IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 3, June 2025, pp. 2347∼2357
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i3.pp2347-2357

CycleGAN for day-to-night image translation: a comparative study

Muhammad Feriansyah Raihan Taufiq, Laksmita Rahadianti
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Article Info

Article history:

Received Oct 2, 2024
Revised Feb 3, 2025
Accepted Mar 15, 2025

Keywords:

CycleGAN
Deep learning
Image processing
Image translation
Synthetic nighttime images

ABSTRACT

Computer vision tasks often fail when applied to night images, because the models are usually trained using clear daytime images only. This creates the need to augment the data with more nighttime image for training to increase robustness. In this study, we consider day-to-night image translation using both traditional image processing approaches and deep learning models. This study employs a hybrid framework of traditional image processing followed by a CycleGAN-based deep learning model for day-to-night image translation. We then conduct a comparative study on various generator architectures in our CycleGAN model. This research compares four different CycleGAN models; i.e., the orginal CycleGAN, feature pyramid network (FPN) based CycleGAN, the original U-Net vision transformer based UVCGAN, plus a modified UVCGAN with additional edge loss. The experimental results show that the orginal UVCGAN obtains an Fréchet inception distance (FID) score of 16.68 and structural similarity index measure (SSIM) of 0.42, leading in terms of FID. Meanwhile, FPN-CycleGAN obtains an FID score of 104.46 and SSIM score of 0.44, leading in terms of SSIM. Considering FPN-CycleGAN's bad FID score and visual observation, we conclude that UVCGAN is more effective in generating synthetic nighttime images.

Corresponding Author:

Laksmita Rahadianti
Faculty of Computer Science, Universitas Indonesia
UI Depok Campus, Depok, Indonesia
Email: laksmita@cs.ui.ac.id

1. INTRODUCTION

Computer vision tasks often perform poorly on night images due to lack of night representation in the training dataset, as clear daytime images are more often used during training [1]. This creates the need for more diverse training data to increase the robustness of deep computer vision models, including nighttime images. Building night image datasets by image capture is a tedious and time consuming task, requiring careful setup to ensure alignment when capturing identical source and target images. This process is made nearly impossible due to the variable conditions that influence the image, such as location, weather conditions, and time of day that will affect components on the image [2]. Several previous studies have used simulators to augment data, with the potential to generate unlimited data for any given scenario [3]. However, the resulting images have a unnatural appearance and does not represent real world scenarios [4]. Thus, alternate ways of generating data is needed.

Image-to-image translation models can be trained to learn the transformation from one image domain to another, which can be exploited to handle tasks such as style transfer, data augmentation, or image restoration. Thus, this technique can also be utilized to augment training data for deep learning models. Image-to-image translation with a supervised deep learning model requires corresponding image pairs from both the source and target domains for a supervised model. Assuming paired images are available, deep learning based methods can create transfer functions by translating image on one domain to another domain by its latent representation. One such architecture is generative adversarial networks (GAN) such as CycleGAN [2].

Nevertheless, we will still require some data to start on any type of training, for which we turn to more traditional image processing methods. Punnappurath [5] proposes a method to create a paired dataset using traditional image processing method to create a synthetic nighttime image from a daytime image input. This can provide a smaller paired dataset to start on a conditional GAN or semi-supervised GAN. This traditional image processing method proves that it can synthesize a nighttime image with various parameters, which is improved further by using other image processing techniques [6]. To ensure the quality of the generated synthetic nighttime images, they must be evaluated using no-reference image quality assessment (NR-IQA) [7]–[10]. When the results are satisfactory, these images can then be used to train GAN models.

There are usually two types of GAN learning tasks, i.e., conditional GAN [11] and unconditional GAN [12]. A conditional GAN model learns generative models based on input images and their corresponding labels similarly to supervised learning [13], while unconditional GAN introduced the contrastive representation learning mechanism [14] and can be considered as unsupervised learning technique. On this research we carried out a semi-supervised GAN [15], namely mixing paired and unpaired images. This GAN model is trained using both a unpaired public dataset as well as our own generated paired dataset which are used together to train our CycleGAN methods [12]. Through this research, we aim to have a better understanding on how to synthesize nighttime image based on traditional image processing method, also utilizing CycleGAN algorithm to perform day-to-night image-to-image translation.

The highlights of the contribution of this research work is as follows:

- This research considers a hybrid method of traditional image processing and a deep learning model, in this case CycleGAN to establish day-to-night image translation.
- We conduct a comparative study of CycleGAN models with varying generator architectures to conduct day-to-night image translation, i.e., original CycleGAN, feature pyramid network (FPN)-CycleGAN (FPN-CycleGAN), and U-Net vision transformer CycleGAN (UVCGAN).
- We include a modified UVCGAN model, with an additional edge feature loss function in the comparison.

