Article

Enhanced Tone Mapping Using Regional Fused GAN Training with a Gamma-Shift Dataset

Sung-Woon Jung, Hyuk-Ju Kwon * and Sung-Hak Lee

School of Electronic and Electrical Engineering, Kyungpook National University, 80 Deahakro, Buk-Gu, Daegu 702-701, Korea; iymo@knu.ac.kr (S.-W.J.); shak2@ee.knu.ac.kr (S.-H.L.)
* Correspondence: olin1223@ee.knu.ac.kr

Abstract: High-dynamic-range (HDR) imaging is a digital image processing technique that enhances an image's visibility by modifying its color and contrast ranges. Generative adversarial networks (GANs) have proven to be potent deep learning models for HDR imaging. However, obtaining a sufficient volume of training image pairs is difficult. This problem has been solved using CycleGAN, but the result of the use of CycleGAN for converting a low-dynamic-range (LDR) image to an HDR image exhibits problematic color distortion, and the intensity of the output image only slightly changes. Therefore, we propose a GAN training optimization model for converting LDR images into HDR images. First, a gamma shift method is proposed for training the GAN model with an extended luminance range. Next, a weighted loss map trains the GAN model for tone compression in the local area of images. Then, a regional fusion training model is used to balance the training method with the regional weight map and the restoring speed of local tone training. Finally, because the generated module tends to show a good performance in bright images, mean gamma tuning is used to evaluate image luminance channels, which are then fed into modules. Tests are conducted on foggy, dark surrounding, bright surrounding, and high-contrast images. The proposed model outperforms conventional models in a comparison test. The proposed model complements the performance of an object detection model even in a real night environment. The model can be used in commercial closed-circuit television surveillance systems and the security industry.

Keywords: gamma shift; generative adversarial network (GAN); tone mapping; regional fusion training (RFT); regional weighted loss (RWL)

1. Introduction

Digital cameras and display equipment have a more limited dynamic range than human vision. As a result, when shooting in a high-contrast outdoor scene, oversaturation or undersaturation occur in some areas due to the limited dynamic range. To overcome such a limited dynamic range and create an image similar to human vision, a high-dynamic-range (HDR) image is required. Debevec et al. [1] proposed a method for creating HDR images using multiple images with different exposures. This method is used for multiple images with different exposure levels to obtain a camera response function (CRF) and generate an HDR image using them. Kwon et al. [2] proposed a method for synthesizing an HDR image using spatial and intensity weighting using two exposure images. These conventional methods require several low-dynamic-range (LDR) images to generate an HDR image, and ghosting artifacts occur when a moving object exists in an image. In addition, cameras capable of shooting HDR images are expensive. Therefore, alternative methods have been proposed for inferring an HDR image from a single LDR image, such as inverse tone mapping (iTM) [3]. Initial iTM methods can improve LDR images, either using individual heuristics or optionally using manual intervention [3–5]; however, the HDR output is frequently different from actual HDR images because these methods

cannot sufficiently compensate for missing information due to sensor saturation, color quantization, and the nonlinearity of CRF [6]. To obtain high-quality HDR images, convolutional neural network (CNN)-based iTM methods have been proposed [6–9]. CNNs have been used in many computer vision applications; CNN-based iTM methods have demonstrated significant performance improvements. Recently, with state-of-the-art techniques, studies related to HDR image conversion have shown a good performance. Li et al. [10] developed a fast multiscale structural patch decomposition to demonstrate how multi-exposure can be adjusted to reduce halo artifacts. Cai et al. [11] used CNN learning as a single image contrast enhancement method; the authors constructed a large-scale multi-exposure image dataset containing 589 high-resolution multi-exposure sequences, which were finely selected with 4413 images. They generated contrast enhancement images for each sequence using 13 representative multi-exposure image fusion and stack-based HDR imaging algorithms and conducted subjective experiments to find the best model for each scene. Wang et al. [12] introduced a method for training neural networks using an illumination map in underexposed photo enhancement using deep illumination estimation. The results of restoring underexposed image pair data showed that image-to-light mapping was learned (instead of image-to-image mapping), leveraging the simplicity of lighting in natural images so that the networks could effectively learn various photo adjustments. To address the problems raised by the preceding cases, we considered the sue of generative adversarial networks (GANs).

