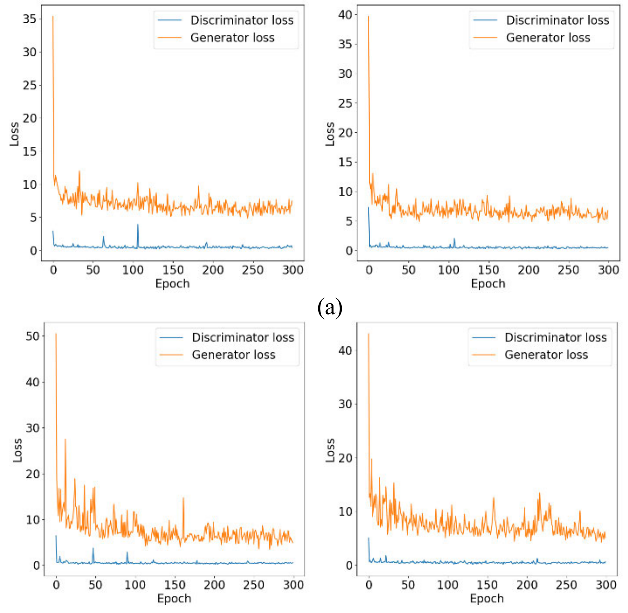
To improve the segmentation performance in low light or nighttime environments, we propose an enhancement-based low light image segmentation method using modified CycleGAN. For a quantitative comparison of the enhancement quality of low light images, the signal-to-noise ratio (SNR) [43], peak-signal-to-noise ratio (PSNR) [44], and structural similarity (SSIM) [45] were used as the evaluation indicators. SNR and PSNR measure the enhancement quality based on the mean squared error (MSE) between two images, and Equations (6)–(8) express the mathematical formulas for MSE, SNR, and PSNR, respectively.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_o\left(i,j\right) - I_e\left(i,j\right)\right]^2 \tag{6}$$

$$SNR = 10log_{10}\left(\frac{\frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_o(i,j)\right]^2}{mn}}{MSE}\right) \tag{7}$$

$$PSNR = 10log_{10}\left(\frac{255^2}{MSE}\right) \tag{8}$$

I_o is a bright and clear daytime image and I_e is the enhanced image generated by the modified CycleGAN. M and n represent the width and height of the image, respectively. Equation (9) expresses the mathematical formula of SSIM.



(a)

(b)

FIGURE 5. Training loss graphs of modified CycleGAN. The left and right figures are the training loss graphs with subset 1 and subset 2 of Table 4, respectively. Training loss graphs with (a) Syn-CamVid database and (b) Syn-KITTI database.

TABLE 5. Quality evaluation of low light image enhancement on CamVid database generated by the proposed method and previous methods.

Methods	PSNR	SNR	SSIM
PLN [46]	11.52	5.05	0.31
Pix2Pix [28]	15.06	8.59	0.35
CycleGAN[16]	19.78	13.31	0.52
Proposed method	22.62	16.15	0.65

$$SSIM = \frac{(2\mu_e\mu_o + C1)(2\sigma_{eo} + C2)}{(\mu_e^2 + \mu_o^2 + C1)(\sigma_e^2 + \sigma_o^2 + C2)} \tag{9}$$

μ_o and σ_o represent the mean and standard deviation of the pixel values of a daytime image, respectively, μ_e and σ_e represent the mean and standard deviation of the pixel values of an enhanced image generated by the modified CycleGAN, respectively, and σ_{eo} is the covariance of the two images. C1 and C2 are positive constants set so that the denominator does not become zero.

In the first experiment, we converted low light images to daytime images by applying the proposed method and the previous methods [16], [28], [46] to the synthesized low light CamVid database. And we compared and analyzed the low light image enhancement qualities of the output images. CycleGAN [16] and Pix2Pix [28] are GAN-based image-to-image translation networks, and the perceptual loss network (PLN) [46] is a single style transfer network without using a discriminator. CycleGAN is a GAN-based network that performs learning with unpaired data, and with the use of cycle consistency loss, image-to-image translation is performed while maintaining the identity of the input images. Pix2Pix is a conditional GAN that performs learning with paired data and uses a generator with a U-Net structure and skip connection. Unlike GAN-based networks, PLN does not use a discriminator and performs a style transfer using perceptual loss based on the features extracted with VGG Net-16 [31]. Table 5 shows the numerical evaluation of the enhancement quality of the output images generated using the proposed and existing methods. Two-fold cross validation was performed for the training and testing processes of all the networks, and the evaluation indicators were measured as the average values of PSNR, SNR, and SSIM obtained by testing subset 1 and subset 2 of the database. As shown in Table 5, the PSNR, SNR, and SSIM values measured by the enhanced images generated by our modified CycleGAN are the highest compared with those of the existing methods.

Figure 6 shows examples of the result images generated by using the proposed method and the existing methods with the Syn-CamVid database as input. The PLN results in Figure 6 show that the enhancement quality was the lowest compared with that of the other methods, and the color information of the objects in the image was largely lost compared with daytime images. In particular, in the lower part of the resulting image, both shape and color information disappeared. Examining the resulting image of Pix2Pix, the noise was considerably removed overall and the shape of the objects in the image was well restored. However, the Pix2Pix network did not properly restore the detail information of the objects and tended to produce blurry results. Compared with Pix2Pix, the resulting images of the original CycleGAN are sharper, and this network can reconstruct the detail information of the objects. However, small objects disappeared and noisy images tended to be produced. Finally, the output images generated by the proposed method in Figure 6 show that the enhancement quality was the best compared with that of the previous methods. In particular, more noise was removed compared with the original CycleGAN result, and the shape and color of the objects were restored more clearly. The experimental results in Table 5 and Figure 6 show that the proposed method has the best enhancement quality in quantitative and visual terms compared with the previous methods.

2) LOW LIGHT IMAGE ENHANCEMENT WITH SYN-KITTI DATABASE

In this section, we converted low light images to daytime images by applying the modified CycleGAN to the Syn-KITTI database and evaluated the enhancement quality quantitatively and visually based on the original daytime KITTI database. In the same way as the previous experiments, the results of the proposed method and the previous methods [16], [28], [46] were compared and analyzed using the Syn-KITTI database. Table 6 quantitatively shows the enhancement quality for the enhanced KITTI database generated by the proposed and previous methods. As shown in Table 6, the proposed method showed the highest PSNR, SNR, and SSIM values compared with the previous methods, indicating that the enhancement performance of our modified CycleGAN was quantitatively superior.

TABLE 6. Quality evaluation of low light image enhancement on the KITTI database generated by the proposed method and previous methods.

Methods	PSNR	SNR	SSIM
PLN [46]	13.02	6.55	0.39
Pix2Pix [28]	12.22	5.75	0.28
CycleGAN[16]	19.31	12.84	0.51
Proposed method	22.73	16.26	0.67

Figure 7 shows examples of the result images generated by using the proposed method and the existing methods with the Syn-KITTI database as input. The PLN results show that the enhancement quality is the lowest, especially in areas with shadows or low brightness. The result of Pix2Pix shows that, although noise is removed, blurry images are still produced. The original CycleGAN recovers the detail information of objects in an image well, but tends to generate a noisy image. Compared with these previous methods, the results of our modified CycleGAN show that the noise and blur are considerably removed from the images and the enhancement quality was visually comparable to the daytime images. The comparative experiments using the Syn-CamVid and Syn-KITTI databases show that the enhancement performance of our proposed method is the best for all the databases compared with the previous methods.

The PLN has the lowest enhancement quality for the two databases because the shape and color information of the objects in the low light image is insufficient to extract features, and the perceptual loss did not work properly during the training of PLN. The generator of the Pix2Pix has a U-Net structure and when the input image passes through the encoder, the size of the feature maps is reduced to 1/256 of the size of the input image. As low light images contain very little information, detail information may be lost when excessively compressing the features of an image. In addition, the L1 loss is used in the training process of Pix2Pix, and low-frequency features tend to be used more for the training because this loss is calculated as an average of pixel value differences. The blurry images of Pix2Pix are attributed to these two reasons. As the original CycleGAN is trained from unpaired data, it is designed to generate output images while maintaining the identity of input images using cycle consistency loss. However, low light input images have very low brightness values, and considerable blur and noise; hence, cycle consistency loss does not work properly and it is difficult to maintain the identity of the input image. Our modified CycleGAN reduces the loss of information in low light images by using LReLU as an activation function. In addition, unlike the original CycleGAN, our network can directly learn the difference between the output and the target by using paired data and paired L1 loss, and generate an output image similar to the daytime image. Consequently, the proposed method showed the best performance in the conversion of low light images to daytime images for the Syn-CamVid and Syn-KITTI databases.