This paper is structured into five sections. The first section serves as the background of the study, discussing its background and objectives. The second section is a literature review that examines research related to the topic. The third section contains the methods used in this research. The fourth section presents the results of the results and the analysis of results. The paper is then ended with a conclusion, encompassing the research results and recommendations for future research.

2. LITERATURE REVIEW

Daytime images are generally easy to capture due to sufficient lighting, while nighttime images are challenging because of low light, which can lead to noise and blur [5]. To create high-quality paired day and night images, it requires the ground truth from the target image to have the same scenery as the input daytime image. Capturing a pair of day and night images requires careful setup to ensure alignment between the image pairs and consumes a lot of time. Therefore we require an alternative method to generate synthetic nighttime images from daytime inputs. Recent research in image-to-image translation, particularly for day-to-night image transformation, has introduced two main approaches: traditional image processing and deep learning.

There are several traditional image processing methods for generating and improving the quality of nighttime images. One proposed method from [5] generates synthetic nighttime images from daytime images. Another approach improves the quality of nighttime images by applying processing techniques such as contrast enhancement, sharpening, and other methods [6]. Additionally, the discrete cosine transform (DCT) can be leveraged to analyze image-level frequency distributions by transforming spatial features into the frequency domain [16]. The DCT helps in understanding the differences in frequency distributions between daytime and nighttime images. A novel method for enhancing low-illumination images aims to improve the visibility of details in dark regions while minimizing artifacts [17]. This method provides clearer details in dark areas while maintaining overall image quality, utilizing a guided filter to refine the transmission map, thus balancing computational efficiency and structural preservation.

In this research we focus on two traditional methods. A proposed method by [5] are used to generate nighttime images from an input of daytime images. This process involves several steps to mimic the appearance of real nighttime scenes. First, the exposure of the daytime image is reduced to simulate the lower light levels typical of night. Next, the scene is relighted using night-specific illuminants, which requires calculating the average brightness of nighttime lighting and applying it to the image. Noise is then added to replicate the visual characteristics of a photo taken in low light. To obtain a visually pleasing of nighttime scene, we should do some image processing on the synthetic nighttime images as done by [6]. Further processing includes RAW image processing, where the image undergoes steps like demosaicing (to convert the image from a sensor's RAW format to a full-color image), white balancing based on Gray world algorithm [18], and denoising to reduce any added or inherent noise. The image is then enhanced for contrast and sharpness to improve its overall quality. These methods aim to create synthetic nighttime images that closely resemble real ones, though they may still differ in quality from actual nighttime photos.

Additionally, we may utilise deep image-to-image translation task to perform day-to-night image translation. A method for translating nighttime images to daytime images using a conditional GAN is presented in [19]. This method employs two generator-discriminator pairs: one for converting night images to day images and another for the reverse process. Another proposed method introduces two-phase consistency network (2PCNet), a model for day-to-night unsupervised domain adaptive object detection [20], which employs a two-phase consistency training approach to mitigate error propagation issues commonly found in student-teacher frameworks. A semantic-aware image-to-image translation method specifically designed for day-to-night image conversion is also proposed [1]. This approach utilizes the encoder of a pre-trained semantic segmentation network to ensure semantic consistency during the translation process, requiring only segmentation predictions from the source domain, making it more practical. Another application of nighttime imagery is detecting parking space status at night by leveraging daytime data [21]. The framework consists of content encoders and specific encoders that learn both domain-invariant and discriminative features.

GAN [22] has presented remarkable outcomes in many applications, such as object detection, style transfer, image synthesis, medical image analysis, etc. The GAN algorithm has been widely used for tasks like day-to-night image translation. GANs consist of two networks: a generator that creates synthetic images and a discriminator that distinguishes between real and fake images. CycleGAN, a specific type of GAN, is particularly effective for day-to-night translation [12]. It learns mapping functions between day and night images using a cycle consistency loss, ensuring that translating an image from one domain to another and back again results in the original image. This research enhances the CycleGAN model by integrating FPN [23] and UVCGAN [24]. FPN improves feature extraction across multiple scales, enhancing object detection and segmentation. UVCGAN combines the strengths of UNet for high-frequency feature extraction with vision transformer (ViT) for capturing low-frequency features, leading to better image translation. The combined model ensures that the output images retain the same dimensions as the input images, providing high-quality results for day-to-night translation.

The concept of edge loss is introduced through the asymmetric CycleGAN aimed at improving unpaired and asymmetric image translation tasks [25]. This method introduces a quantitative metric to determine whether the translation task is symmetric or asymmetric and employs different sizes of generators to adapt to these asymmetric translations. A key feature of this method is the introduction of an edge-retain loss, which enhances the quality of generated images by retaining necessary structural details, particularly edges, from the input images. This edge-retain loss is computed using a pre-trained edge detector and is combined with the original CycleGAN loss to form the total loss function. The results demonstrate that the asymmetric CycleGAN with edge-retain loss significantly improves the visual quality of generated images compared to the original CycleGAN and symmetric CycleGAN on various tasks.