GANs can generate new images using a competing network of generators and discriminators and have been proven to have the ability to not only increase image resolution but also convert them to new images suitable for use in various fields, such as underwater photo transformation, art, culture, security, and the military [13]. GANs trained for image-to-image translation can map parts of an image to a target image via methods such as texture conversion [14]. When converting an input image to a target image, the expected result image is mapped to the RGB channel of the input image by referring to the target image's pattern.

However, it is difficult to obtain sufficient numbers of paired-image datasets using image-to-image conversion. The Berkeley AI Research laboratory's CycleGAN tackled this problem by adding a cycle consistency loss, which feeds back generated images via a reverse function to ensure transitive parity with their source images [15]. Consequently, the main image-to-image conversion problem—the difficulty in obtaining target images with spatial correspondence to input images—was effectively tackled. In addition, CycleGAN produces efficient training results using relatively small amounts of data.

An HDR image conversion algorithm takes images of various exposure values (EVs) or expresses all image information that can be visually recognized by humans using a local area [16]. However, after the training module is developed, a GAN can perform HDR tone mapping even on single image inputs. The model proposed in this study creates an HDR image using a CycleGAN with information obtained from a single image. However, the case of HDR image tone mapping is not straightforward. The conventional CycleGAN trains on the RGB channel [17,18], and thereby distorts the color and degrades the image quality. Therefore, correction is required when using a conventional GAN module, resulting in difficulty in HDR image generation. The proposed model mitigates this problem by introducing several techniques.

First, the proposed model exclusively trains on a luminance channel to reduce color distortion or noise; then, a dataset is augmented with a gamma shift to learn various tone levels.

Second, a regional weighted loss (RWL) performs training for tone compression in the local area of images. The proposed weight map uses a Gaussian filter, which helps in distinguishing the dark and bright areas of an image. As the dark and bright areas vary depending on the value of the gamma shift, the weight function can be used differently based on the gamma value.

Finally, we propose the use of a regional fusion training (RFT) method designed for providing a difference in restoration speed according to the brightness of a region [17,18]. The restoration speed is the degree of improvement in the local bright and dark areas of a training image. This method learns a target image's appropriate luminance range value and can enhance image quality more than existing methods. RFT, similar to RWL, is affected by the gamma shift value, resulting in the function being modified, and the input images' color values are retained as the resulting images' color values. In addition, when inputting images into the completed module, global luminance-level tuning is used to learn the biased average luminance level from that of the target image.

The proposed model's local contrast restoration performance is better than that of conventional models. A positive effect is also observed on the results of foggy, dark surround, and high-contrast images. The proposed model can incorporate the proposed HDR GAN into Mask R-CNN for improved detection performance in dark environments [19]. Moreover, closed-circuit television (CCTV) surveillance systems using this model can enhance their real-time image detection.

2. Background

2.1. Inverse Tone Mapping

iTM is a method that converts a single LDR image into an HDR image. It expands the contrast range of an LDR image using a single LDR image with missing information. Missing information arises due to the limited dynamic range of sensors, under- or over-sensor saturation, the nonlinearity of the CRF, and quantization. Therefore, reconstructing an HDR image from missing information is an ill-posed problem, and it is difficult to obtain a high-quality HDR image. Initial iTM models used the density of a light source to expand dynamic ranges. Banterle et al. [3] used a linear interpolation of an original LDR image and an expanded LDR image using an expand map. They used the inversion of the global photographic tone reproduction operator to generate the expanded LDR image. The expand map is a weighting map that estimates the highlight regions and avoids blocky transition between the regions when generating an HDR image. Rempel et al. [4] proposed a model that could improve the range of legacy videos and photographs for viewing HDR displays. This model converts a nonlinear LDR image into an HDR image using inverse gamma and contrast scaling. Then, brightness enhancement and an edge-stopping function are used to compensate for the saturated regions. Huo et al. [5] proposed an iTM model based on the human visual system. It adopted the retina response of the human visual system and used an LDR image to achieve local luminance. The LDR image and local luminance were used in the inversed local retina response to reconstruct an HDR image. However, the above iTM models still exhibit the problem of being ill-posed due to insufficient information (data), and the HDR images resulting from these models have limited quality.

