mathematics

MDPI

Article

Enhanced Night-to-Day Image Conversion Using CycleGAN-Based Base-Detail Paired Training

Dong-Min Son, Hyuk-Ju Kwon and Sung-Hak Lee *

School of Electronic and Electrical Engineering, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 41566, Republic of Korea; forhollow@knu.ac.kr (D.-M.S.); olin1223@knu.ac.kr (H.-J.K.)
* Correspondence: shak2@ee.knu.ac.kr; Tel.: +82-53-950-7216

Abstract: Numerous studies are underway to enhance the identification of surroundings in nighttime environments. These studies explore methods such as utilizing infrared images to improve night image visibility or converting night images into day-like representations for enhanced visibility. This research presents a technique focused on converting the road conditions depicted in night images to resemble daytime scenes. To facilitate this, a paired dataset is created by augmenting limited day and night image data using CycleGAN. The model is trained using both original night images and single-scale luminance transform (SLAT) day images to enhance the level of detail in the converted daytime images. However, the generated daytime images may exhibit sharpness and noise issues. To address these concerns, an image processing approach, inspired by the Stevens effect and local blurring, which align with visual characteristics, is employed to reduce noise and enhance image details. Consequently, this study contributes to improving the visibility of night images by means of day image conversion and subsequent image processing. The proposed night-to-day image translation in this study has a processing time of 0.81 s, including image processing, which is less than one second. Therefore, it is considered valuable as a module for daytime image translation. Additionally, the image quality assessment metric, BRISQUE, yielded a score of 19.8, indicating better performance compared to conventional methods. The outcomes of this research hold potential applications in fields such as CCTV surveillance systems and self-driving cars.

Keywords: image-to-image translation; CycleGAN; Pix2Pix; luminance adaptation transform; image processing; Stevens effect

MSC: 68T45

Citation: Son, D.-M.; Kwon, H.-J.; Lee, S.-H. Enhanced Night-to-Day Image Conversion Using CycleGAN-Based Base-Detail Paired Training. *Mathematics* **2023**, *11*, 3102. https://doi.org/10.3390/math11143102

Academic Editor: Stefano De Marchi

Received: 12 June 2023
Revised: 1 July 2023
Accepted: 12 July 2023
Published: 13 July 2023

1. Introduction

Image-to-image translation is a field that utilizes generative models to map input image sets to output image sets. Traditionally, this task was addressed through image processing, computer vision, or computer graphic techniques; however, these approaches often yielded unsatisfactory results and led to various issues [1]. To overcome these limitations and enhance performance, researchers have developed deep-learning-based methods for image-to-image translation. Two prominent methods in this domain are Pix2Pix and CycleGAN. These deep learning networks are based on generative adversarial networks (GANs) [2]. GAN-based image-to-image translation has emerged as the preferred approach, demonstrating impressive results. This technique finds applications in diverse fields, such as converting labels into tangible objects or transforming black and white images into color photographs. Pix2Pix is a specific type of conditional GAN (CGAN) that learns the mapping between a conditional vector (or an input image) and its corresponding paired image in the training data. In a broader sense, Pix2Pix falls under the umbrella of CGAN, where the emphasis is on learning image-to-image translation using a paired dataset consisting of condition images and their corresponding target images, rather than

relying solely on a condition vector [3,4]. The objective of Pix2Pix is to generate an image that closely resembles the original by minimizing the pixel-to-pixel difference between the generated image and the real image. This is achieved through the utilization of CGAN loss and L1 loss.

As described above, Pix2Pix requires a paired dataset, making it difficult to achieve satisfactory image conversion performance when data are limited. Therefore, CycleGAN addresses this limitation by enabling image-to-image translation with unpaired datasets, eliminating the need for paired datasets [5]. CycleGAN consists of two generators. The first generator (G) transforms input images into output images, ensuring forward consistency. The second generator (F) transforms output images back into input images, ensuring backward consistency. This structure, referred to as a cycle structure, returns to itself after cycling through generators G and F, thus establishing cycle consistency. The mode collapse problem [6], wherein GAN forgets the characteristics of inputs and produces the same result image for all inputs, can be mitigated to some extent due to the circular structure. CycleGAN introduces not only GAN loss but also cycle consistency loss and identity loss to enable more natural image transformations [7]. Cycle consistency loss involves reconstructing the image created by the generator and calculating the difference between the original image and the generated image as a loss measure. Additionally, an identity loss method is employed to create an image with colors similar to the real image, encouraging the fake image to imitate the color of the real image to a larger extent. Day and night image conversion, the subject of this study, can be performed with CycleGAN. When converting to a day-to-night image, CycleGAN removes existing information or darkens the bright areas in the day image. However, since non-existent information must be newly generated when converting a night image into a day image, the resulting converted day image may appear awkward.

"Cross-domain car detection using unsupervised image-to-image translation: from day to night" [8] is a research paper that utilizes unsupervised image-to-image translation techniques to address the car detection problem in nighttime scenarios. The paper focuses on bridging the domain gap between daytime and nighttime environments by employing CycleGAN for learning the transformation between day and night images. Specifically, it generates nighttime-like images by applying the annotation information obtained from daytime images, which are used as training data for car detection in nighttime environments. The main finding of the paper is that the trained model performs better in detecting cars in nighttime images compared to models trained solely on daytime data. The key point of this paper is to reduce the domain gap between daytime and nighttime environments to enhance the performance of car detection algorithms in nighttime scenarios. However, the generation of nighttime-like images mentioned in the paper poses a challenge, as it tends to produce unrealistic and fanciful illuminations. The core of day-to-night translation is to darken the objects in daytime images and apply nighttime lighting to make them appear as nighttime images. Although generating fake night images is easier than generating fake day images, object information is lost. Nevertheless, the proposed method in this paper minimizes the loss of object information by employing paired training and post-processing techniques to enhance the visibility of objects during the night-to-day image translation process.

Another method for night-to-day image-to-image translation is ToDayGAN, which demonstrates improved results compared with CycleGAN when transforming night-to-day images [9]. ToDayGAN utilizes a modified ComboGAN [10] network that incorporates the concept of CycleGAN, enhancing performance by employing three different discriminators. The modified network only requires two domain transformations: day and night. The fake B image generated by the A-to-B generator is fed as input to the B-to-A generator. This modified structure yields superior performance for night-to-day image translation. In the discriminator, traditional GANs compare fake and real images. However, this paper introduces discriminator networks with the same structure to compare fake and real images using three methods: Blur-RGB, Grayscale, and xy-Gradients. Similar to CycleGAN,

ToDayGAN can cause images to resemble daytime scenes while preserving the sky regions, which are ambiguous in night images. However, object details in night images are not translated as clearly.

The focus of this study is night-to-day image-to-image translation, specifically emphasizing the representation of object information on the road from the perspective of a driver at night. The proposed method aims to reveal detailed information, particularly the shape of objects on the road, in night images. The training of the proposed method was conducted using paired data consisting of base and detailed images. Additionally, the Stevens effect was applied as a post-processing method to enhance the clarity of object details in the night image [11]. The daytime image generated through this method exhibits increased noise due to the enhancement of details in dark areas where objects are not visible in the original nighttime image. To mitigate this noise, a weighted difference local blur map was defined. This map reduces noise while preserving the details enhanced by the Stevens effect in other regions. Finally, a region of interest (ROI) was defined based on the driver's viewpoint, and the image quality metrics were compared with conventional methods. Given that the study evaluates the driver's visual perception within a limited range of vision—similar to studies on geometric visual models [12]—the attentiveness of drivers in augmented reality visualization materials in HUD (head-up display) technology [13], or human visual characteristics in car-following models [14,15], the image quality evaluation focuses on the ROI area of the image. Therefore, the result of this study compares image feature metrics by cropping the image to the object-oriented ROI area.

2. Materials and Methods

2.1. Paired Dataset Training Using CycleGAN

In this study, the training for unpaired day-to-night image translation was conducted using the CycleGAN network, resulting in the creation of a paired dataset. Furthermore, the generated fake night images from the unpaired day-to-night image translation module of CycleGAN were paired with real day images to construct a paired dataset for the creation of a paired night-to-day image translation module using CycleGAN. The proposed method utilizes the CycleGAN structure, which consists of a ResNet generator [16] and a PatchGAN discriminator [4]. The structure of CycleGAN is visually depicted in Figure 1.