D. TESTING OF SEGMENTATION NETWORKS WITH SYNTHESIZED LOW LIGHT CAMVID DATABASE

1) LOW LIGHT IMAGE SEGMENTATION WITHOUT ENHANCEMENT

In the experiments in Sections V.C, as the first step, we converted low light images to daytime images using the modified CycleGAN and obtained enhanced databases similar to the original daytime databases. In the second step, we performed multi-class segmentation with the output images generated by the modified CycleGAN. In this study, our final goal was to improve semantic segmentation performance in low light or nighttime environments. Therefore, our experiment focused on the extent of improvement of segmentation performance compared with low light databases. For a quantitative evaluation of the segmentation performance, we used the pixel accuracy (Pixel Acc), mean class accuracy (Class Acc), and mean intersection over union (Mean IOU) used in [1, 2] as the evaluation indicators. They are calculated as the ratio of true positive (TP), false positive (FP), and false negative (FN). These evaluation indicators are expressed in Equations (10)–(12).

$$Pixel\ Acc = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} (FP_i + TP_i)} \tag{10}$$

$$Class\ Acc = \frac{1}{L}\sum_{i=1}^{L}\left(\frac{TP_i}{FP_i + TP_i}\right) \tag{11}$$

$$Mean\ IOU = \frac{1}{L}\sum_{i=1}^{L}\left(\frac{TP_i}{FP_i + TP_i + FN_i}\right) \tag{12}$$

L represents the number of class labels. TP$_i$ represents the number of pixels that have been correctly predicted when the real ground-truth label class was i and the prediction result was predicted as class i accordingly. FP_i represents the number of pixels with incorrect prediction results from the pixels with class i. FN$_i$ represents the number of pixels with the incorrect prediction results among the pixels with the real ground-truth label as class i. Pixel Acc in Equation (10) represents a correctly predicted ratio of the prediction results of the segmentation network for all classes. Class Acc of Equation (11) represents a value calculated by averaging the accuracy of the prediction results for each class. Mean IOU in Equation (12) represents the average value of the ratio of intersections over unions for each class. In this section, we measured how much the segmentation performance is dropped in a low light environment compared with a daytime environment when no image enhancement method is used. Table 7 shows the segmentation performances measured for the daytime CamVid and Syn-CamVid databases using the state-of-the-art segmentation networks. All the segmentation networks of Table 7 are trained from scratch with each database. As shown in Table 7, all four networks showed high segmentation performance for the original daytime CamVid database. However, the segmentation performances for the low light CamVid database were significantly dropped when the image enhancement method was not used.

TABLE 7. Comparisons of segmentation performances with daytime CamVid and Syn-CamVid databases (All networks are trained from scratch with each database) (unit: %).

Methods	Pixel Acc		Class Acc		Mean IOU	
	Daytime	Low light	Daytime	Low light	Daytime	Low light
FCN [1]	90.22	61.27	65.34	26.35	57.49	19.26
SegNet [2]	89.14	64.32	72.99	50.19	60.42	31.73
PSPNet [17]	92.85	74.65	75.43	37.17	67.63	29.11
ICNet [30]	90.59	68.17	70.36	35.78	61.1	27.05

Figure 8 shows examples of segmentation results that were tested with the four segmentation networks using the Syn-CamVid database. The images in the second row of Figure 8 are the inputs to the segmentation networks, the low light images belonging to the Syn-CamVid database. Low light images have very low illumination; hence, the color information of objects is very limited. Furthermore, due to noise and blur, the shape of the objects is deformed and the boundaries between them are unclear. Therefore, training of segmentation networks becomes very difficult and segmentation performance is greatly reduced. Considering the segmentation results of the four networks, we can observe that objects of large size such as sky, building, and road show object shapes to some extent, but small objects have not been segmented. Furthermore, the boundaries between the objects are unclear overall and the shape information of the objects disappears. From the results in Table 7 and Figure 8, we can observe that the segmentation performance in the low light environment is significantly reduced compared with that in the daytime environment.

2) ABLATION STUDY (LOW LIGHT IMAGE SEGMENTATION WITH IMAGE ENHANCEMENT)

In our experiments using the Syn-CamVid database, we observed that the segmentation performance dropped significantly when the semantic segmentation was performed directly without applying the image enhancement method. We used a modified CycleGAN combined with a segmentation network to tackle this problem. First, the modified CycleGAN is used to enhance the low light image and then the segmentation network is used to output the segmentation results. We conducted a comparative experiment using four segmentation networks to confirm that the segmentation performance is improved when our modified CycleGAN and a segmentation network are combined. Table 8 shows the measured segmentation performance before and after

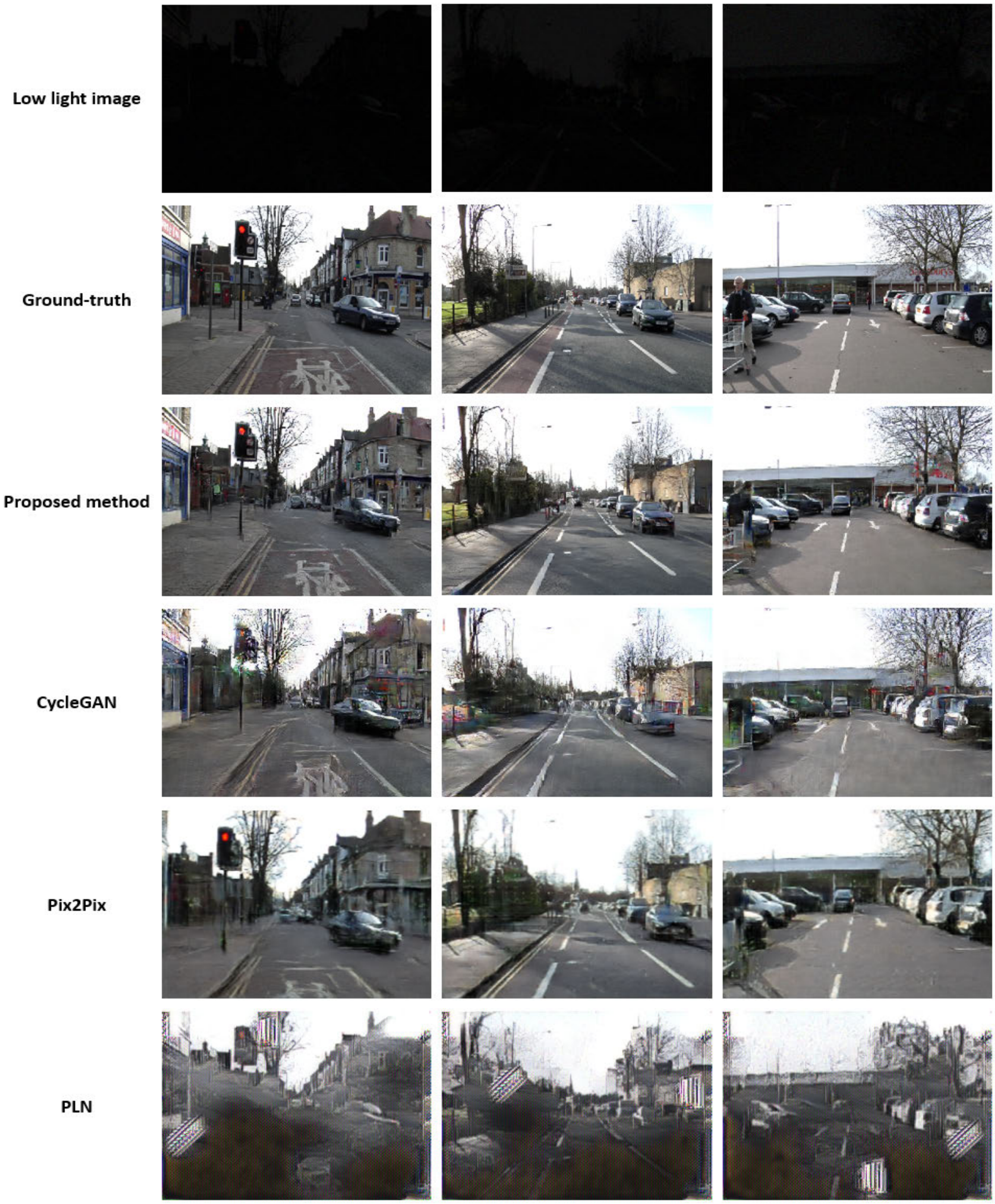


FIGURE 6. Examples of low light image enhancement results of the Syn-CamVid database. The first row shows the input images of the source domain, the second row shows each low light image and the accumulated ground-truth daytime images of the target domain, and the third row to the last row show the enhancement result images obtained by the proposed method, CycleGAN, Pix2Pix, and PLN, respectively.

applying the enhancement technique using the modified CycleGAN.