To obtain high-quality HDR images, CNN-based iTM models have been proposed. Eilertsen et al. [7] proposed HDRCNN, which focuses on recovering missing information in overexposed areas using a hybrid dynamic range autoencoder while ignoring underexposed areas. Endo et al. [6] proposed deep reverse tone mapping to mitigate the ill-posed problem that occurs when directly inferring a 32-bit HDR image from an 8-bit LDR image. The core of this model is to composite LDR images shot with different exposures based on supervised learning and merge them to reconstruct an HDR image. A three-dimensional deconvolution network was used to learn relative changes in pixel values according to exposure changes. Marnerides et al. [8] proposed ExpandNet, which consists of global, local, and dilation branches. The global branch has a pyramid-shaped structure that gradually decreases from the input to output layers to reduce the dimension of an input image and obtain abstract features. The local and dilation branches provide localized processing without down-sampling to capture high-frequency and neighboring features. Liu et al. [9] proposed three specialized CNNs to recover an HDR image from an LDR image: Dequantization-Net for reducing noise and contouring artifacts, Linearization-Net

for estimating the CRF and converting nonlinear LDR images into linear irradiance, and Hallucination-Net for recovering missing contents caused by dynamic range clipping.

2.2. Generative Adversarial Network

A basic GAN has a unique structure, wherein two neural networks—a generator and a discriminator—compete [13]. The generator takes a source image and noise as random inputs and mimics the source image as a feature of the target image, whereas the discriminator compares the target image with the generated image. The generator attempts to create a fake image similar to the target image to deceive the discriminator. At this time, if the discriminator and generator compete, the discriminator will outperform the generator. Therefore, the discriminator's training performance should be fixed at a certain level. In addition, the discriminator should have a fully connected classifier at the end. A CNN or long short-term memory can be used as either the generator or discriminator [20,21]. This competitive structure is expressed as an adversarial loss, which becomes an objective function [13]. The purpose of GAN training is to reduce the generator loss, which is given by:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[log\, D(x)] + \mathbb{E}_{z \sim p_{data}(z)}[log\big(1 - D\big(G(z)\big)\big)], \tag{1}$$

where $D(\cdot)$ represents the discriminator function, x x represents the target image, z represents a random input, $G(\cdot)$ represents a function of the generator, and $\mathbb{E}$ represents the input dataset. Notably, as the generator's x mapping value increases, the discriminator's x mapping value decreases. The objective function is optimized as the value of the discriminant function increases. Therefore, the discriminator optimization function is given by:

$$D_{opt}(x) = \frac{p_{data}(x)}{p_{data}(x) + p_G(x)}, \tag{2}$$

where $p_{data}(\cdot)$ and $p_G(\cdot)$ have values between 0 and 1, meaning that the closer the discrimination result for G is to 0, the closer the value of $D_{opt}(\cdot)$ is to 1. Figure 1 shows the structure of the basic GAN. Generator $G(\cdot)$ receives a random image z and randomly generates an image. Discriminator $D(\cdot)$ receives both the generated image from $G(\cdot)$ and the target image x. To achieve the function objective x, the loss value of the discriminator (max_D) should be large; thus, the loss value of the generator (min_G) must be decreased. After determining max_D and min_G, the final objective function value $O(\cdot)$ is obtained.

Figure 1. Structure of a generative adversarial network diagram.

2.3. Pix2Pix

An image-to-image conversion model, known as pix2pix, is used [14] to overcome the limitations of the conventional GAN model. This network not only learns the mapping from the input to output images but also learns a loss function to train this mapping [2]. The aim is to work with two classes of images—A and B—such that A is chosen to be similar to B. The objective functions of pix2pix are as follows:

$$L1(G) = \big\|B - G_{fake}(A)\big\|_1, \tag{3}$$

$$G_{pix2pix} = O(G, D) + L1(G), \tag{4}$$

where $O(G, D)$ is the conventional GAN objective function value and $L1(\cdot)$ is the loss value between the target image B and the generated (fake) image $G_{fake}(\cdot)$. However, this model requires training images to be associated in pairs. For instance, in order to convert

a natural photographic image into a Monet painting style, a large amount of data drawn directly from the Monet style dataset is required for style description.