First, CycleGAN comprises two generators and two discriminators because it is an image conversion of two domains. In Pix2Pix, the generator exhibits a U-Net structure; the biggest feature of this structure is the skip connection, which reduces the loss of detail information at a low level. However, because of its limited depth, high-level features such as abstract information cannot be extracted [17]. Therefore, the generator in CycleGAN uses ResNet. CycleGAN's ResNet generator possesses a deep network depth and can preserve the detailed parts of data images to a larger extent by utilizing skip connections. As shown in Figure 1, CycleGAN's generator consists of three parts: encoder, transformer, and decoder, with the transformer part incorporating residual blocks (ResNet), as depicted in Figure 1.

The discriminator in CycleGAN follows a similar approach to PatchGAN, which is similar still to Pix2Pix. The patch size of PatchGAN is determined by the receptive field size, which is determined by the number of convolution layers in the discriminator. The final output of PatchGAN can be obtained by either averaging all values in the last feature map or using a one-dimensional scalar value as the output throughout the layers. In CycleGAN's PatchGAN discriminator, a one-dimensional scalar value is used as the final output to determine the authenticity of an image. This approach is computationally efficient and incurs low cost. The patch size in CycleGAN is typically around 70×70, enabling the application of the patch-level discriminator structure to images of various sizes with fewer parameters. An activation function is applied in the last layer of the discriminator to classify the results. CycleGAN does not employ a separate sigmoid function or activation function for classification, but rather utilizes mean square error loss, borrowed from LSGAN [18], to

determine the authenticity of the image. The structure of CycleGAN ensures the effective determination and updating of the loss value.

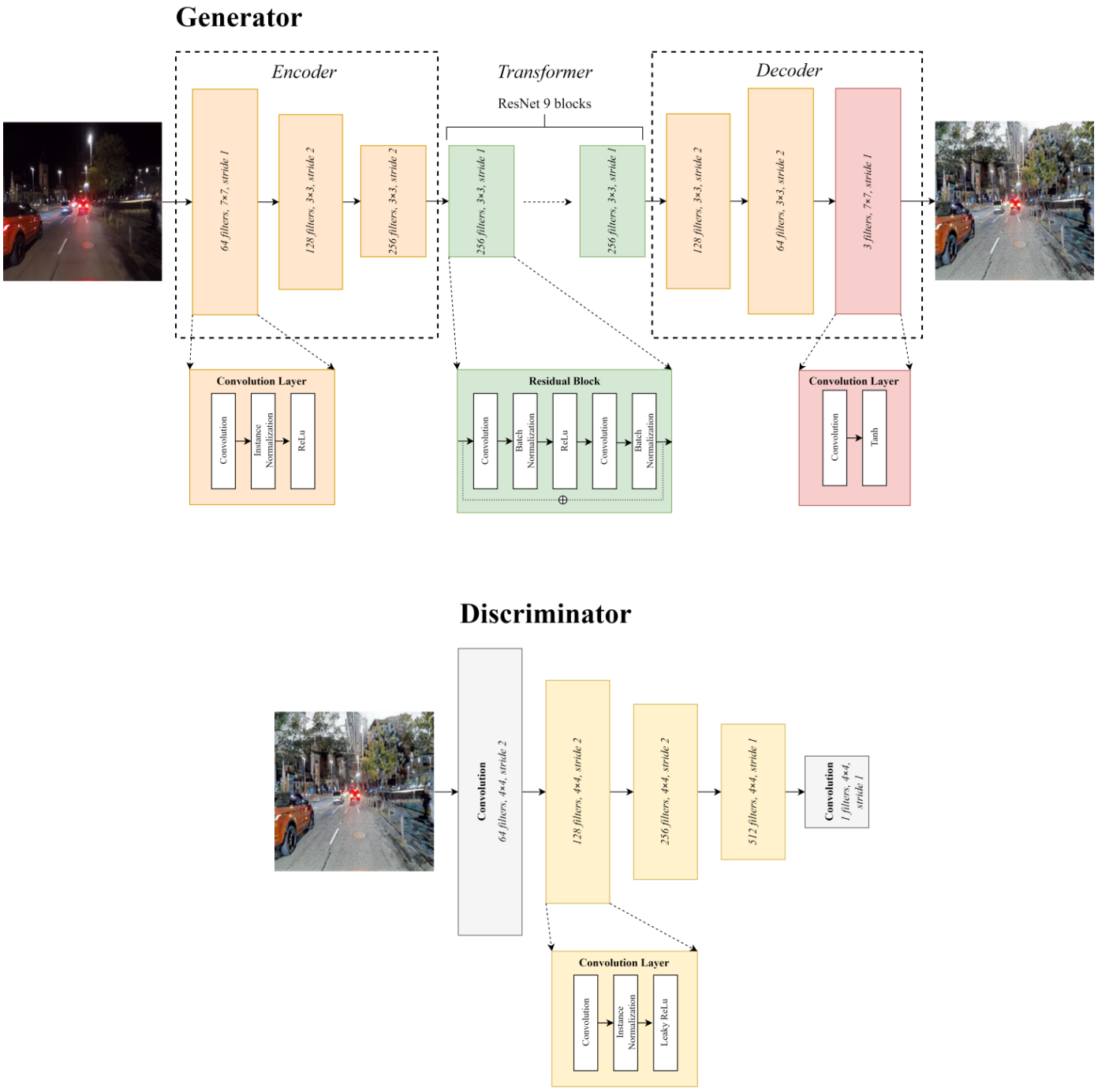

Figure 1. CycleGAN's generator and discriminator architecture.

The proposed method involves the use of two CycleGANs. The first CycleGAN training focuses on unpaired learning for day-to-night image translation, which serves as a data augmentation task to compensate for the limited number of paired day and night images. Subsequently, the second CycleGAN training is performed for night-to-day image translation and detail enhancement. In the first CycleGAN module, the night images, which serve as paired images to the day images, are generated using base night images that have been blurred with a bilateral filter. Real day images corresponding to the night images are

processed with a single-scale luminance adaptation transform (SLAT) [19] to construct the training dataset for the second CycleGAN training.

Although CycleGAN is designed for unpaired image-to-image translation, the night-to-day image translation is performed using paired learning. This is because the results obtained from paired learning with CycleGAN outperform those obtained from paired learning with Pix2Pix. Figure 2 illustrates this comparison, where training with the same paired dataset resulted in a blurry image when using Pix2Pix (Figure 2b), whereas the image generated with the CycleGAN pair module (Figure 2c) is slightly clearer. The improved clarity in the resulting image of CycleGAN, compared with Pix2Pix, is attributed to CycleGAN's ability to address the mode collapse problem through cycle consistency loss and its ability to preserve color using identity loss.

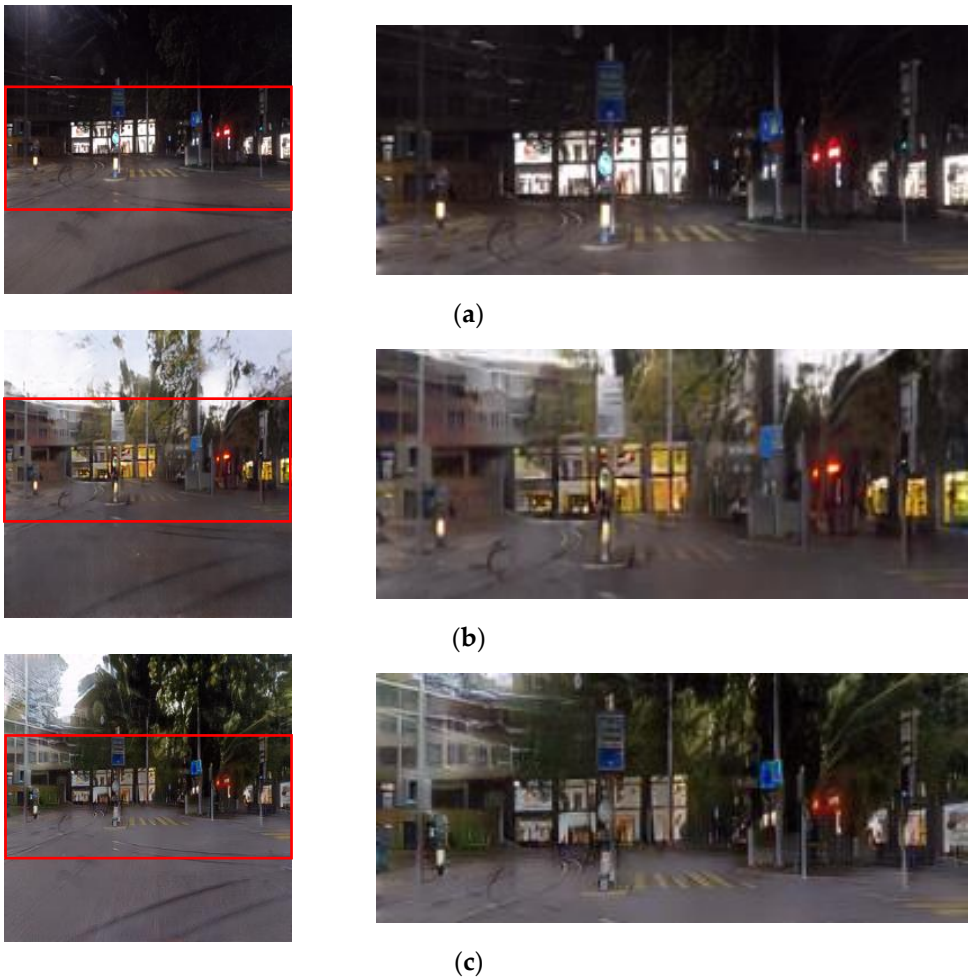

Figure 2. Image-to-image translation module's result image comparison: (**a**) real night image; (**b**) Pix2Pix's result image; (**c**) paired CycleGAN result image.