As shown in Table 8, the segmentation performance without enhancement was very low for all evaluation indicator values. In contrast, we can observe that the performance was greatly improved for all the segmentation networks after applying the enhancement technique using the modified CycleGAN. PSPNet, in particular, showed the best performance with a pixel accuracy of 88.58%, class accuracy of 63.97%, and mean IOU of 55.31%. Through comparative experiments using the Syn-CamVid database, we confirmed that the modified CycleGAN not only enhanced the low light image well, but also significantly improved the semantic segmentation performance. Figure 9 shows examples of the segmentation results using the Syn-CamVid database, which were tested with models combining our modified CycleGAN with each segmentation network. The first and third columns show the results of testing the low light images

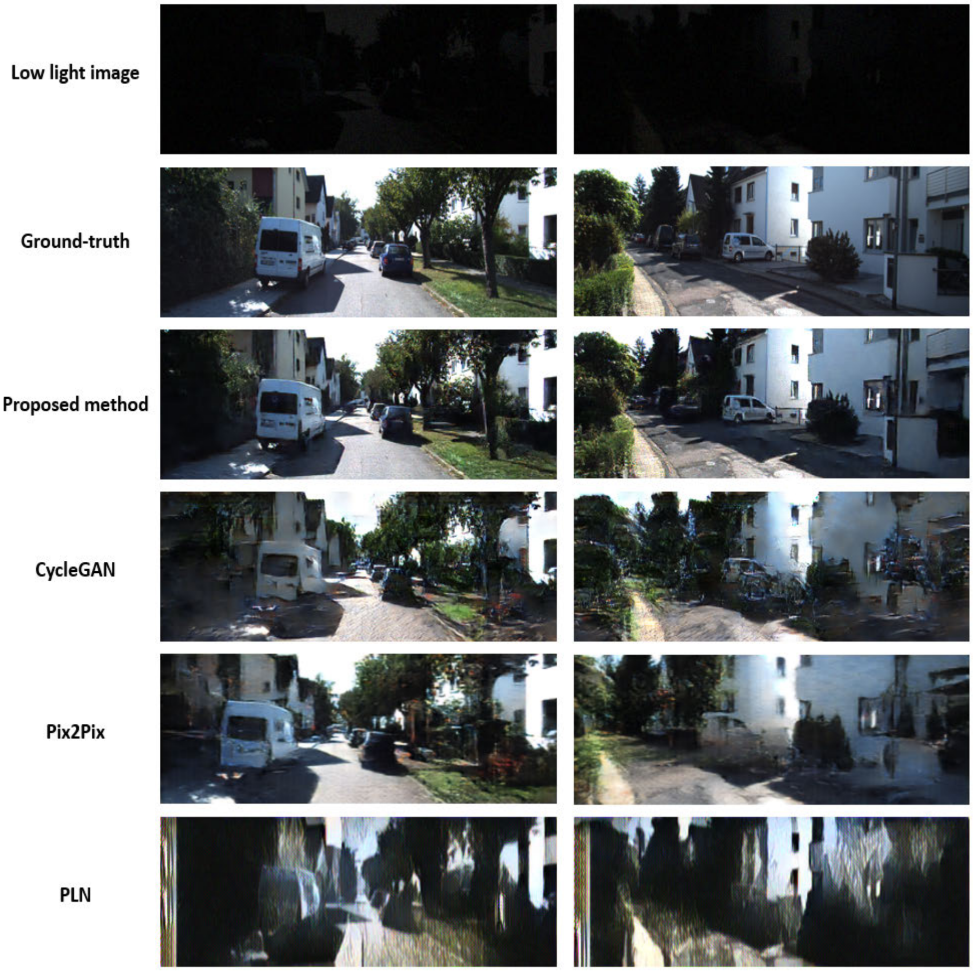


FIGURE 7. Examples of low light image enhancement results of the Syn-KITTI database. The first row shows the input images of the source domain, the second row shows each low light image and the accumulated ground-truth daytime images of the target domain, and the third row to the last row show the enhancement result images obtained by the proposed method, CycleGAN, Pix2Pix, and PLN, respectively.

TABLE 8. Semantic segmentation performances of the Syn-CamVid database before and after using the modified CycleGAN ("A" indicates no enhancement, and "B" indicates enhancement by the modified CycleGAN) (unit: %).

Methods	Pixel Acc		Class Acc		Mean IOU	
	A	B	A	B	A	B
FCN [1]	61.27	83.41	26.35	48.79	19.26	40.92
SegNet [2]	64.32	84.5	50.19	56.97	31.73	47.51
ICNet [30]	68.17	87.92	35.78	62.18	27.05	53.49
PSPNet [17]	74.65	88.58	37.17	63.97	29.11	55.31

without enhancement with each of the four segmentation networks.

Figure 9 shows examples of the segmentation results using the Syn-CamVid database, which were tested with models combining our modified CycleGAN with each segmentation network. The first and third columns show the results of testing the low light images without enhancement with each of the four segmentation networks. The second and fourth columns are the result of testing images enhanced with the modified CycleGAN. As shown in figure 9, when segmentation was performed directly on low light images without using the modified CycleGAN, the performance results were very poor. In contrast, when segmentation was performed by combining our modified CycleGAN on low light images, the resulting images were clearer and each object was well distinguished. Comparing the result images with the ground-truth images shows that the segmentation result with the enhancement applied is similar to the ground-truth image.

3) COMPARISONS OF LOW LIGHT IMAGE SEGMENTATION ACCORDING TO ENHANCEMENT NETWORK

In previous experiments, we confirmed that the segmentation performance in a low light environment was the highest when using the modified CycleGAN as the enhancement network and PSPNet as the segmentation network. Therefore, we set

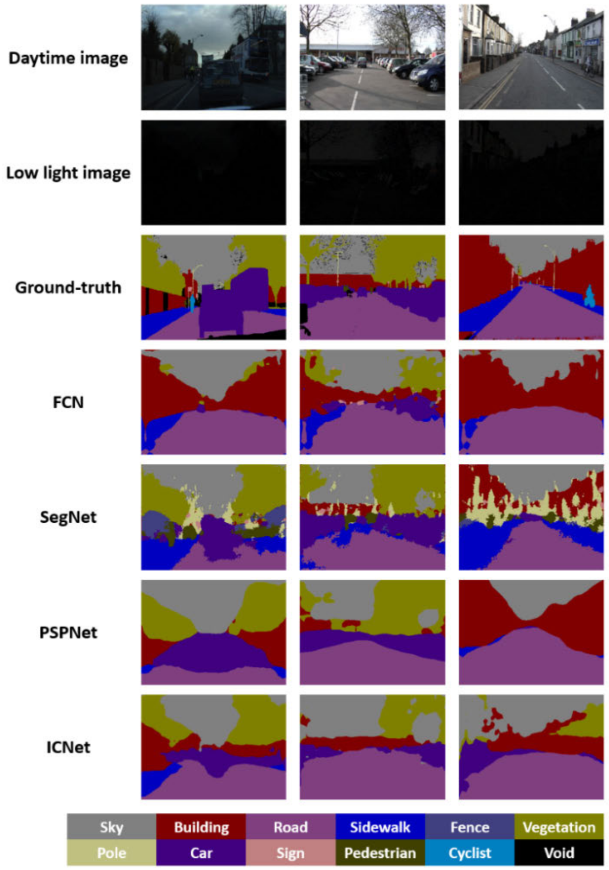


FIGURE 8. Examples of segmentation results of Syn-CamVid database without an enhancement method. The first row shows the ground-truth daytime images, the second row shows low light input images paired with each daytime image, the third row shows the ground-truth label images, and the fourth row to the last row show the segmentation result images of FCN, SegNet, PSPNet, and ICNet, respectively.

the model that combined the modified CycleGAN and PSPNet as our final model. In this section, the segmentation performance was compared using models combining four different enhancement networks and one segmentation network. For the enhancement network, Pix2Pix, PLN, CycleGAN, and modified CycleGAN were used, and for the segmentation network, PSPNet was used.

In Table 9, we compare the segmentation performance for the Syn-CamVid database using models combining PSPNet and four different enhancement networks. Among the three existing enhancement networks, the segmentation performance of PSPNet model combined with CycleGAN was higher. Accordingly, we can observe that CycleGAN shows superior performance in a low light image environment compared with Pix2Pix and PLN, and it can be confirmed that the performance can be further improved when combined with the segmentation network.