2.4. CylcleGAN

The CycleGAN approach enables the deep learning (DL) of unrelated image pairs by adding a cycle consistency loss incentive [15]. Smaller training datasets than those required by pix2pix are effective in coloring objects within images, whereas boundaries are also relatively well preserved. The goal of CycleGAN is to reduce the loss of cycle consistency [15,18].

The term "reconstruction" is used to refer to a generator's restoration of a source image when the generated image is given. A reconstruction function that can convert the source image into the target image and vice versa is given by:

$$L_{cyc}(G,F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x) - x)\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|F(F(y) - y)\|_1]. \tag{5}$$

$O_G(\cdot)$ represents the generator's objective function, where $G(\cdot)$ represents a forward-directed generator that creates a fake image of the target image y with the source image x and $F(\cdot)$ represents a reverse generator that converts the target image y into the input image x. This peculiar structure also changes the hostile loss function as follows:

$$O_G(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[log\, D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[log(1 - D_Y(G(x))], \tag{6}$$

where G and D_Y, are the generator and discriminator outputs. X and Y are the source and target data sets. x and y denote the source and target images, respectively. The reconstructor objective function of CycleGAN is given by:

$$O_F(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data}(x)}[log\, D_X(x)] + \mathbb{E}_{y \sim p_{data}(y)}[log(1 - D_X(F(y))]. \tag{7}$$

These reconstructions reduce the amount of data required for training while preserving the boundaries of the resulting images. The proposed model adopts this approach [22,23].

The generator and reconstructor model is called U-Net, and it uses a skip connection [7,15]. When an image is down-sampled and subsequently up-sampled, specific pixel information is lost. The skip connection directly transfers relevant information from the encoder to the decoder, resulting in a significantly sharp image, thereby enabling accurate predictions. When deployed, along with (5), (6), and (7), the final objective function of CycleGAN is as follows:

$$O_{cycle}(G, F, D_X, D_Y) = O_G(G, D_Y, x, y) + O_F(F, D_X, y, x) + \lambda L_{cyc}(G, F). \tag{8}$$

2.5. Exposure Value GAN

EV GAN uses image dataset with different exposure values [17,18]. This work trains the model to transform dark images into region-specific brightness values by inputting increasing EVs as EV-3, EV-4, and EV-5. EV GAN trains luminance channels in images and trains dark and bright regions of images using regional weight maps. For dark-area learning in the L channel of LDR images, regional weight maps derive a weighted loss map WB from the inverse Gaussian map in the loss function L for the target and reference images. The loss function then concentrates on the dark regions of the image using WB. The modified loss function is as follows:

$$W_B = 1 - N(F(B)), \tag{9}$$

$$L = ||W_B\,(A - B)\,||_1, \tag{10}$$

where a Gaussian filter of size 11 × 11 for a blurred or LDR reference image B is expressed as F(B); then, the image is normalized by $N(\cdot)$ and reversed. In addition, the RFT of EV GAN works to generate a target image by fusing the image being trained in reconstruction with the image trained in the previous step.

The purpose of EV GAN is to convert the training data into results with a good contrast performance by making the dark regions of dark images very bright and the relatively bright regions slightly brighter. The overall performance of EV GAN is good for dark images.

3. Proposed Model

3.1. Motivation

The conventional technique for creating an HDR image involves shooting two or more pictures with different areas of brightness in the same composition. Thus, every time an HDR image is created, the process must prepare data with different brightness information, which is difficult to accommodate in real-time videos. In addition, deep training models can be easily applied to real-time video after training is complete. Therefore, this study aims to propose a tone compression filter based on GANs to brighten the dark areas of an image and preserve bright areas, regardless of complex processes and data.

GAN models—deep training models—have proven performance in the image generation field [13–15,18,22–25]. The proposed model is based on GAN models. It involves gamma shift, RWL, and RFT. GANs require paired training data. Obtaining actual data for GAN training is difficult; however, virtual datasets similar to real datasets that can be used for experiments exist [26].

Pix2pix solved the color distortion problem with identity loss, but it is difficult to implement because training and target images must be paired, which requires a lot of data. The proposed model solved the problem of pix2pix based on CycleGAN.