3. METHODOLOGY

This study aims to generate synthetic nighttime images for an augmented dataset. This is accomplished by modelling day-to-night image translation using CycleGAN. We then conduct a comparative study on various generator architectures in CycleGAN for the best result.

3.1. Dataset

In order to train the CycleGAN, we need a dataset of day-to-night image pairs to train the model. Since we have established that these pairs are difficult to obtain, in this research we use two sources of day and night images, a synthetic nighttime dataset and the Berkeley deep drive (BDD) dataset [26]. The synthetic nighttime dataset used is generated by our proposed framework as described in section 3.1.1.

3.1.1. Synthetic nighttime dataset

The initial dataset used in this study is from [5]. This dataset contains a total of 70 daytime images in raw format (.dng). The spatial resolution of each image is 3024×4032 pixels in RGB format. Most of the images consist of outdoor and indoor scenes. The outdoor scenes are captured under street lighting, while the indoor scenes are captured under regular indoor illumination. For building the night illumination dictionary we need a graycard image under different night illuminations. This night illumination dictionary is important to relight the synthetic nighttime image with night illumination. An example of the dataset is shown on Figures 1(a) and 1(b).

(a) (b)

Figure 1. Dataset overview: (a) daytime images and (b) images of gray card

The synthetic images are generated using a framework of two steps, i.e., the synthetic nighttime image generation step and nighttime image enhancement step. This two-step framework is illustrated in Figure 2. The synthetic nighttime image generation step models a lower exposure and generates a corresponding dark image. The output of this step becomes the input for the nighttime image enhancement step. The nighttime image enhancement step improves the nighttime quality of the overall image.

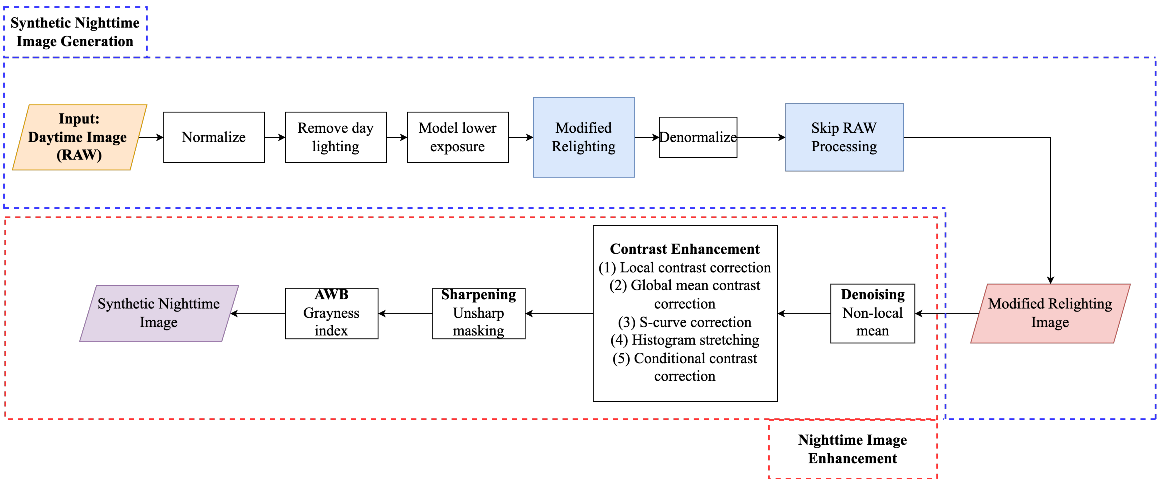

Figure 2. Framework for generating the synthetic nighttime dataset

The synthetic nighttime image generation step mostly adapts from [5], with a modified relighting step. This modification is required because default parameter in the source code generates image that differs from the one shown in [5]. In our modification, we try to generate random night illuminant position. The night illuminants are obtained by randomly sampling from a 2D multivariate Gaussian distribution of joint chromaticity values that fit around database of night illuminants, i.e from images of gray card on Figure 1(b). To obtain the synthetic nighttime image we loop this sample multiple times until the color is balanced. We keep this fixed sample and adjust the color of the lights that have been obtained from the modeled 2D multivariate Gaussian. According to Punnappurath *et al.* [5], a total of seven local lights are generated, with the first one being ambient background lighting. On this framework with modified relighting, we only use the ambient background lighting due to the fact that the generated light with local light does not look natural. Hence, resulting a modified relighting image.