Another reason for using pair learning in CycleGAN is that unpaired learning can lead to distortions in traffic light colors. Figure 3 presents a comparison of four methods: Pix2Pix, unpaired CycleGAN, paired CycleGAN, and the proposed method. Unpaired CycleGAN suffers from information deficiency compared with paired learning, which can result in color distortions of traffic lights and even the disappearance of signs. Figure 3c displays the result images obtained from training with unpaired CycleGAN. Upon closer examination of the cropped image, it is evident that the traffic light color is faint and the representation of the traffic light itself is inadequate. Moreover, in the remaining paired training images, such as Pix2Pix, paired CycleGAN, and the proposed method module, there is relatively better preservation of traffic light color. Therefore, it can be observed

that the representation of traffic lights in the result images from paired training methods is superior to that of the unpaired CycleGAN module. Specifically, the result images from the proposed method module not only enhance the details of objects but also depict signage more clearly, as can be seen in the images.

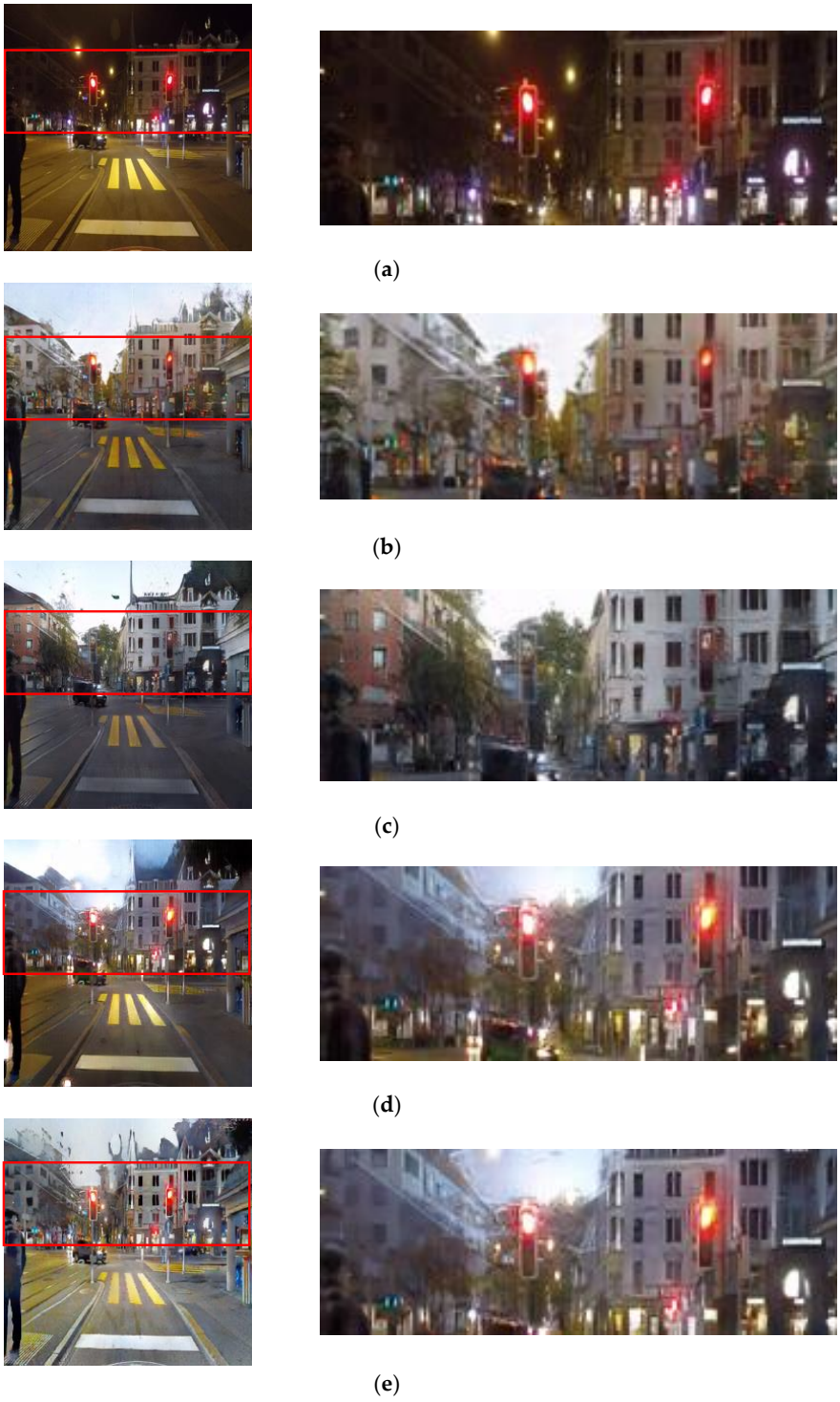

Figure 3. Result image comparison of the image-to-image translation module: (**a**) real night image; (**b**) Pix2Pix's result image; (**c**) unpaired CycleGAN result image; (**d**) paired CycleGAN result image; (**e**) proposed method's result image.

The proposed training method in this study aims to preserve the original colors and objects of the input images while enhancing object details. To achieve this, paired CycleGAN training was utilized. The block diagram in Figure 4 illustrates the proposed method, which involves unpaired CycleGAN training to generate fake nighttime images. These fake night images were then paired with daytime images to create a day–night paired dataset. In the next step of CycleGAN pair training, the fake night images were blurred using a bilateral filter. The original fake night images and the blurred fake night images were subtracted from each other, resulting in base night images, which represent the differences between the images.