From the four models in Table 9, our final model combining the modified CycleGAN with PSPNet showed the highest performance. This indicates that our modified CycleGAN contributed more to improving segmentation performance than the original CycleGAN. Figure 10 shows an example of segmentation result images for the Syn-CamVid database obtained using PSPNet models combined with four different enhancement networks separately. Segmentation results using Pix2Pix, PLN, and CycleGAN as an enhancement network is quite good for large objects such as trees, buildings, and roads, but poor for smaller objects. In comparison, the resulting images of our final model showed high segmentation quality not only for large objects but also for small objects such as pedestrian or sign; these results are superior to those of other models. In addition, comparing with ground-truth images, the segmentation results obtained using our final model are the most similar to the ground-truth images.

As a final experiment with the Syn-CamVid database, we performed a t-test [47] for showing the significance of performance difference between our method and the second-best method (original CycleGAN [16] +PSPNet) as shown in Figure 11. In the null hypothesis for the t-test, it is assumed that there is no difference between the accuracy of our method and that of the second-best method. As shown in Figure 11, the p-values of Pixel Acc, Class Acc, and Mean IOU for this t-test were 9.74×10^{-3}, 8.29×10^{-3}, and 9.39×10^{-3} (less than 0.01), respectively. This shows that the null hypothesis is rejected at a 99% confidence level indicating that there is a significant difference at this confidence level between the performances (Pixel Acc, Class Acc, and Mean IOU) of our method and those of the second-best method. In addition, for analyzing the reliability of the observed phenomena in descriptive statistics, we performed the Cohen's d method [48], by which the size of the difference between the two models was demonstrated using the effect size [49]. It is calculated based on the average difference between the performance of our method and that of the second-best method, which is divided by standard deviation. Generally, effect size is classified as small, medium, and large defined by Cohen's d values of 0.2, 0.5, and 0.8 respectively. The experimental results in Figure 11 show the Cohen's d values of 14.2 (Pixel Acc), 15.4 (Class Acc), and 14.4 (Mean IOU). Because these Cohen's d values are close to 0.8, the results show that the differences between the performances of our method and those of the second-best one are large in effect size.

E. TESTING OF SEGMENTATION NETWORKS WITH SYNTHESIZED LOW LIGHT KITTI DATABASE

1) LOW LIGHT IMAGE SEGMENTATION WITHOUT ENHANCEMENT

In our comparative experiments using the Syn-CamVid database, we proved that the proposed method can significantly improve the segmentation performance in low light environments. In this section, segmentation performance was measured using the second open database, the Syn-KITTI database, in the same way as the experiment with the first database. As a first step, we measured how much the segmentation performance is dropped in the low light condition compared with the daytime environment.

TABLE 9. Comparisons of segmentation performances for the Syn-CamVid database according to different enhancement networks (PSPNet [17] is used as the base model for segmentation) (unit: %).

Methods	Pixel Acc	Class Acc	Mean IOU
Pix2Pix [28] + PSPNet	82.64	51.99	42.83
PLN [46] + PSPNet	84.31	53.30	44.70
Original CycleGAN[16] + PSPNet	85.09	57.96	48.49
Modified CycleGAN + PSPNet (proposed method)	88.58	63.97	55.31

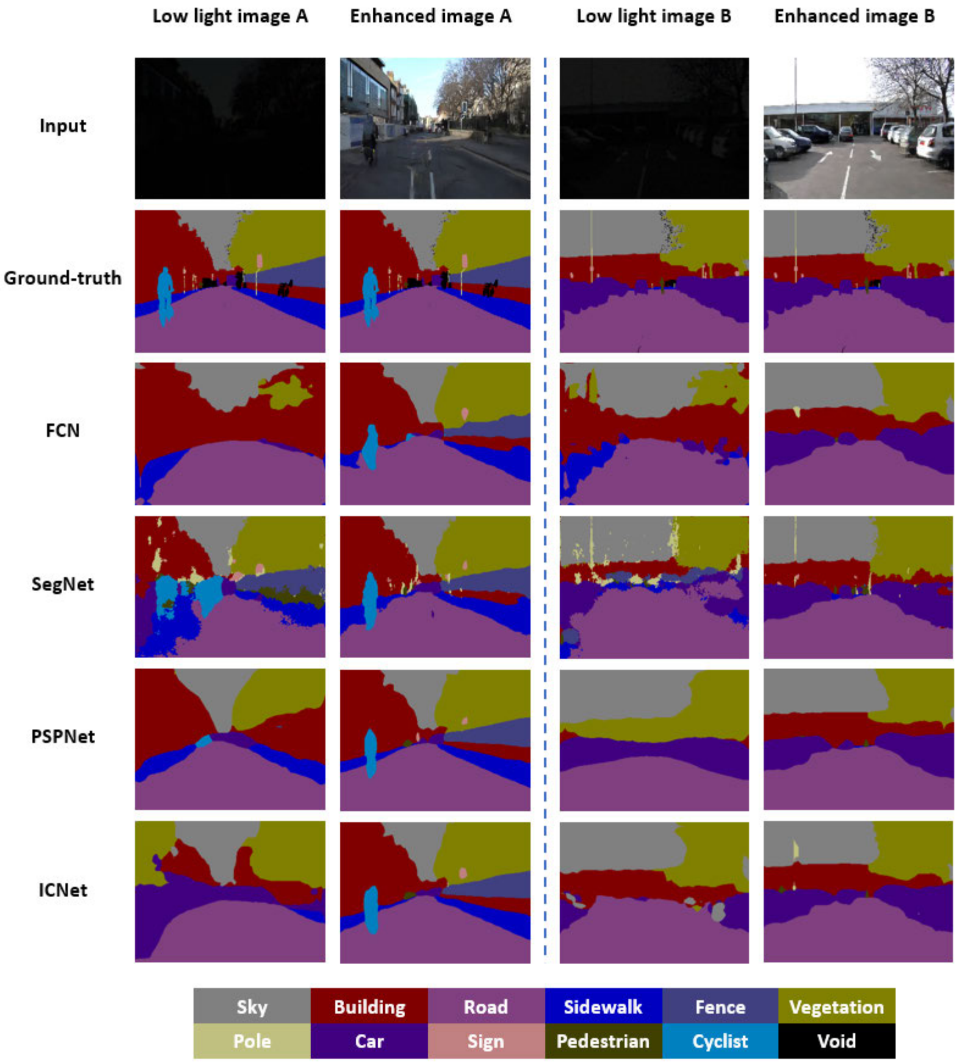


FIGURE 9. Examples of segmentation results of the Syn-CamVid database before and after using the modified CycleGAN.

Table 10 shows the segmentation performances measured for the daytime KITTI and Syn-KITTI databases using four segmentation networks. All the segmentation networks of Table 10 are trained from scratch with each database. As with the previous experiments using the CamVid database, all the four networks showed high segmentation performance for the daytime KITTI database. However, the segmentation performances for the low light KITTI database are significantly lower.

Figure 12 shows examples of segmentation result images tested with the four segmentation networks using the Syn-KITTI database. The images in the second row of Figure 12 are inputs to the segmentation networks and are the low light images belonging to the Syn-KITTI database.