The nighttime image enhancement step mostly adapts from [6], however, we do not modify the image processing pipeline. In the second framework there are four main processes, namely, denoising denotes removing the image noises, contrast enhancement is for enhancing the image contrast, unsharp masking is for sharpening the image caused by the denoising process, and lastly AWB (grayness index) [27] is for measuring and controlling how close the white balance is to the true gray value under certain conditions. We don't modify the algorithm because the purpose of this framework is to create a visually pleasing nighttime image. However, there are several parameters that need to be tuned. For example, the β value from local contrast correction (LCC) [28] can be tuned for research purposes to see how it affects the image.

The results of our synthetic nighttime image is shown in Figure 3. To ensure the results of the generated dataset is satisfactory, we evaluate each generated image using NR-IQA metrics, i.e. Fréchet inception distance (FID) and structural similarity index measure (SSIM) score. Each of these metrics aim to objectively score the image quality. More detail on these metrics are discussed on section 3.3. The modified relighting image is illuminated by ambient background and randomize night illumination. Hence resulting a different night illumination on specific output. The local lights from [5] are fixed, and the artificial light does not look visually pleasing. Later, it will affect the CycleGAN mapping function. We use only the modified relighting image output for the synthetic nighttime dataset, as the other output type makes the image appear too blue and unnatural. Furthermore, it is quantitatively proven that the modified relighting image has better quality compared to the other images. For the modified relighting image, the FID and SSIM scores are 5.2 and 0.4, respectively. Meanwhile, the contrast enhancement image has the second-best quality, with FID and SSIM scores of 6.7 and 0.3, respectively.

Figure 3. The generated synthetic nighttime image

3.1.2. Berkeley deep drive translation dataset

The BDD dataset contains unpaired day and night images with a spatial resolution of 1280×720 pixels in RGB [26]. One of the main reasons for using BDD dataset is to see how it generates traffic lights and street lamps. Also, we combine BDD dataset and the resulted synthetic nighttime image to test if the learned features from deep learning model affects the night aspect of synthethic nighttime image from traditional image processing methods. Later, we perform a semi-supervised GAN by which it need both a paired and unpaired datasets. The paired datasets of day and night images (resulted synthetic nighttime image), while the unpaired datasets comes from BDD dataset and its also day and night images.

3.2. Experimental design

The experimental design of this research consists of performing deep image-to-image translation using a semi-supervised CycleGAN. Semi-supervised GAN uses both paired and unpaired images. For paired data, we used the synthetic nighttime dataset as described in section 3.1.1. For unpaired data, we use the BDD dataset described in section 3.1.2. This results in 70 paired images and 100 unpaired images, respectively, for utilizing semi-supervised CycleGAN is illustrated on Figure 4.

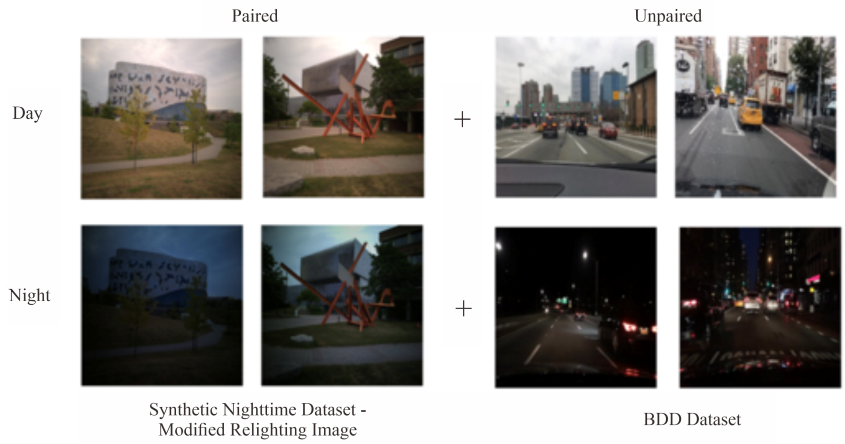

Figure 4. An overview of the paired & unpaired datasets

The training process of the semi-supervised CycleGAN model involves feeding each image domain with its corresponding input. For the daytime domain, we feed both the generator and the discriminator of the daytime image. However, we feed the synthetic nighttime image to the day-to-night generator, and it acts as a label for the paired daytime image. The forward cycle-consistency is required for the model to do identity mapping. Identity mapping refers to how the model translates one image domain to another domain while keeping the color information. By doing this, both the daytime image and the reconstructed daytime image should have the same scenery. The overall process on the nighttime domain is the same as on the daytime domain.

3.2.1. CycleGAN for day-to-night image translation

Several deep learning CycleGAN model is built with different generator architectures to see how well it performs on generating nighttime images. In this study we compare 4 different CycleGAN architectures, i.e., original CycleGAN, FPN-CycleGAN, and UVCGAN. Additionally, we also introduce a new combination of UVCGAN with edge feature loss that will be explained more in section 3.2.2. All compared architectures are listed in Table 1.