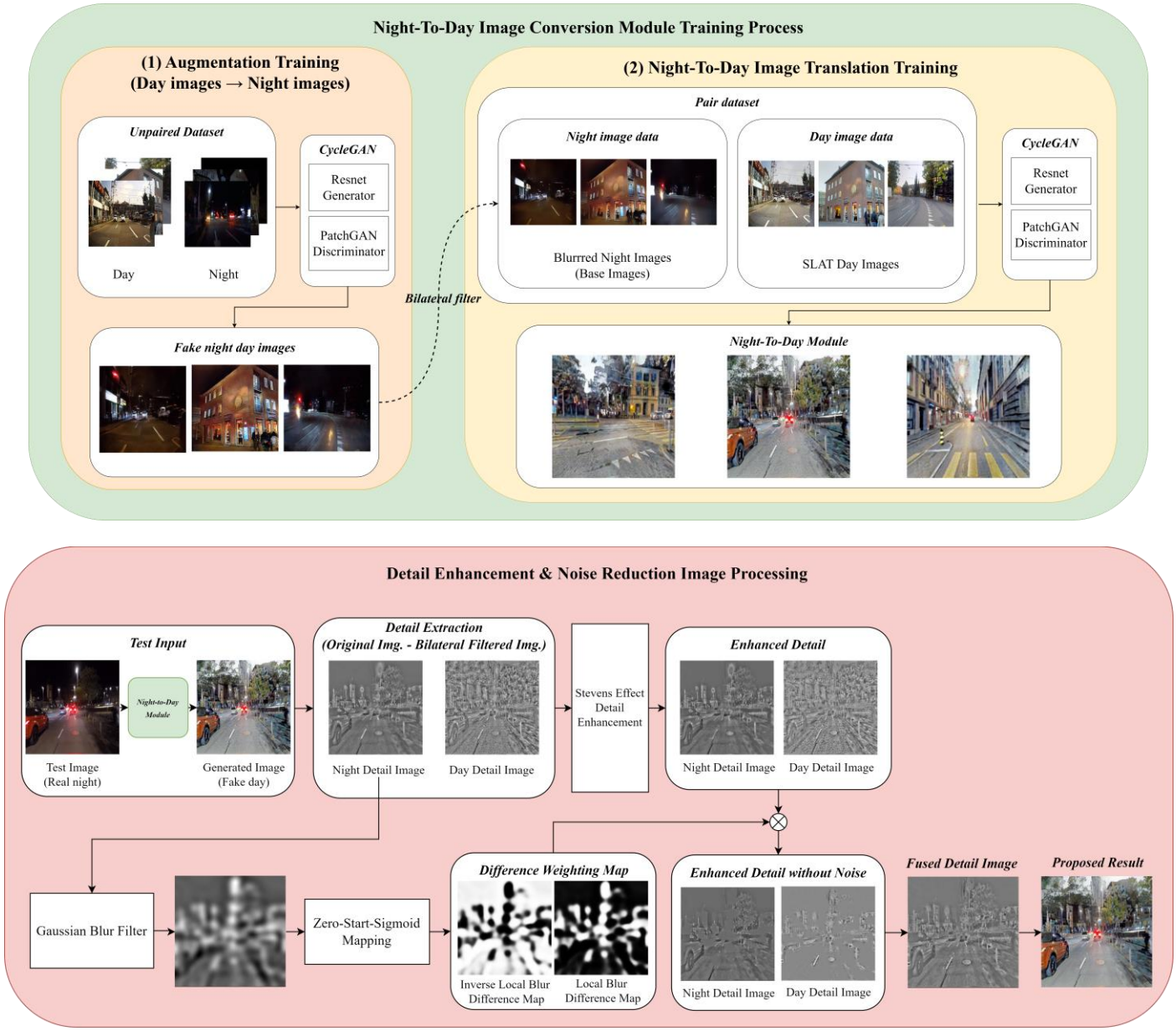

Figure 4. Block diagram of the proposed method.

To perform both day conversion training and detail training simultaneously, SLAT processing was applied to the real daytime image corresponding to the base fake night image to construct a paired dataset. SLAT processing is an image processing technique that preserves detail areas while improving local contrast. It involves two main processes: local tone mapping of luminance channels and chrominance compensation of chrominance

channels. However, in this study, the chrominance channel compensation method was omitted to simplify the SLAT processing. Instead, the RGB channels of the real daytime image were separated, and the luminance components of each channel were used for local tone mapping. Figure 5 illustrates the process of SLAT processing. Overall, the proposed training method combines unpaired CycleGAN training, bilateral filtering, and SLAT processing to create a paired dataset and enhance object details while preserving the original colors and objects of the input images.

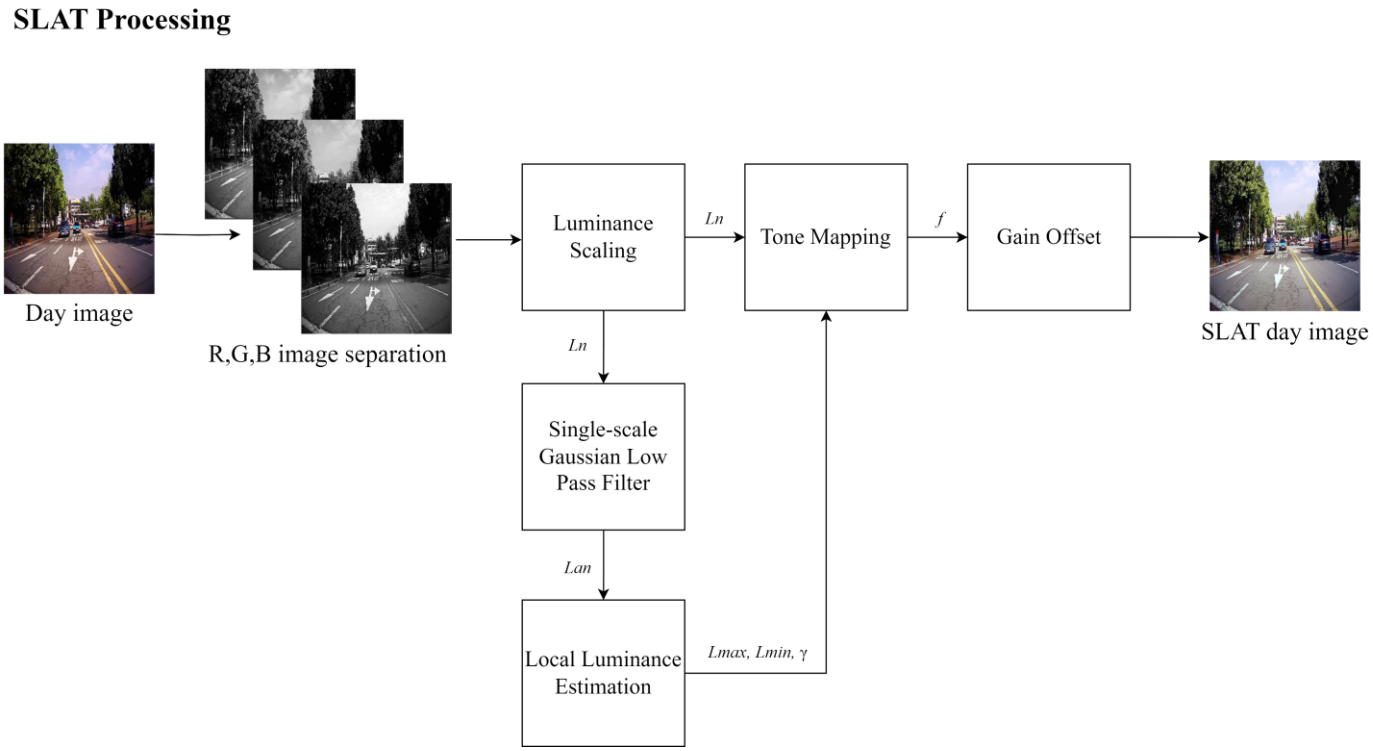

Figure 5. Single-scale luminance adaptation transform in RGB channel process.

SLAT adjusts the visually compensated gamma value according to the local adaptive luminance level. The luminance level is divided into minimum and maximum luminance levels. To set the maximum value of the luminance channel to 100, luminance scaling normalization is performed because local luminance estimation is designed according to an adaptation luminance of 100 cd/m^2 as the surrounding luminance. Next, a single-scale Gaussian low pass filter is applied to each RGB luminance scaled image to create an adaptation luminance image (L_{an}) that is essentially an image of the surrounding area. Luminance scaling normalization and normalized adaptation luminance are calculated using Equations (1) and (2).

$$L_n = (I - \min(I))/(\max(I) - \min(I)) \times 100, \quad (1)$$

where L_n is a normalized luminance scaling input image, and I is the input luminance image. The RGB image is the luminance image in this study. Therefore, the R channel L_n, G channel L_n, and B channel L_n are generated.

$$L_{an} = Gauss\_blur(L_n)/\max(Gauss\_blur(L_n)) \times 100, \quad (2)$$

where L_{an} is the normalized adaptation luminance calculated by a single-scale Gaussian low pass filter, and $Gauss\_blur$ is a Gaussian low pass filter. This L_{an} surrounding image is also generated by the RGB channel.

Next, local luminance estimation is performed to calculate the L_{max}, L_{min}, and γ of each RGB channel. Here, the γ value is the visual gamma value that affects SLAT based on the Bartelson–Breneman brightness function [10]. Values can be obtained using Equations (3)–(5):

$$L_{min} = 0.00212 + 0.0185 L_{an}^{1.0314}, \tag{3}$$

$$L_{max} = 25.83 + 30.82 L_{an}^{0.6753}, \tag{4}$$

$$\gamma = 0.444 + 0.045 \times \ln(L_{an} + 0.6034), \tag{5}$$

where L_{min} is the minimum luminance level, L_{max} is the maximum luminance level, and γ is the visual gamma value.

Applying the visual gamma value to the Bartelson–Breneman brightness function curve and applying it to the calculated luminance image values thus far yields the SLAT image, as expressed in Equation (6):

$$f = (L_n - L_{min})/(L_{max} - L_{min})^{\gamma}, \; SLAT = 99.9 \times f + 0.1, \tag{6}$$

where f is the Bartelson–Breneman brightness function curve with the visual gamma value, $SLAT$ is an image processed by a single luminance adaptation transform, L_n is a normalized luminance scaling input image, L_{min} is the minimum luminance level, L_{max} is the maximum luminance level, and γ is the visual gamma value. The SLAT result image enhances the local contrast and desaturates the bright area.