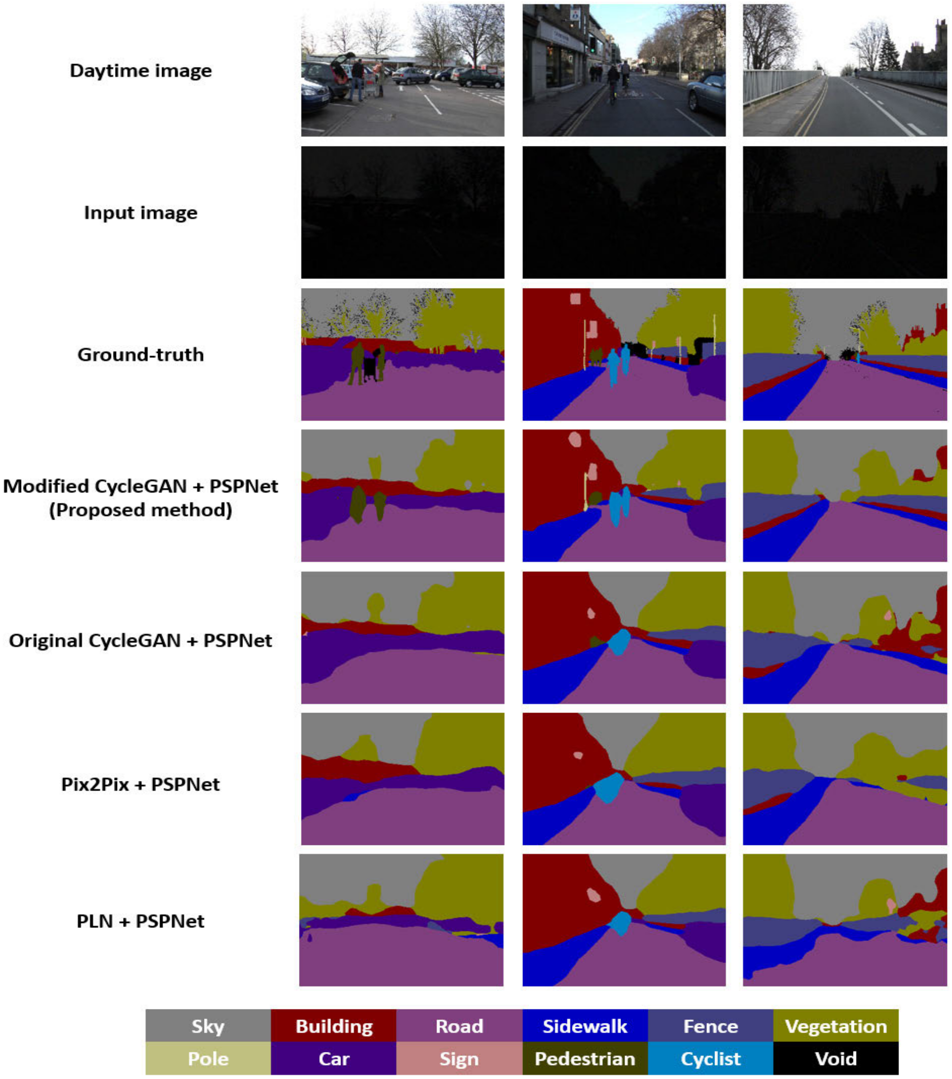
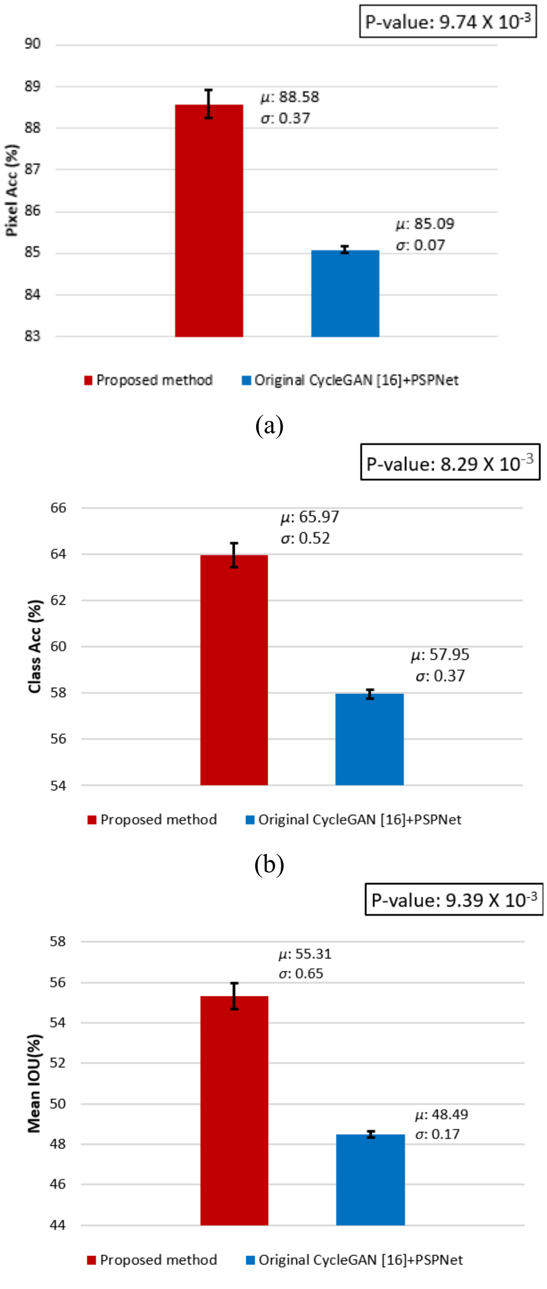


FIGURE 10. Examples of segmentation results for the Syn-CamVid database according to different enhancement networks (PSPNet [17] is used as the base model for segmentation).

TABLE 10. Comparisons of segmentation performances with the daytime KITTI and Syn-KITTI database (All networks are trained from scratch with each database) (unit: %).

Methods	Pixel Acc		Class Acc		Mean IOU	
	Daytime	Low light	Daytime	Low light	Daytime	Low light
FCN [1]	88.92	68.75	61.14	36.51	53.89	28.67
SegNet [2]	87.13	55.64	62.99	41.71	54.99	26.76
PSPNet [17]	87.66	57.24	68.59	32.15	55.92	21.64
ICNet [30]	81.39	54.5	56.94	36.01	44.3	22.77

Considering the segmentation result images of the four networks, we can observe that large objects with many pixel numbers such as sky, tree, and building show object shape to some extent, but small objects with small pixel numbers are not segmented. Furthermore, the boundaries between the objects are unclear overall and the shape information of the objects disappears. From the results in Table 10 and Figure 12, we observed that segmentation performance was significantly reduced in the low light environment compared with that in the daytime environment not only for the CamVid database but also for the KITTI database.



(a)

(b)

(c)

FIGURE 11. T-test performance of our method and the second-best method using the Syn-CamVid database. (a) Pixel Acc, (b) Class Acc, and (c) Mean IOU.

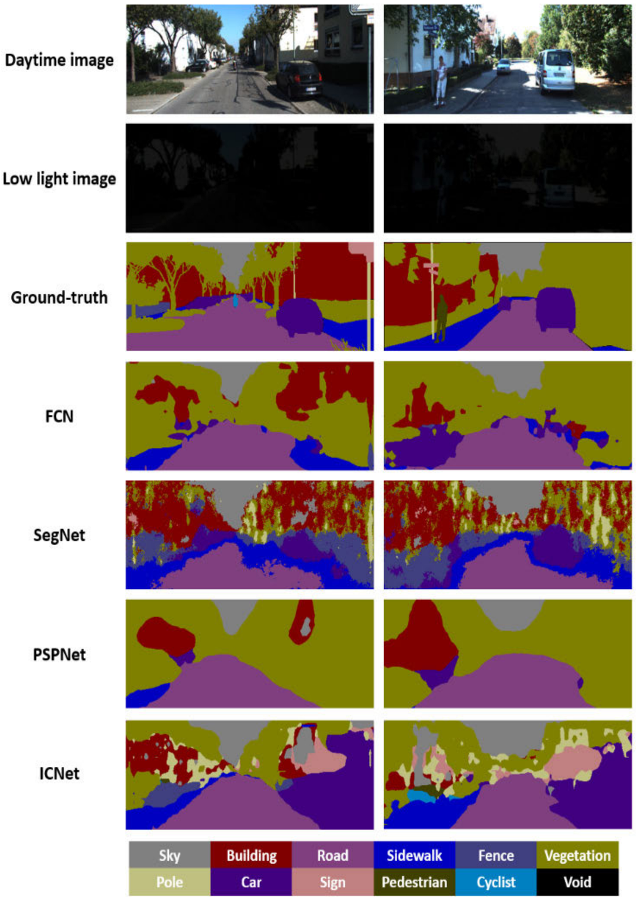


FIGURE 12. Examples of segmentation results of the Syn-KITTI database without an enhancement method. The first row shows the ground-truth daytime images, the second row shows low light input images paired with each daytime image, the third row shows the ground-truth label images, and the fourth row to the last row show the segmentation result images of FCN, SegNet, PSPNet, and ICNet, respectively.

2) ABLATION STUDY (LOW LIGHT IMAGE SEGMENTATION WITH ENHANCEMENT)

To improve the segmentation performance for the Syn-KITTI database, we used the modified CycleGAN, an enhancement network, in combination with four segmentation networks separately. Table 11 shows the measurement of segmentation performance before and after applying the enhancement technique using the modified CycleGAN. As shown in Table 11, the segmentation performance without enhancement was very low for all the evaluation indicator values. In contrast, the performance was greatly improved for all the segmentation networks after applying the enhancement technique using our modified CycleGAN. Especially, PSPNet showed the best performance among the four networks with a pixel accuracy of 80.48%, class accuracy of 62.9%, and mean IOU of 47.63%. In a comparative experiment using the second database, we proved that the modified CycleGAN not only showed good enhancement performance for the Syn-KITTI database, but also significantly improved segmentation performance. Figure 13 shows the examples of segmentation results using the Syn-KITTI database tested with the combined models.