Table 1. Various CycleGAN generator

CycleGAN	Generator
Original CycleGAN	Base generator (UNet)
FPN-CycleGAN	Feature pyramid network
UVCGAN	UNet-vision transformer
UVCGAN with edge feature loss	UNet-vision transformer

3.2.2. Modified UVCGAN with edge loss

To enhance the structural quality on the generated nighttime image of the original UVCGAN [24], we attempt to define loss using edge data. This edge data is computed using the Canny operator [29], inspired by [25]. A UNet model denoted by E_f on Figure 5 is fixed and obtain the edge feature to generate a proper nighttime image.

Edge is an important part of an image and should get some attention while building a deep image-to-image translation model to enhance the visualization. In this research, we use a pre-trained weight that act as the prior-knowledge to perform a better image generation. More specifically we use $L1$ distance of the edges between input and the generated images. The edge feature loss is formulated as:

$$\begin{aligned}\mathcal{L}_{Edge}\left(G_{Day}, G_{Night}, Day, Night\right) = {} & \mathbb{E}_{Day\sim p_{data}(Day)}\left[\left\|E_f\left(G_{day}(Day)\right) - E_f(Day)\right\|_1\right] \\ & + \mathbb{E}_{Night\sim p_{data}(Night)}\left[\left\|E_f\left(G_{day}(Night)\right) - E_f(Night)\right\|_1\right] \quad (1)\end{aligned}$$

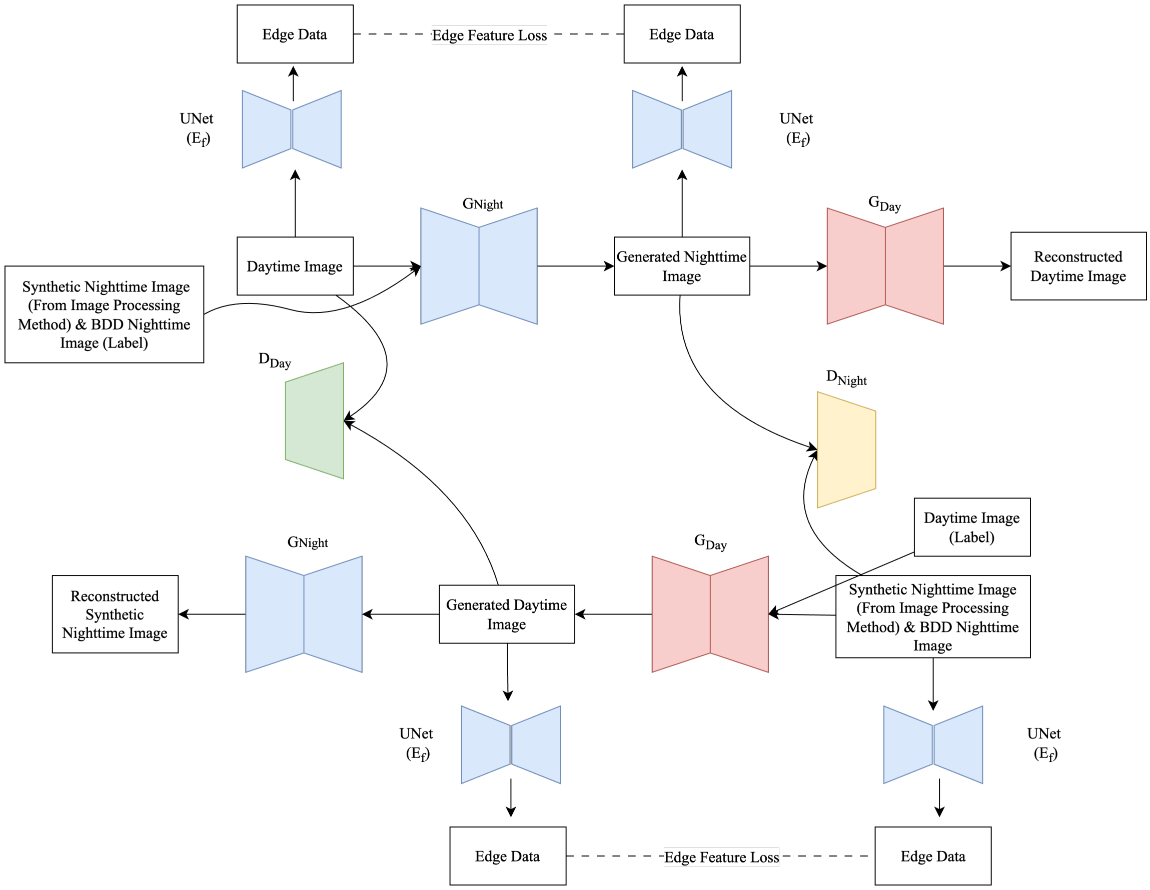

Figure 5. Modified UVCGAN with edge feature loss

3.3. Evaluation metrics

We evaluate the generated nighttime image across various CycleGAN architectures by using FID [30] and SSIM [31]. FID is a metric introduced to evaluate the diversity of images generated by generative models, particularly in the area of GAN [32]. FID measures the dissmiliarity between the probability distributions of the real and generated images in a high-dimensional space learned by Inception v3 model. A lower FID score indicates that the generated images are closer to the real images, represents a better images quality. SSIM evaluates the similiarity between two images using structural information in the image. The SSIM score value ranges from -1 to 1, where 1 indicates perfect similiarity, 0 indicates no similiarity, and -1 indicates perfect dissimiliarity.