The proposed model uses a gamma shift strategy for the L-channel dataset to train numerous tones, effectively reducing the color space distortion introduced by the conventional CycleGAN model. Initially, an experiment was conducted with EVs as an input to the neural network [18]. This model performed well for dark images, but bright images were distorted. Therefore, gamma shift was applied to input images to train wide information.

Next, the RWL was used to train the dataset's local brightness. This function was modified from EV GAN. In the conventional method, dark images are source images. Therefore, the RWL was implemented to enhance dark images. In the modified method, because an image with a gamma value less than 1 was added, RWL equations for bright and dark images were independently designed. This function primarily trained the brightness of the local area in the target data.

Finally, an RFT method was used to balance the restoration speed for each local domain. The RFT function was used to adjust the improvement of dark and bright areas to a certain level within the same epoch; it fused the results of the previous and current epochs. This method was advantageous for training the local tone of images.

First, source images were constructed using gamma shift. Next, each image's weighted loss map was obtained from the luminance channels of the source and target images. The above maps were used by the generator to train the local brightness values of both images. Then, the generator generated images. Finally, the restoration speed of local brightness was adjusted in the RFT step. The RGB test image was separated by luminance and chrominance channels. The luminance channel, to which mean gamma tuning was applied, was used for testing. The chrominance channel was a fusion obtained through training.

3.2. Gamma Shift Dataset for Tone Compression Generative Adversarial Network Training

This section describes how the gamma shift was applied to training datasets. Training HDR images requires various EVs. The gamma shift dataset is proposed to train a wide range of brightness information. In the conventional CycleGAN model, all color channels are used for training input, and the generator causes color distortion. However, unnecessary color information need not be trained. Training only the luminance space reduces the training time and avoids the color distortion problem. When the module was trained, only the luminance channel *L* was used, whereas *ab* color channels were maintained.

Gamma values of 1.2, 1.4, and 1.6 were used with similar brightness levels for the low-level dataset. Figure 2 is an example of a dataset used for training. Figure 2a–c show the target data, EV dataset, and gamma shift dataset, respectively. It is assumed that the partial brightness values of the improvement images to be learned are random. However, Figure 2b is increasingly dimly composed of EV-3, EV-4, and EV-5. Nonetheless, although there were bright areas, the amount of learning in these areas was expected. Figure 2c, on the other hand, is a dataset of the proposed model that has varying degrees of exposure in bright and dark areas, providing an advantageous environment for training networks to learn various brightness ranges.Its will be the appearance of over saturation in some parts of the image when learning bright areas by increasing the EV. Information is lost in the over saturated region. It causes color distortion in the image generated by CycleGAN.

That is to say that it has limitations in learning a wide range of luminance channels. Figure 3a shows that EV shift image information was irrecoverable in high-level luminance. When dark images are inputted, GANs attempt the output bright images due to their nature of attempting to create target images. Inputting a bright image into such a trained module saturates the bright area in the resulting image. As shown in Figure 3a, the EV is generated similarly to the expected image for low-level regions, but the resulting value does not match the expected image for high-level regions, with it instead losing information. Meanwhile, the proposed model learns a wide range of brightness exposure information with no information loss, which is appropriate for training on luminance maps and beneficial for training the local tone of images to approximate the expected image (Figure 3b). Therefore, training image tone compression over a wide luminance range with the gamma shift dataset was necessary.
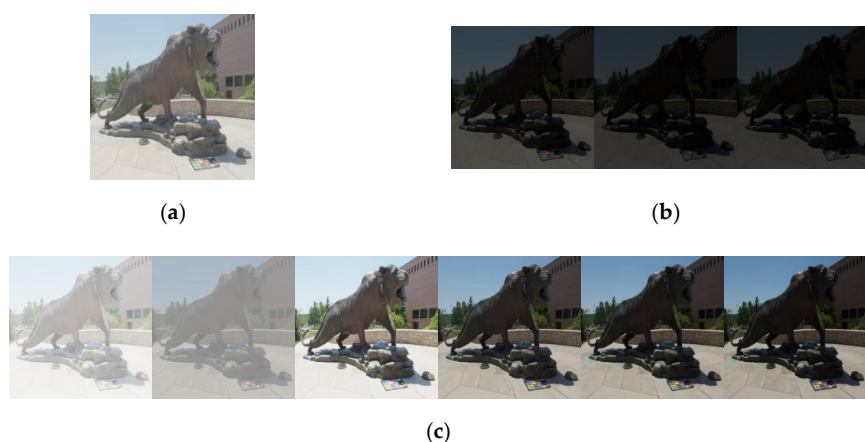