In Figure 6, a comparison between the SLAT-processed image and the original image has been conducted to examine the presence of artifacts such as ringing effects, where erroneous boundaries are formed in the surrounding regions, resulting in a loss of local contrast and a decrease in image sharpness. When comparing the SLAT-processed image with the Canny edge image, it was observed that the SLAT-processed image exhibited minimal formation of erroneous boundaries, and there were areas where the local contrast increased, leading to better representation of details compared to the original image. Therefore, utilizing SLAT processing on low-light images before training could potentially assist in enhancing the representation of object areas.

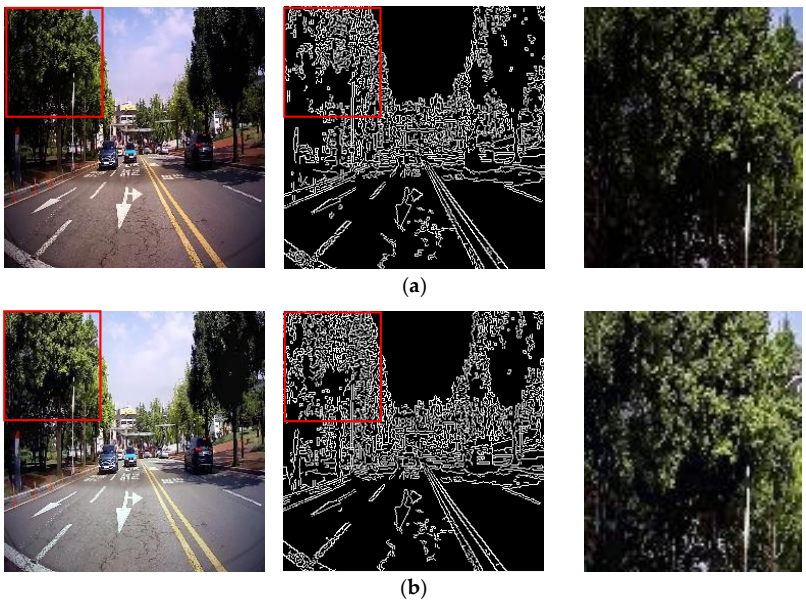

Figure 6. Image comparison of original image and SLAT image: (**a**) original image with canny edge; (**b**) SLAT image with canny edge.

The real daytime images, after being processed with SLAT, and the fake nighttime base images generated through unpaired CycleGAN training were utilized as paired images for training in the paired CycleGAN. The resulting paired CycleGAN (base-detail model) effectively enhances the local contrast and restores fine details in the daytime image, leading to improved object identification.

Figure 7 presents a comparison between the result images of unpaired CycleGAN and the base-detail model. The base-detail model effectively describes objects by enhancing their details, but it also introduces noise. Excessive enhancement of details can lead to noticeable noise in the resulting images. Therefore, it is crucial to employ image processing techniques that preserve details while reducing noise.

(**a**) (**b**)

Figure 7. Result image comparison of unpaired CycleGAN and base-detail model: (**a**) unpaired CycleGAN result; (**b**) base-detail model result.

2.2. Detail Enhancement and Noise Reduction

To address the increased noise in the resulting image from the base-detail model, image processing techniques are employed to reduce noise while enhancing detail information. As illustrated in Figure 4, the proposed approach utilizes the Stevens effect and a local blur map to achieve this. Initially, the detail information is extracted from both the real nighttime images and the resulting daytime images from the base-detail model by obtaining the difference image. This difference image contains the detailed information of both the night and day images, obtained by subtracting the original image from the image blurred with a bilateral filter. Subsequently, the Stevens effect is applied to each detail image of the day and the night, enhancing the sharpness of the detail components while considering visual characteristics. The mathematical expressions for the Stevens effect are defined in Equations (7)–(9). These equations provide the framework for enhancing the details and improving the visual quality of the images.

$$La = 0.2 \times base,\ k = \frac{1}{5La + 1}, \tag{7}$$

where *base* is an image with bilateral filter blur, k is the adjustment factor, and La is 20% of the adaptation luminance.

$$FL = 0.2k^4(5La) + 0.1\left(1 - k^4\right)^2(5La)^{1/3}, \tag{8}$$

where FL is a factor of various luminance-dependent appearance effects that is used to calculate the detail enhancement of the Stevens effect.

$$detail_{stevens} = detail^{(FL+0.8)^{0.25}}, \tag{9}$$

where $detail_{stevens}$ is a sharper detail image based on the application of the Stevens effect to the detail image, and *detail* refers to the detail extraction image. The Stevens effect is utilized to refine the details by considering the anticipated modifications in luminance

levels. Specifically, as the luminance level rises, there is an augmentation in local perceptual contrast, resulting in the visual enhancement of details. The alteration in detail, enhanced through the application of the Stevens effect, is illustrated in Figure 8. Notably, the image with the Stevens effect applied appears noticeably clearer, reflecting the effectiveness of this technique in enhancing visual clarity.

Figure 8. Stevens effect image adaptation: (**a**,**c**) original images; (**b**,**d**) Stevens effect adapted images.

To mitigate noise in the resulting image from the base-detail model, the proposed method employs a local blur weight map to reduce noise while preserving essential details. The generation of the local blur weight map relies on the differential image of the night image. This choice is motivated by the fact that regions with pixel values close to zero in the difference image indicate the absence of detail information, while nonzero pixel values correspond to regions with detail information. Thus, the difference image serves as a criterion for selecting regions where noise can be locally reduced while preserving detail. The difference image of the nighttime image is essentially the same as the nighttime detail image mentioned earlier.

In order to use the difference image as a blur map, it undergoes blurring with a Gaussian blur filter to create a visually blurry representation. To render the blurred difference image suitable as a weight map, the values of the difference image are stretched. It is important to avoid extreme separations in the local blur weight map, as they can introduce unnatural representation of image details and image distortion artifacts. To address this, the values of the local blur weight map are mapped to a sigmoid-type graph, which reduces the occurrence of such unnatural phenomena. However, since the standard sigmoid function starts within a range smaller than zero, a modified sigmoid function is employed in this context. Specifically, the sigmoid function is set to start from zero. The resulting difference blur map, generated using this modified sigmoid function formula, is represented by Equation (10).

$$Diff_{map} = \left\{ 1 + \left(\frac{Img_{diff} \times (1 - 0.5)}{0.5 \times \left(1 - Img_{diff}\right)} \right)^{-5} \right\}^{-1}, \tag{10}$$

where $Diff_{map}$ is the local blur difference weight map and Img_{diff} is an input image of the Gaussian blurred difference image. The application of the sigmoid function to the local blur weight map and its subsequent use for blurring is aimed at achieving a more natural appearance in the resulting image.

In essence, a locally applicable local blur weight map is necessary to reduce noise and preserve details. This weight map can be generated from the difference image of the nighttime image. However, if the weight map follows a linear step-like form, it can lead to image distortion or unnatural artifacts. By mapping the weight map values to a sigmoid function, a non-linear local blur weight map can be created, mitigating unnatural artifacts in the image while locally attenuating noise.

To enhance details in both the nighttime detail image and the daytime detail image, the Stevens effect is applied. Consequently, two local blur weight maps are required: one corresponding to the nighttime detail image and the other to the daytime detail image. The local blur weight map for the nighttime detail image can be obtained using the weight map mentioned earlier. However, for the daytime detail image, the local blur weight map needs to be in an inverse form. This is because regions with pixel values close to zero in the difference image of the nighttime image potentially indicate the presence of objects in the corresponding regions of the daytime image. Thus, it is necessary to preserve these regions in the daytime image. By utilizing the inverse local blur weight map, it becomes possible to incorporate the daytime image's detail information in areas where the nighttime image lacks detail, thus preserving details throughout the entire image. This relationship is expressed mathematically as shown in Equation (11).

$$Day_{fused\_detail} = \left(Diff_{map} \times Night_{stevens\_detail}\right) + \left(\left(1 - Diff_{map}\right) \times Day_{stevens\_detail}\right), \tag{11}$$

where Day_{fused_detail} is a synthesized detail image processed based on the Stevens effect. $Diff_{map}$ is an image representing a local blur weight map using a difference image of a night image as shown in Equation (10), and $(1 - Diff_{map})$ is an inverse local blur map image.