4. RESULTS AND DISCUSSION

The approach to this training process is performed by training the model from end-to-end without using a pre-trained model. This is done to ensure the model learns particular features from the given datasets. The training is performed on Google Colab Pro using the NVIDIA A100 GPU. Some hyperparameters are used for the training process, such as the input size, epochs, learning rate, and batch size. We built four CycleGAN model with different generator architecture. We use a total of 170 and 50 images for training and testing, respectively. On the training data, we use 70 paired images and 100 unpaired images. We resize the image to $256 \times 256\ pixels$ to prevent resource exhaustion while doing the training process. We use the Adam optimizer with the learning rate value of 0.0002. For the orginal and FPN generator, we use a total of 9 ResNet blocks, as this setting is commonly employed in image translation tasks utilizing CycleGAN. While UVCGAN with or without edge feature loss, we use a total of 12 transformer block. For all types of generators, we also initialize weight kernel on convolution 2D layer that based on Gaussian distribution. We use L1 loss function to the generator, and L2 loss function to the discriminator.

After training the model on proposed datasets described in section 3.2., we tested several samples. For the nighttime dataset generated using traditional image processing methods discussed in section 3.1.1. The resulting nighttime images could not be accurately described as 'night-like' due to the variance in night illuminance. This variability made it challenging for the model to learn the fixed properties of night illuminance. In contrast, the generated nighttime images from the BDD dataset exhibited a more authentic nighttime visualization. This improvement can be attributed to the dataset's unpaired data condition, where the nighttime images were captured directly by a camera rather than generated. The generated synthetic nighttime image by using various generator of CycleGAN are illustrated on Figure 6.

Figure 6. The semi-supervised CycleGAN results for all generators

All models generated 'night-like' images except for the FPN-CycleGAN model. The evaluation scores using FID and SSIM revealed significant insights into each model's performance. As on Table 2, FPN-CycleGAN model showed inconsistency in its results, achieving the highest FID score, which indicates poor image quality compared to the other models. The generated images from FPN-CycleGAN were of low quality, leading to the conclusion that it performs the worst for this type of task. In terms of evaluation metrics, UVCGAN with the UNet-ViT generator outperformed the others, achieving the best generated nighttime images based on the FID metric. Additionally, UVCGAN attained superior training scores across both metrics, showcasing its effectiveness. Although FPN-CycleGAN achieved a better test score on SSIM, its high FID score indicates its inadequacy for the day-to-night image-to-image translation task due to its complexity. By adding an additional loss function, namely edge feature loss, to UVCGAN, we hypothesized that the model's performance would improve by enhancing the overall structural quality through better edge feature representation. However, our results show that the model performed slightly worse than UVCGAN without the edge feature loss. This decline in performance is due to the added complexity, as the model now has to learn an additional task of differentiating edge data between the real and generated images. Therefore, we conclude that UVCGAN is the best model for generating nighttime images, as it consistently provides high-quality results in comparison to other models.

We also perform a qualitative evaluation on CycleGAN implementation using UVCGAN with edge loss as shown in Figure 7(a) and without edge loss as shown in Figure 7(b). The illuminance reconstruction on UVCGAN with edge feature loss is dispersed and does not converge on one position. This is also the case on UVCGAN, but While on UVCGAN, but more controllable. The illuminance dispersion on UVCGAN with edge feature loss is due to the high frequency feature on the clouds that is detected by edge feature by the model. However, this still does not prove that adding an edge loss to UVCGAN could be an issue. As shown in Figure 7 on the right side, we can see that the illuminance is more controllable by the edge feature loss. The illuminance on the roof by UVCGAN without edge feature loss does not converge on a fixed position, therefore an edge feature loss is needed to fix the structural quality of an image.

Table 2. FID and SSIM for various type of CycleGAN generator

Generator	FID		SSIM	
	Train	Test	Train	Test
Vanilla CycleGAN	35.61	62.54	0.44	0.38
FPN-CycleGAN	71.64	104.46	0.48	**0.44**
UVCGAN	**13.67**	**16.68**	**0.49**	0.42
UVCGAN with edge feature loss	21.83	47.79	0.45	0.40

(a)

(b)

Figure 7. Qualitative evaluation of UVCGAN (a) with edge loss and (b) without edge loss.