The first and third columns are the results of testing the low light images without enhancement with each of the four segmentation networks. The second and fourth columns are the result of testing images enhanced with the modified CycleGAN. As shown in Figure 13, the segmentation performance was very low when the test was performed on low light images without the modified CycleGAN. In contrast,

TABLE 11. Segmentation performances of the Syn-KITTI database before and after using the modified CycleGAN ("A" indicates no enhancement, and "B" indicates the enhancement by the modified CycleGAN) (unit: %).

Methods	Pixel Acc		Class Acc		Mean IOU	
	A	B	A	B	A	B
FCN [1]	68.75	80.88	36.51	51.31	28.67	43.17
SegNet [2]	55.64	78.5	41.71	49.66	26.76	40.64
PSPNet [17]	57.24	80.48	32.15	62.9	21.64	47.63
ICNet [30]	54.5	76.01	36.01	51.61	22.77	39.3

TABLE 12. Comparisons of segmentation performances for the Syn-KITTI database according to different enhancement networks (PSPNet [17] is used as the base model for segmentation) (unit: %).

Methods	Pixel Acc	Class Acc	Mean IOU
Pix2Pix [28] + PSPNet	70.11	40.07	30.6
PLN [46] + PSPNet	71.61	41.13	32.55
Original CycleGAN [16] + PSPNet	74.33	42.88	34.89
Modified CycleGAN + PSPNet (proposed method)	80.48	62.9	47.63

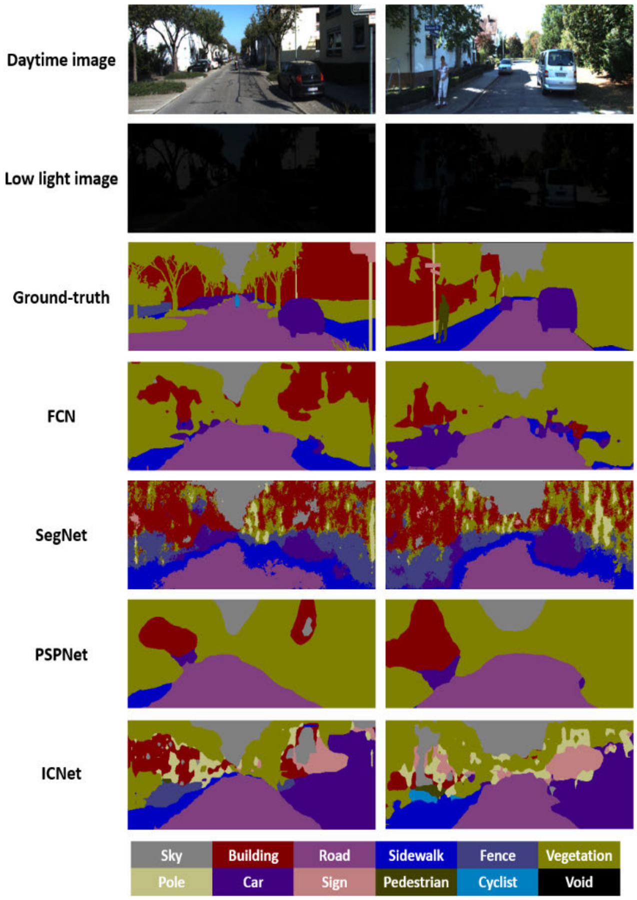


FIGURE 13. Examples of segmentation results of the Syn-KITTI database before and after using the modified CycleGAN.

TABLE 13. Comparisons of FLOPs and parameters with the proposed method and original CycleGAN.

	#Params	#FLOPs
Modified CycleGAN (proposed method)	2.04×10^{6}	26.78×10^{9}
Original CycleGAN[16]	11.38×10^{6}	116.45×10^{9}

when the test was performed by combining our modified CycleGAN and segmentation networks for low light images, the segmentation result images were clearer and more accurate. Comparing with the ground-truth image, the segmentation result images with enhancement are more similar to the ground-truth image.

3) COMPARISONS OF LOW LIGHT IMAGE SEGMENTATION ACCORDING TO ENHANCEMENT NETWORK

In this section, we compared the segmentation performance of the Syn-KITTI database using PSPNet models combined with four different enhancement networks separately: Pix2Pix, PLN, CycleGAN, and modified CycleGAN. Among the four models in Table 12, our final model, which combines the modified CycleGAN and PSPNet, showed the highest performance for all the evaluation indicators. This indicates that our modified CycleGAN contributed more to improving the segmentation performance than the previous methods.

Figure 14 shows an example of segmentation result images for the Syn-KITTI database obtained using PSPNet models combined with four different enhancement networks separately. Segmentation results using Pix2Pix, PLN, and CycleGAN as an enhancement network is quite good for large objects with many pixel numbers such as trees, buildings, and roads. However, the performance was poor for objects of small size with a small number of pixels, such as sign and pole. In comparison, the resulting images of our final model showed fairly high segmentation quality for small objects

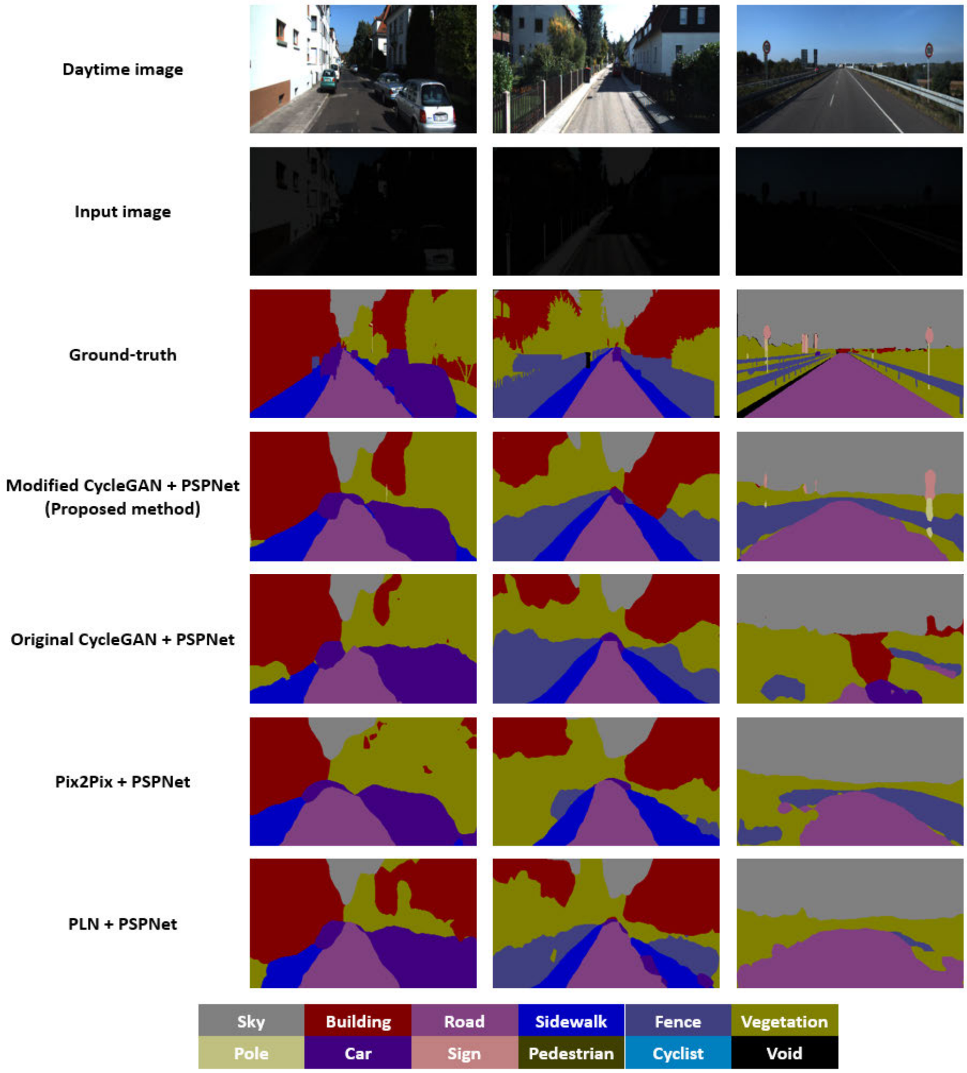
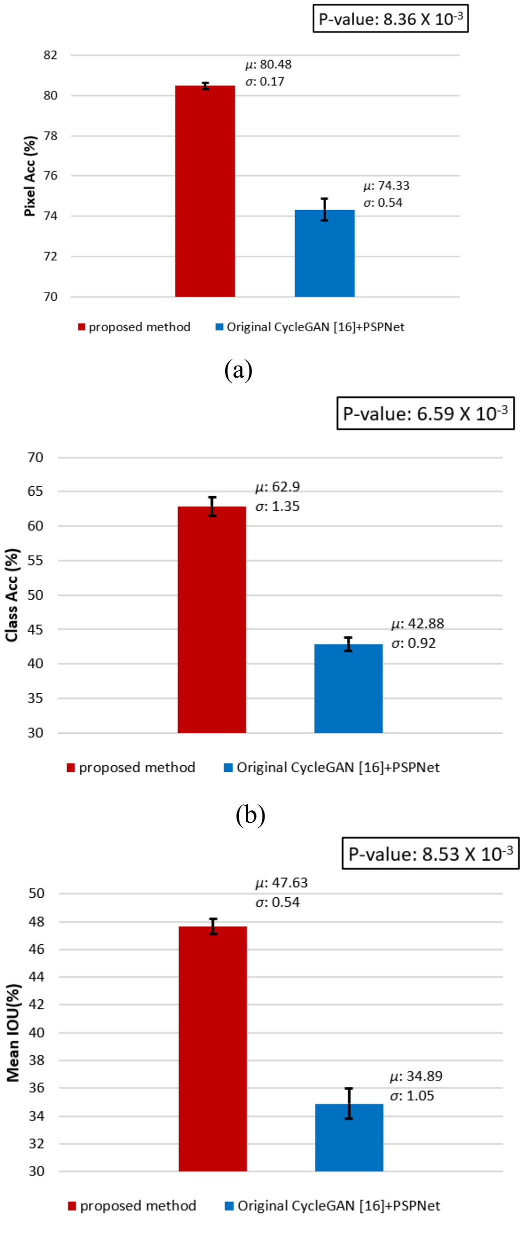