Finally, as shown in Equation (12), the proposed result image is generated by adding the base image of the generated day image and the enhanced detail image.

$$Night2Day\ image = Day_{fused\_detail} + Day_{base}, \tag{12}$$

where $Night2Day\ image$ is a result of the proposed method, Day_{fused_detail} is an enhanced detail image based on the Stevens effect as shown in Equation (11), and Day_{base} is an image of the base-detail model's result without detailed information.

In conclusion, the proposed method employs a paired CycleGAN module to convert nighttime images into daytime images. The fake daytime images generated in this module undergo detail enhancement using the Stevens effect and a local blur weight map, which

takes into account visual characteristics. This enhancement process aims to improve the visual detail information while reducing noise. As a result, the final converted daytime image allows for better recognition of objects even in low-light conditions. The Night2Day image translation results are presented in Figure 9. It is evident from Figure 8 that the Night2Day image exhibits improved and sharper details compared with the existing CycleGAN result image. Moreover, it can be observed that the noise is reduced and the detail areas are preserved in comparison to the image generated by the base-detail module.

(**a**) (**b**) (**c**)

Figure 9. Result image comparison of unpaired CycleGAN and base-detail model (paired CycleGAN) with the proposed method: (**a**) unpaired CycleGAN result; (**b**) base-detail model's result; (**c**) proposed method's result.

3. Experimental Results

3.1. Dataset, Computer, and Software Specifications

GAN-based image-to-image translation methods were implemented on a PC with the following specifications: Intel i9-10980XE 3.00 GHz CPU, 256 GB RAM, and an NVIDIA GeForce RTX 3090 GPU. The proposed deep learning network was based on CycleGAN using the PyTorch framework. For optimization, CycleGAN utilized the Adam optimizer with β parameters set to 0.5 and 0.999, respectively. The batch size was set to 1, and the learning rate was initialized to 0.0002, which linearly decreased every 100 epochs. The total number of epochs was set to 250, and the image crop size was fixed at 256 $\times$ 256 pixels.

The training dataset for the unpaired Day2Night CycleGAN module consisted of 6450 day images, comprising 1200 images from our dataset and 5250 images from the Dark Zurich [20] and ACDC (Adverse Conditions Dataset with Correspondences) [21] datasets. Additionally, 7744 night images were included in the training dataset, comprising 5400 images from our dataset and 2344 images from the Dark Zurich and ACDC datasets. These night images were used to generate fake night images through the unpaired Day2Night CycleGAN.

The dataset for the base-detail model comprised 4852 real day images, consisting of 1005 images from our dataset and 3847 images from the Dark Zurich and ACDC datasets. Paired with these real day images were 4852 fake night images generated by the unpaired Day2Night CycleGAN. Finally, a set of 20 real night test images was used for evaluating the performance of the base-detail model.

3.2. Night-to-Day Image Translation Simulation Result Comparison

Deep learning methods for night-to-day image translation include Pix2Pix, which is a paired learning method, CycleGAN, which is an unpaired learning method, and ToDayGAN, a variant of CycleGAN with modified architecture. In comparison to the results obtained from these conventional methods, the proposed method demonstrates superior expressive capabilities in capturing details. The fake daytime images generated by the proposed method exhibit enhanced detail representation compared with the results produced by the other methods.

In Figure 10, it is observed that CycleGAN and ToDayGAN, which are unpaired learning methods, struggle to accurately represent the red color of the traffic light. In contrast, the paired learning method Pix2Pix and the proposed method excel in expressing the red color of the traffic light. As mentioned earlier, unpaired learning methods generate new information through learning, which can result in higher information loss compared with paired learning. Moreover, in the ROI where objects are present from the perspective of a car driver, the depiction of the road surface with a pedestrian, as well as the representation of buildings, is more effectively described by the proposed method compared with other methods.

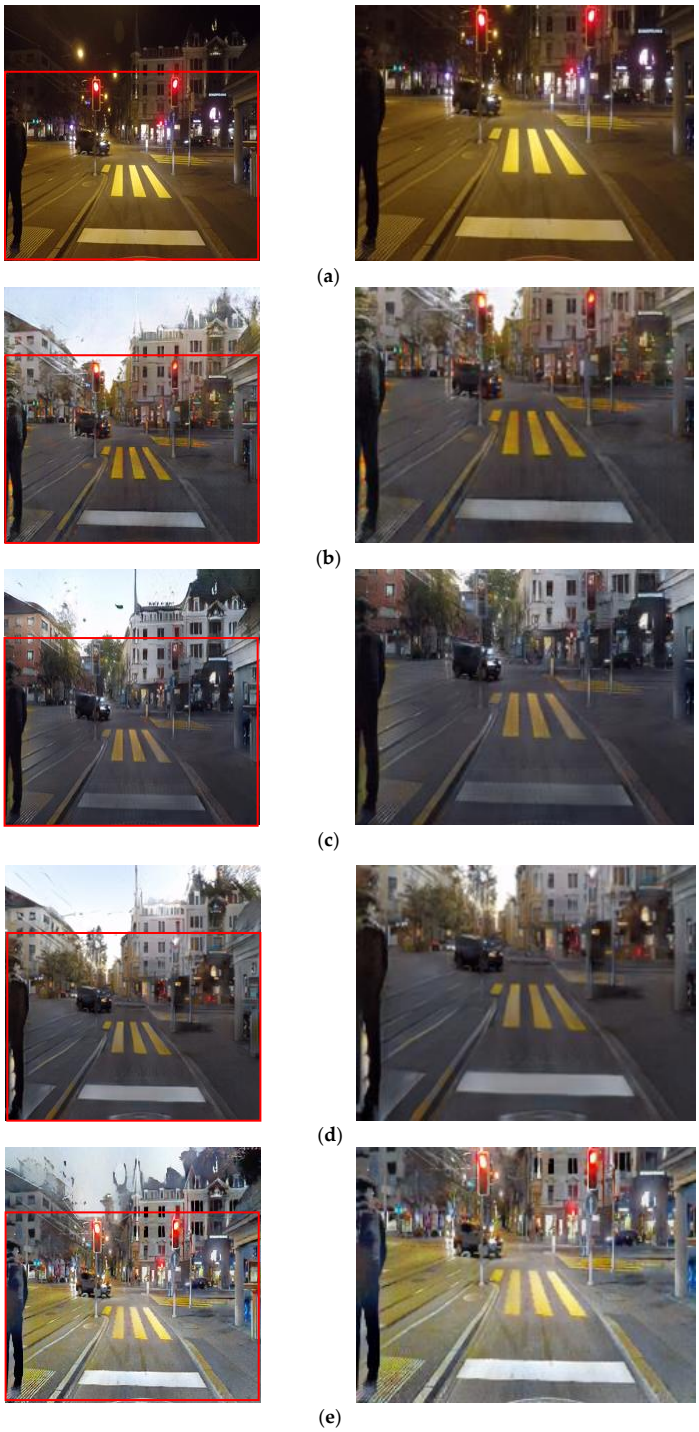

Figure 10. Night-to-day image translation comparison: (**a**) night image; (**b**) Pix2Pix; (**c**) CycleGAN; (**d**) ToDay GAN; (**e**) proposed method.

Figure 11 highlights an important observation regarding the ROI. In the result images obtained from conventional methods, such as the unpaired model, the car backlights are not accurately identified. This poses a potential risk in real-world scenarios, particularly when using these models for night-to-day transformation in car image sensors, as it may compromise safety and increase the risk of car accidents. Furthermore, the proposed method demonstrates sharper white lanes compared with other methods. When comparing the yellow crosswalk lanes, the proposed method exhibits the closest similarity to the night image among the existing methods. These findings emphasize the effectiveness and improved performance of the proposed method in accurately representing key elements and maintaining visual fidelity in crucial areas of the image.