5. CONCLUSION

Capturing paired images for performing deep image-to-image translation task is tedious and requires careful setup to ensure alignment between the image pairs. Therefore, we need an automated method to obtain the image pairs from one domain to another domain. This paper achieves this using CycleGAN for day-to-night image translation. To obtain the best nighttime image, we compare four generator architrectures in the CycleGAN framework; i.e original CycleGAN, FPN-CycleGAN, UVCGAN, and a modified UVCGAN with an edge feature loss. The CycleGANs are trained using a dataset of day and night images, built from two sources, i.e., paired day and night images generated by our proposed image processing framework with a modified relighting procedure; and the BDD dataset. Our experiments show that in general the UVCGAN model improves day-to-night image translation task in terms of FID and SSIM. UVCGAN outperforms other CycleGAN generator with an FID test score of 16.68 and SSIM of 0.42. The original CycleGAN obtains a FID test score of 62.54 and SSIM of 0.38, UVCGAN with edge feature loss is slightly worse in terms of evaluation metric scores with an FID of 47.79 and SSIM of 0.4, while FPN-CycleGAN obtains an FID score of 104.4 and SSIM score of 0.44. Comparative analysis on this research with the baseline methods shows that UVCGAN model improves day-to-night image translation task in terms of FID and SSIM. It is obtained that UVCGAN outperforms other CycleGAN generator, even original CycleGAN as baseline methods, while FPN-CycleGAN has the worst performance due to the complexity of its model. Although FPN-CycleGAN leading in terms of SSIM, it has a poor visual appearance. This is due to the model complexity and unbound trained weight of its model. However, by adding edge feature loss to UVCGAN we can improve the structural quality of an image. As shown, the illuminance dispersion can be better controlled by the model, yet UVCGAN with edge feature loss is slightly worst in terms of evaluation metric scores, but still performs better in generating nighttime images. Hence, it is concluded UVCGAN performs better in term of generating synthetic nighttime image.

FUNDING INFORMATION

This research was funded by Hibah Riset Internal Faculty of Computer Science, Universitas Indonesia, contract number NKB-8/UN2.F11.D/HKP.05.00/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad F. R. Taufiq	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Laksmita Rahadianti	✓	✓			✓	✓				✓	✓	✓	✓	✓

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author, [MFRT], on request.

REFERENCES

[1] D. Shiotsuka *et al.*, "GAN-based semantic-aware translation for day-to-night images," *2022 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, USA, 2022, pp. 1–6, doi: 10.1109/ICCE53296.2022.9730532.

[2] J. Lee, D. Shiotsuka, G. Bang, Y. Endo, T. Nishimori, K. Nakao, and S. Kamijo, "Day-to-night image translation via transfer learning to keep semantic information for driving simulator," *International Association of Traffic and Safety Sciences Research*, vol. 47, no. 2, pp. 251–262, 2023, doi: 10.1016/j.iatssr.2023.04.001.

[3] A. Dosovitskiy, G. Ros, F. Codevilla, A. M. López, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, Oct. 2017, pp. 1–16.

[4] E. Santana and G. Hotz, "Learning a driving simulator," *arXiv-Computer Science*, pp. 1–8, 2016

[5] A. Punnappurath, A. Abuolaim, A. Abdelhamed, A. Levinshtein, and M. S. Brown, "Day-to-night image synthesis for training nighttime neural ISPs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, 2022, pp. 10759–10768, doi: 10.1109/CVPR52688.2022.01050.

[6] S. Zini, C. Rota, M. Buzzelli, S. Bianco, and R. Schettini, "Back to the future: a night photography rendering ISP without deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023, pp. 1465–1473, doi: 10.1109/CVPRW59228.2023.00151.

[7] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012, doi: 10.1109/TIP.2012.2214050.

[8] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, March 2013, doi: 10.1109/LSP.2012.2227726.

[9] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," *2015 Twenty First National Conference on Communications (NCC)*, Mumbai, India, 2015, pp. 1–6, doi: 10.1109/NCC.2015.7084843.

[10] Y. Liu, K. Yang, and H. Yan, "No-reference image quality assessment method based on visual parameters," *Journal of Electronic Science and Technology*, vol. 17, no. 2, pp. 171–184, 2019, doi: 10.11989/JEST.1674862X.70927091.

[11] T. -C. Wang, M. -Y. Liu, J. -Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.

[12] J. -Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[13] H. Madokoro *et al.*, "Semantic segmentation of agricultural images based on style transfer using conditional and unconditional generative adversarial networks," *Applied Sciences*, vol. 12, no. 15, Aug. 2022, doi: 10.3390/app12157785.