FIGURE 14. Examples of segmentation results for the Syn-KITTI database according to different enhancement networks (PSPNet [17] is used as the base model for segmentation).

such as poles or signs and for large objects, which indicates superior performance to those of other models. Table 12 and Figure 14 show that our final model has the highest segmentation performance for the Syn-KITTI database and produced the greatest performance improvement over the results of the existing enhancement networks. As a final experiment with the Syn-KITTI database, we performed a t-test for showing the significance of performance difference between our method and the second-best method (Original CycleGAN [16]+PSPNet). As shown in Figure 15, the p-values of Pixel Acc, Class Acc, and Mean IOU for this t-test were 8.36×10^{-3}, 6.59×10^{-3}, and 8.53×10^{-3} (less than 0.01), respectively. This shows that the null hypothesis is rejected at a 99% confidence level indicating that there is a significant difference at this confidence level between the performances (Pixel Acc, Class Acc, and Mean IOU) of our method and those of the second-best method. Furthermore, we performed the Cohen's d method. The experimental results in Figure 15 show the Cohen's d values of 15.3 (Pixel Acc), 17.3 (Class Acc) and 15.2 (Mean IOU). As these Cohen's d values are close to 0.8, the results show that the differences between the performances of our method and those of the second-best method are large in effect size.

F. COMPUTATIONAL COST AND PROCESSING TIME

In this section, computational cost and processing time are measured for our final model, and our modified CycleGAN and original CycleGAN are comparatively analyzed. For the comparison of the computational cost of the networks, we used floating-point operations (FLOPs) and parameters (Params) as the evaluation indicators. #FLOPs and #Params represent the total number of FLOPs and Params, respectively. The values of these two evaluation indicators were measured using the profiler library of TensorFlow framework (version 1.8.0) [40].



(a)

(b)

(c)

FIGURE 15. T-test performance of our method and the second-best method using the Syn-KITTI database. (a) Pixel Acc, (b) Class Acc, and (c) Mean IOU.

First, we compared the computational costs of the modified CycleGAN, the enhancement network of our final model, and the original CycleGAN. We reduced the computational cost of our network by modifying the residual blocks of the original CycleGAN into a bottleneck structure and reducing the number of the blocks from nine to six. To compare this numerically, Table 13 shows the number of FLOPS and parameters of the modified CycleGAN and the original CycleGAN. As shown in Table 13, our modified CycleGAN shows a reduction of approximately 5.58 times in terms of #Params and approximately 4.35 times in terms of #FLOPs compared with the original CycleGAN.

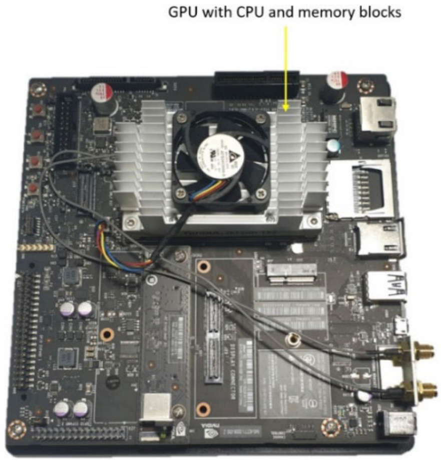


FIGURE 16. Jetson TX2 embedded system.

Second, we compared the average processing times of the modified CycleGAN and the original CycleGAN for the two databases. Table 14 shows the average processing time per image of our modified CycleGAN in a desktop environment. For the Syn-CamVid and Syn-KITTI databases, the average processing times per image were 34.63 ms and 38.66 ms, respectively. The processing time was reduced by approximately 8 to 14 ms compared with that of the original CycleGAN. As the next experiment, the average processing time was measured in the Jetson TX2 embedded system [50], which is widely used for on-board deep learning processing as shown in Figure 16. Jetson TX2 has an NVIDIA Pascal^{TM}-family GPU (256 CUDA cores), having 8 GB of memory shared between the central processing unit (CPU) and GPU, and 59.7 GB/s of memory bandwidth; it uses less than 7.5 W of power. As shown in Table 14, for the Syn-CamVid and Syn-KITTI databases, the average processing times per image were 200.97 ms and 238.93 ms, respectively. The processing time was reduced by approximately 157 to 202 ms compared with that of the original CycleGAN.

As the final experiment, we measured the average processing time of our final model for the two databases. Table 15 shows the average processing time per image measured using our proposed method for the Syn-CamVid database. The average processing time per image was 149.71 ms in the desktop environment and 1172.61 ms in the Jetson TX2 embedded system. Table 16 shows the average processing time per image measured using our proposed method for the Syn-KITTI database. The average processing time per image was 186.26 ms in the desktop environment and 1350.23 ms in the Jetson TX2 embedded system. The average processing time using the Jetson TX2 embedded system is longer than that using the desktop computer due to its limited computing resources. However, this result shows that our

TABLE 14. Comparisons of average processing time for the proposed method and original CycleGAN (unit: ms).

	Database	Original CycleGAN [16]	Modified CycleGAN
Desktop computer	Syn-CamVid	46.86	34.63
	Syn-KITTI	52.38	38.66
Jetson TX2 embedded system	Syn-CamVid	357.68	200.97
	Syn-KITTI	440.95	238.93

TABLE 15. Average processing time of the proposed method using the Syn-CamVid database (unit: ms).

	Modified CycleGAN	PSPNet	Total
Desktop computer	34.63	115.08	149.71
Jetson TX2 embedded system	200.97	971.64	1172.61

TABLE 16. Average processing time of the proposed method using the Syn-KITTI database (unit: ms).

	Modified CycleGAN	PSPNet	Total
Desktop computer	38.66	147.6	186.26
Jetson TX2 embedded system	238.93	1111.3	1350.23

method is also applicable to embedded systems with limited computing resources.

VI. CONCLUSION

In this study, we discussed semantic segmentation methods in low light environments, which have not been studied extensively so far. Unlike most of the prior studies, we proposed an enhancement-based multi-class segmentation method in a low light environment. The proposed model is divided into two parts: an enhancement network and a segmentation network. For the enhancement network, we used a new network that modified the original CycleGAN for low light image enhancement. The proposed modified CycleGAN modified the generator structure of the original CycleGAN and added a paired L1 loss. This increased the enhancement quality and reduced the computational cost and processing time compared with those of the original CycleGAN. Comparative experiments using synthesized low light CamVid and KITTI databases showed that our modified CycleGAN had the best low light image enhancement quality compared with other existing enhancement networks. In addition, using a model combining the proposed modified CycleGAN and state-of-the-art segmentation networks, we demonstrated that segmentation performance can be significantly improved in low light environments. Among those, the proposed final model that combined the modified CycleGAN and PSPNet showed the best segmentation performance.

In future, we plan to increase the segmentation performance in the nighttime environment to a similar level as the segmentation performance in the daytime environment. In training process of the proposed enhancement network, we would improve the segmentation performance by using the output of segmentation network as the value of the loss function. In addition, we would design a segmentation model that can operate in real time in nighttime environment by applying various model pruning techniques.

REFERENCES

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

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

[3] Y. Wang and J. Ren, "Low-light forest flame image segmentation based on color features," *J. Phys. Conf. Ser.*, vol. 1069, no. 1, Aug. 2018, Art. no. 012165.