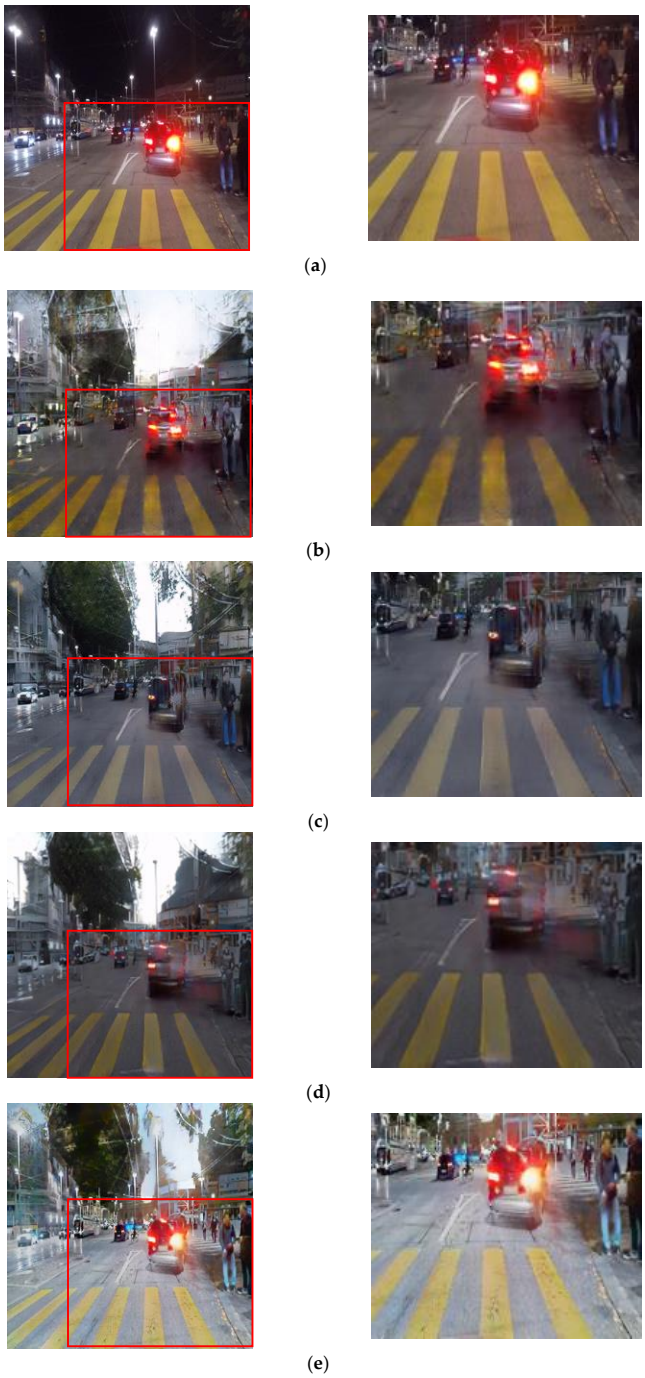

Figure 11. Night-to-day image translation comparison: (**a**) night image; (**b**) Pix2Pix; (**c**) CycleGAN; (**d**) ToDay GAN; (**e**) proposed method.

In Figure 12, it is evident that the proposed method excels in representing the color of the red car on the left, while other methods tend to display a blackish tone for the red color. Additionally, the proposed method accurately captures the color of the vehicle backlight, whereas other methods exhibit a noticeable reduction in backlight color. Moreover, the resulting images from the proposed method showcase a higher level of detail in lane markings, and even intricate features such as manhole covers are depicted with clarity. This enhanced ability to represent the road environment in finer detail demonstrates the superior descriptive capabilities of the proposed method compared with conventional methods.

Figure 12. Night-to-day image translation comparison: (**a**) night image; (**b**) Pix2Pix; (**c**) CycleGAN; (**d**) ToDay GAN; (**e**) proposed method.

In Figure 13, it can be observed that unpaired learning models such as unpaired CycleGAN and ToDayGAN struggle to accurately represent the red traffic signal and face difficulties in areas where white and yellow lane markings are mixed. In contrast, the resulting image from the proposed method effectively captures the information from the nighttime image, preserving the distinct colors of yellow and white lane markings. Other methods often depict the yellow markings as faded or incorrectly represent them as white markings. However, when representing areas outside the ROI, particularly the sky, conventional methods tend to depict the sky more accurately than the proposed method. The proposed method, in contrast, excels in representing objects within the driver's attentive field rather than prioritizing areas such as the sky or regions outside the ROI.

Figure 13. Night-to-day image translation comparison: (**a**) night image; (**b**) Pix2Pix; (**c**) CycleGAN; (**d**) ToDay GAN; (**e**) proposed method.

In Figure 14, the proposed method demonstrates excellent representation of the backlight area and accurately captures the color of the road lines based on the information from the nighttime image. Moreover, the proposed method effectively depicts the signs around the road, outperforming other methods in this aspect. Overall, the proposed method exhibits superior performance and would be a suitable choice for installation as a day-to-night conversion module in a vehicle sensor module. Its ability to accurately represent objects, traffic lights, and vehicle backlights renders it highly advantageous for ensuring safe driving through vehicle sensors.

Figure 14. Night-to-day image translation comparison: (**a**) night image; (**b**) Pix2Pix; (**c**) CycleGAN; (**d**) ToDay GAN; (**e**) proposed method.

3.3. Metrics and Score Comparisons

The results of each module were objectively compared using five image assessment metrics, including four no-reference image quality metrics and one image sharpness metric. The design of the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) algorithm is based on the observation that a distorted natural scene will lead to the distortion of the statistics of the image pixels [22]. When an image is processed using mean subtraction and contrast normalization to represent it as a Gaussian distribution, the increase in image distortion leads to a corresponding deviation from the shape of a Gaussian distribution. Based on this observation, BRISQUE is defined as a metric that determines how close an image is to a natural scene by assessing the level of distortion in the image. BRISQUE extracts features such as brightness, contrast, and texture, among others, from an image and evaluates them. A lower BRISQUE score indicates a better quality image.

The Natural Image Quality Evaluator (NIQE) is based on the Natural Scene Static model. It utilizes a pretrained model to extract statistical features from an image and measures the quality of the image based on these features [23]. NIQE examines the average characteristics of clean, undistorted images and compares these characteristics with those of a test image. A closer match between the characteristics of the test image and the reference clean images indicates a higher quality image. A lower NIQE score indicates a higher quality image.

The Contrast-Enhancement-Based Contrast-Changed Image Quality measure (CEIQ) is a quality metric that addresses contrast distortion, which is often overlooked by traditional no-reference metrics [24]. It specifically focuses on measuring the quality of images considering contrast changes. The test image is histogram equalized, after which the structural similarity index (SSIM) is used to measure the structural similarity between the original test image and the equalized image. Additionally, the histogram entropy and cross-entropy between the original test image and the equalized image are computed. Histogram entropy represents the diversity of color distribution within an image, whereas cross-entropy measures the difference in histogram distribution between two images. By utilizing these entropies, the evaluation of color distribution changes can be conducted as a result of image enhancement. Finally, the quality score of the image is inferred by feeding the extracted features into a support vector machine (SVM) regression module, which has been trained using the obtained scores. By combining all the features discussed so far, the overall quality score of the image is evaluated. In the case of CEIQ, a higher score indicates better image quality.

Michelson contrast [25] is an index that utilizes the difference between the highest and lowest brightness values in an image. It provides a quantitative measurement of how well the bright and dark areas are distinguished from each other. A higher contrast value indicates a larger intensity difference, resulting in sharper boundaries and improved visual perception of details in the image.

S3 (Spectrum and Spatial Sharpness) calculates a map of perceived sharpness, with higher values indicating more perceptually sharper areas [26]. This metric considers both the sharpness in the spatial and spectral domain (frequency domain) of the image to evaluate its quality. S3 is used to assess the tradeoff between sharpness enhancement and loss. A higher S3 score indicates a sharper image that closely resembles the original.

In the proposed method, SLAT processing is applied during base-detail training. Consequently, the images generated by the base-detail model exhibit increased local contrast and enhanced sharpness compared to the results obtained using conventional methods. These improvements are visually apparent.

Table 1 compares the existing methods, Pix2Pix, CycleGAN, and TodayGAN. Additionally, the proposed post-processing techniques, the Stevens effect and the local blur weight map, are applied to the TodayGAN method for comparison. The results of the base-detail model, which excludes the post-processing component of the proposed method, are also compared. Observations show that the base-detail model, without the post-processing step, achieves higher scores in the contrast distortion metric (CEIQ), local contrast measure-

ment (Michelson contrast), and sharpness metric (S3) compared to the existing methods. However, the base-detail model performs lower in metrics related to naturalness and color distortion (BRISQUE and NIQE) compared to the proposed method. Moreover, the proposed method demonstrates superior scores compared to the existing methods. Therefore, the proposed post-processing techniques effectively reduce noise while enhancing the naturalness of the images. In conclusion, as intended in this paper, the proposed method improves the sharpness of objects in images and enhances the overall quality of the images.

Table 1. Comparison of average image assessment scores (up arrow: higher is better; down arrow: lower is better).