[14] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *37th International Conference on Machine Learning*, vol. 119, 2020, pp. 1597–1607.

[15] H. Kerdegari, M. Razaak, V. Argyriou, and P. Remagnino, "Smart monitoring of crops using generative adversarial networks," *Computer Analysis of Images and Patterns*, pp. 457–468, 2019, doi: 10.1007/978-3-030-29888-3_45.
[16] Z. Xie *et al.*, "Boosting night-time scene parsing with learnable frequency," *IEEE Transactions on Image Processing*, vol. 32, pp. 2386–2398, 2023, doi: 10.1109/TIP.2023.3267044.
[17] Z. Shi, M. Zhu, B. Guo, M. Zhao, and C. Zhang, "Nighttime low illumination image enhancement with single image using bright/dark channel prior," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 1–15, 2018, doi: 10.1186/s13640-018-0251-4.
[18] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1–26, 1980, doi: 10.1016/00160032(80)900587.
[19] B. Adhikari, K. C. Hari, and S. Thapa, "Night to day and day to night image transfer using generative adversarial network," in *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 3, pp. 1053–1060, 2022.
[20] M. Kennerley, J. -G. Wang, B. Veeravalli, and R. T. Tan, "2PCNet: Two-phase consistency training for day-to-night unsupervised domain adaptive object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 11484–11493, doi: 10.1109/CVPR52729.2023.01105.
[21] W. -Z. Zheng, V. -H. Tran, and C. -C. Huang, "D2NA: Day-to-night adaptation for vision based parking management system," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 1573–1577, doi: 10.1109/ICASSP40776.2020.9053567.
[22] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *International Journal of Information Management Data Insights*, vol. 1, no. 1, 2021, doi: 10.1016/j.jjimei.2020.100004.
[23] Y. Li, J. Lu, and X. Meng, "Combining feature pyramid and CycleGAN for image generation," *Journal of Physics: Conference Series*, 2023, pp. 1–8, doi: 10.1088/1742-6596/2646/1/012033.
[24] D. Torbunov *et al.*, "UVCGAN: UNet vision transformer cycle-consistent GAN for unpaired image-to-image translation," *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, USA, 2023, pp. 702–712, doi: 10.1109/WACV56688.2023.00077.
[25] H. Dou, C. Chen, X. Hu, L. Jia, and S. Peng, "Asymmetric CycleGAN for image-to-image translations with uneven complexities," *Neurocomputing*, vol. 415, pp. 114–122, 2020, doi: 10.1016/j.neucom.2020.07.044.
[26] F. Yu *et al.*, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 2020, pp. 2633–2642, doi: 10.1109/CVPR42600.2020.00271.
[27] Y. Qian, J. -K. Kämäräinen, J. Nikkanen, and J. Matas, "On finding gray pixels," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, USA, 2019, pp. 8054–8062, doi: 10.1109/CVPR.2019.00825.
[28] N. Moroney, "Local color correction using non-linear masking," in *IS&T/SID Eighth Color Imaging Conference*, vol. 8, 2000, pp. 108–111, doi: 10.2352/CIC.2000.8.1.art00021.
[29] J. Canny, "A computational approach to edge detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.
[30] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, Long Beach, California, USA, 2017, pp. 6629–6640, doi: 10.5555/3295222.3295408.
[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[32] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," *30th Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain, 2016, pp. 1–9.

BIOGRAPHIES OF AUTHORS

Muhammad Feriansyah Raihan Taufiq earned his Bachelor of Science (B.Sc.) degree in system and instrumentation physics in 2022, and his Master of Computer Science (M.CS.) degree in 2025 from the University of Indonesia. He is currently pursuing a master's degree in computer science at the same university. To strengthen his knowledge of artificial intelligence, he has taken courses in the area of machine learning and deep learning. He also gained hands-on experience in solving deep learning problems through his bachelor's thesis, which focused on reconstructing hyperspectral images of bisbul leaves using deep learning techniques, applying his skills to real-world challenges in the field of deep computer vision. He can be contacted at email: muhammad.feriansyah21@ui.ac.id.

Laksmita Rahadianti received her B.CS. degree from University of Indonesia in 2009, her M.Sc. degree from Erasmus Mundus Color in Informatics and Media Technology master's program from the University Saint Étienne, France, University of Granada, Spain, and Gjøvik University College (now Norwegian University of Science and Technology) in 2012, and her Ph.D. degree in computer science and engineering from Nagoya Institute of Technology, Japan, in 2018. She has been part of the teaching staff at Faculty of Computer Science, Universitas Indonesia since 2018, and Associate Professor since 2021. Her research interests include computer vision, photometric image analysis, image restoration, and image quality. She has handled various research projects and supervised multiple student theses in various fields of machine learning and computer vision. She can be contacted at email: laksmita@cs.ui.ac.id.