[4] V. Haltakov, J. Mayr, C. Unger, and S. Ilic, "Semantic segmentation based traffic light detection at day and at night," in *Proc. German Conf. Pattern Recognit.*, Aachen, Germany, Oct. 2015, pp. 446–457.

[5] O. Alpar, "Corona segmentation for nighttime brake light detection," *IET Intell. Transp. Syst.*, vol. 10, no. 2, pp. 97–105, Mar. 2016.

[6] T. Soumya, "A moving object segmentation method for low illumination night videos," in *Proc. World Congr. Eng. Comput. Sci.*, San Francisco, CA, USA, Oct. 2008, pp. 1–5.

[7] S.-H. Lee, S. I. Han, and M. G. Kang, "Object detection in low illumination environment," in *Proc. Int. Conf. Comput. Graphs. Vis. Comput. Vis.*, Pilsen, Czech Republic, Feb. 2009, pp. 33–38.

[8] J. Li, F. Zhang, L. Wei, T. Yang, and Z. Lu, "Nighttime foreground pedestrian detection based on three-dimensional voxel surface model," *Sensors*, vol. 17, no. 10, p. 2354, Oct. 2017.

[9] T. Kancharla, P. Kharade, S. Gindi, K. Kutty, and V. G. Vaidya, "Edge based segmentation for pedestrian detection using NIR camera," in *Proc. Int. Conf. Image Inf. Process.*, Shimla, India, Nov. 2011, pp. 1–6.

[10] M. Akther, M. K. Ahmed, and M. Z. Hasan, "Detection of Vehicle's number plate at nighttime using iterative threshold segmentation (ITS) algorithm," *Int. J. Image, Graph. Signal Process.*, vol. 5, no. 12, pp. 62–70, Oct. 2013.

[11] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Nighttime sky/cloud image segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 345–349.

[12] D. Dai and L. V. Gool, "Dark model adaptation: Semantic image segmentation from daytime to nighttime," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Maui, HI, USA, Nov. 2018, pp. 3819–3824.

[13] C. Sakaridis, D. Dai, and L. Van Gool, "Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 7374–7383.

[14] L. Sun, K. Wang, K. Yang, and K. Xiang, "See clearer at night: Towards robust nighttime semantic segmentation through day-night image conversion," 2019, *arXiv:1908.05868*. [Online]. Available: http://arxiv.org/abs/1908.05868

[15] A. Valada, J. Vertens, A. Dhall, and W. Burgard, "AdapNet: Adaptive semantic segmentation in adverse environmental conditions," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2017, pp. 4644–4651.

[16] 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)*, Venice, Italy, Oct. 2017, pp. 2223–2232.

[17] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2881–2890.

[18] G. J. Brostow, J. Fauqueur, and R. Cipolla, ''Semantic object classes in video: A high-definition ground truth database,'' *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, Jan. 2009.

[19] 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.

[20] P.-R. Chen, H.-M. Hang, S.-W. Chan, and J.-J. Lin, ''DSNet: An efficient CNN for road scene segmentation,'' 2019, *arXiv:1904.05022*. [Online]. Available: http://arxiv.org/abs/1904.05022

[21] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, ''Dual attention network for scene segmentation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 3146–3154.

[22] M. Arsalan, D. S. Kim, M. B. Lee, M. Owais, and K. R. Park, ''FRED-net: Fully residual encoder–decoder network for accurate iris segmentation,'' *Expert Syst. Appl.*, vol. 122, pp. 217–241, May 2019.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4700–4708.

[24] G. Lin, A. Milan, C. Shen, and I. Reid, ''RefineNet: Multi-path refinement networks for high-resolution semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1925–1934.

[25] *Synthesized Low Light Cambridge-driving Labeled Video Database (Syn-CamVid), Synthesized Low Light Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago (Syn-KITTI) Database, and Algorithm Including CNN Models*. Accessed: Jan. 10, 2020. [Online]. Available: http://dm.dgu.edu/link.html

[26] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ''Generative adversarial nets,'' in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014, pp. 1–9.

[27] A. L. Maas, A. Y. Hannun, and A. Y. Ng, ''Rectifier nonlinearities improve neural network acoustic models,'' in *Proc. Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, pp. 1–6.

[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)*, Honolulu, HI, USA, Jul. 2017, pp. 1125–1134.

[29] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, ''Least squares generative adversarial networks,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2794–2802.

[30] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, ''ICNet for real-time semantic segmentation on high-resolution images,'' in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, Sep. 2018, pp. 405–420.

[31] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, May 2015, pp. 1–14.

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

[33] I. Krešo, D. Čaušević, J. Krapac, and S. Šegvić, ''Convolutional scale invariance for semantic segmentation,'' in *Proc. German Conf. Pattern Recognit.*, Hannover, Germany, Sep. 2016, pp. 64–75.

[34] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, ''BDD100K: A diverse driving dataset for heterogeneous multitask learning,'' May 2018, *arXiv:1805.04687*. [Online]. Available: http://arxiv.org/abs/1805.04687

[35] M. J. Milford and G. F. Wyeth, ''SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights,'' in *Proc. IEEE Int. Conf. Robot. Autom.*, Saint Paul, MN, USA, May 2012, pp. 1643–1649.

[36] Y. P. Loh, X. Liang, and C. S. Chan, ''Low-light image enhancement using Gaussian process for features retrieval,'' *Signal Process., Image Commun.*, vol. 74, pp. 175–190, May 2019.

[37] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, ''MSR-net: Low-light image enhancement using deep convolutional network,'' Nov. 2017, *arXiv:1711.02488*. [Online]. Available: http://arxiv.org/abs/1711.02488

[38] K. G. Lore, A. Akintayo, and S. Sarkar, ''LLNet: A deep autoencoder approach to natural low-light image enhancement,'' *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[39] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.

[40] *TensorFlow*. Accessed: Apr. 11, 2019. [Online]. Available: https://www.tensorflow.org/

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

[42] *NVIDIA GeForce GTX 1080*. Accessed: Dec. 10, 2019. [Online]. Available: https://www.geforce.com/hardware/desktop-gpus/geforce-gtx-1080/specifications

[43] T. Stathaki, *Image Algorithms and Applications*. Cambridge, MA, USA: Academic, 2008.

[44] D. Salomon, *Data Compression: The Complete Reference*, 4th ed. New York, NY, USA: Springer-Verlag, 2006.

[45] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, ''Image quality assessment: From error visibility to structural similarity,'' *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[46] J. Johnson, A. Alahi, and L. Fei-Fei, ''Perceptual losses for real-time style transfer and super-resolution,'' in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 694–711.

[47] *Student's T-Test*. Accessed: Dec. 10, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Student%27s_t-test

[48] J. Cohen, ''A power primer,'' *Psychol. Bull.*, vol. 112, no. 1, p. 155, Jul. 1992.

[49] S. Nakagawa and I. C. Cuthill, ''Effect size, confidence interval and statistical significance: A practical guide for biologists,'' *Biol. Rev.*, vol. 82, no. 4, pp. 591–605, Nov. 2007.

[50] *Jetson TX2 Module*. Accessed: Dec. 10, 2019. [Online]. Available: https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/

SE WOON CHO received the B.S. degree in electronics and electrical engineering from Dongguk University, Seoul, South Korea, in 2017, where he is currently pursuing the combined course of M.S. and Ph.D. degrees in electronics and electrical engineering. He designed the semantic segmentation system in low light environments-based on convolutional neural networks, analyzed results of experiments, and wrote the original article. His research interests include biometrics and pattern recognition.

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

JA HYUNG KOO received the B.S. degree in electronics and electrical engineering from Dongguk University, Seoul, South Korea, in 2016, where he is currently pursuing the combined course of M.S. and Ph.D. degree in electronics and electrical engineering. He helped the experiments and analysis. His research interests include biometrics and pattern recognition.

MUHAMMAD ARSALAN received the B.S. degree in computer engineering from COMSATS University Islamabad, Pakistan, in 2012, and the M.S. degree in computer science from NCBA&E, Lahore, Pakistan, in 2016. He is currently pursuing the Ph.D. degree in electronics and electrical engineering with Dongguk University, Seoul, South Korea. He helped the implementation of modified CycleGAN. His research interests include computer vision and deep learning.

KANG RYOUNG PARK received the B.S. and M.S. degrees in electronic engineering and the Ph.D. degree in electrical and computer engineering from Yonsei University, Seoul, South Korea, in 1994, 1996, and 2000, respectively. Since March 2013, he has been a Professor with the Division of Electronics and Electrical Engineering, Dongguk University. He supervised this research and revised the original article. His research interests include image processing and biometrics.