Metric	Pix2Pix	CycleGAN	ToDayGAN	ToDayGAN + Post-Processing	Base-Detail Model	Proposed Method
BRISQUE↓	20.755	19.879	25.246	20.814	25.92	19.796
NIQE↓	3.617	3.814	3.974	4.061	4.079	3.594
CEIQ↑	3.006	3.001	2.989	2.943	3.387	3.338
Michelson contrast↑	0.969	0.97	0.969	0.942	0.999	0.987
S3↑	0.219	0.248	0.169	0.166	0.326	0.267

4. Discussions

The proposed method enhances local contrast and detail information during the transformation of nighttime to daytime images. Figures 9 and 12 demonstrate that it accurately preserves the colors of important traffic signals without distortion, unlike unpaired learning modules such as CycleGAN and ToDayGAN, which suffer from color distortion due to the limitations of unpaired training. Additionally, as shown in Figures 10 and 11, the proposed method produces sharper representations of vehicle backlights compared with conventional methods, enabling better identification of the situation of preceding vehicles and reducing the risk of accidents. In Figure 13, it is expected that the proposed method will perform well in rendering objects such as signs that are challenging to see at night, as demonstrated by other methods. Moreover, Figures 9–13 illustrate that the proposed method excels in detailing road lines, manholes, pedestrian crossings, vehicles, and pedestrians, enhancing the visibility of objects that are typically difficult to perceive during the nighttime. This capability is expected to significantly assist drivers by providing additional visual information. The proposed method specifically focuses on capturing the road situation from the driver's perspective and prioritizes the depiction of objects on the road. Consequently, the research emphasizes the rendering capabilities of objects on the road. However, compared with conventional methods, there are certain issues, such as the inclusion of trees in the sky area or the presence of various noise artifacts. These issues arise from the attempt to transform unseen areas in nighttime images into trees through the learning process.

Additionally, it takes 0.18 s on average for night-to-day image translation in the condition of 256 × 256 resolution, and the post-processing time is about 0.63 s per image. Therefore, the total processing time required for conversion is about 0.81 s on average. Consequently, it may pose some challenges to implementing it in real time on a vehicle. Future research should aim to improve the representation of objects over the driver's entire viewing area, including the sky area. Future studies should also include performing night-to-day video conversion in real time and exploring and tracking ROI objects in the converted video stream. The post method plans to optimize the algorithm to reduce the processing time to milliseconds, enabling seamless real-time video translation and facilitating object tracking. This would allow for efficient integration of the proposed method in real-world applications.

5. Conclusions

This study focuses on the translation of nighttime images to daytime images. To leverage the benefits of paired training, the proposed method generates synthetic nighttime

images for organizing a paired dataset through unpaired training. The training process involves using fake nighttime images as the base images and enhancing the detail and local contrast of the converted daytime images by incorporating a base-detail model. The resulting images from the base-detail model may contain noise that is addressed by employing a local blur weight map to reduce noise while preserving details. Additionally, to improve the details of nighttime objects in the converted daytime images while considering human visual characteristics, the Stevens effect is applied as a post-processing step. This enhances the detail information of nighttime objects in the converted images. One notable drawback of conventional methods is the removal or diminishment of backlights and lighting components in daytime scenes, which can significantly impact the quality of the converted images. However, the proposed method successfully preserves important elements in the night-to-day image conversion, including the clear representation of traffic signals, road signs, and the backlights of preceding vehicles, without distortion. Consequently, this research enables the conversion of nighttime images to daytime images from the perspective of a car driver during nighttime driving. This provides drivers with sufficient visibility, ultimately facilitating the prevention of accidents while driving.

Author Contributions: Conceptualization, S.-H.L.; methodology, S.-H.L. and D.-M.S.; software, D.-M.S.; validation, S.-H.L. and D.-M.S.; formal analysis, S.-H.L. and D.-M.S.; investigation, S.-H.L. and D.-M.S.; resources, S.-H.L. and D.-M.S.; data curation, S.-H.L., H.-J.K. and D.-M.S.; writing—original draft preparation, D.-M.S.; writing—review and editing, S.-H.L.; visualization, D.-M.S.; supervision, S.-H.L.; project administration, S.-H.L.; funding acquisition, S.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Korea (NRF-2021R1I1A3049604) and supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C3007327).

Data Availability Statement: Dark Zurich Dataset [20]: https://www.trace.ethz.ch/publications/2019/GCMA_UIoU/. ACDC Dataset [21]: https://acdc.vision.ee.ethz.ch/. All dataset is made freely available to academic and non-academic entities for non-commercial purposes.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pang, Y.; Lin, J.; Qin, T.; Chen, Z. Image-to-Image Translation: Methods and Applications. *IEEE Trans. Multimed.* **2021**, *24*, 3859–3881. [CrossRef]
2. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *arXiv* **2014**, arXiv:1406.2661.
3. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
4. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv* **2016**, arXiv:1611.07004.
5. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
6. Srivastava, A.; Valkov, L.; Russell, C.; Gutmann, M.U.; Sutton, C. VEEGAN: Reducing Mode Collapse in GANs Using Implicit Variational Learning. *arXiv* **2017**, arXiv:1705.07761.
7. Taigman, Y.; Polyak, A.; Wolf, L. Unsupervised Cross-Domain Image Generation. *arXiv* **2016**, arXiv:1611.02200.
8. Arruda, V.F.; Paixão, T.M.; Berriel, R.F.; De Souza, A.F.; Badue, C.; Sebe, N.; Oliveira-Santos, T. Cross-Domain Car Detection Using Unsupervised Image-to-Image Translation: From Day to Night; Cross-Domain Car Detection Using Unsupervised Image-to-Image Translation: From Day to Night. *arXiv* **2019**, arXiv:1907.08719.
9. Anoosheh, A.; Sattler, T.; Timofte, R.; Pollefeys, M.; Van Gool, L. Night-to-Day Image Translation for Retrieval-based Localization. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019. [CrossRef]
10. Anoosheh, A.; Agustsson, E.; Timofte, R.; Van Gool, L. ComboGAN: Unrestrained Scalability for Image Domain Translation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 896–8967. [CrossRef]
11. Kuang, J.; Johnson, G.M.; Fairchild, M.D. iCAM06: A refined image appearance model for HDR image rendering. *J. Vis. Commun. Image Represent.* **2007**, *18*, 406–414. [CrossRef]

12. Choi, K.; Byun, G.; Kim, A.; Kim, Y. Drivers' Visual Perception Quantification Using 3D Mobile Sensor Data for Road Safety. *Sensors* **2020**, *20*, 2763. [CrossRef] [PubMed]
13. Tönnis, M.; Sandor, C.; Klinker, G.; Lange, C.; Bubb, H. Experimental Evaluation of an Augmented Reality Visualization for Directing a Car Driver's Attention. In Proceedings of the Fourth IEEE and ACM International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2005, Vienna, Austria, 5–8 October 2005; Volume 2005, pp. 56–59.
14. Andersen, G.J.; Sauer, C.W. Optical information for car following: The driving by visual angle (DVA) model. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2007**, *49*, 878–896. [CrossRef] [PubMed]
15. Jin, S.; Wang, D.-H.; Huang, Z.-Y.; Tao, P.-F. Visual angle model for car-following theory. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 1931–1940. [CrossRef]
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 630–645.
18. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Smolley, S.P. Least Squares Generative Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2813–2821. [CrossRef]
19. Kwon, H.-J.; Lee, S.-H.; Lee, G.-Y.; Sohng, K.-I. Luminance adaptation transform based on brightness functions for LDR image reproduction. *Digit. Signal Process.* **2014**, *30*, 74–85. [CrossRef]
20. Sakaridis, C.; Dai, D.; Van Gool, L. Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *arXiv* **2019**, arXiv:1901.05946.
21. Sakaridis, C.; Dai, D.; Van Gool, L. ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10745–10755. [CrossRef]
22. Sun, T.; Zhu, X.; Pan, J.-S.; Wen, J.; Meng, F. No-Reference Image Quality Assessment in Spatial Domain. In *Genetic and Evolutionary Computing*; Advances in Intelligent Systems and Computing; Springer: Cham, Swizterland, 2015; pp. 381–388. [CrossRef]
23. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]
24. Yan, J.; Li, J.; Fu, X. No-Reference Quality Assessment of Contrast-Distorted Images Using Contrast Enhancement. *arXiv* **2019**, arXiv:1904.08879.
25. Michelson, A.A. *Studies in Optics*; The University of Chicago Press: Chicago, IL, USA, 1962.
26. Vu, C.T.; Chandler, D.M. S3: A Spectral and Spatial Sharpness Measure. In Proceedings of the 2009 First International Conference on Advances in Multimedia, MMEDIA 2009, Colmar, France, 20–25 July 2009; pp. 37–43.