(**a**)

(**b**)

(**c**)

Figure 2. Data organization for the experiment: (**a**) source data, (**b**) exposure value (EV) data [18], and (**c**) gamma shift data.
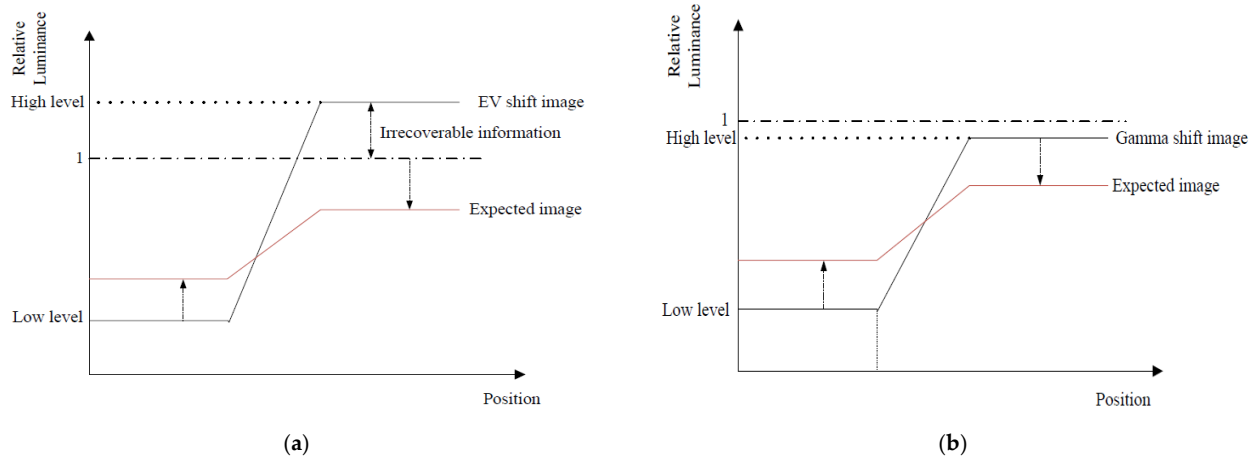
Figure 3. Difference in tone compression: (**a**) EV shift and (**b**) gamma shift.

3.3. Regional Weighted Loss Function

CycleGAN exhibits the phenomenon of changing the overall brightness tone. Therefore, a local weighted loss function for local tone training was proposed as the second training step. When the local tone area is trained from the L channel of a target image, the region of interest (ROI) comprises areas with a high local tone contrast. The RWL is the difference between the source and target images' local brightness. The luminance weight map of an image is devised using GF_i obtained using a Gaussian filter $g(\cdot)$ and normalization. Equation (12) obtains W_i to train the dark and bright areas according to the gamma values of input images:

$$GF(i) = N[g(i)], \tag{11}$$

$$W(i,\gamma) = \begin{cases} 1 - GF(i), if\ \gamma \geq 1 \\ GF(i),\ if\ 0 < \gamma < 1 \end{cases}, \tag{12}$$

where GF represents a Gaussian filter of 11×11 size, i represents the luminance of images, and $N(\cdot)$ represents the normalization function.

A weight value $W(\cdot)$ must be obtained according to the brightness level based on the gamma value of 1 because the training data are in the form of a gamma shift. γ represents the gamma value of the source image.

These were performed to reduce the spatial consistency and brightness difference between the target and generated images.

$$RWL_G(G, D_Y, x, y) = \mathbb{E}_{y \sim p_{data}(y)}\big[log\ D_Y\big(y \times W(y,\gamma)\big)\big] + \mathbb{E}_{x \sim p_{data}(x)}\big[log\big(1 - D_Y\big(G(x) \times W(y,\gamma)\big)\big], \tag{13}$$

where RWL_G represents the RWL of generator loss value and x and y represent the gamma shift source and target images, respectively.

$$RWL_F(F, D_X, y, x) = \mathbb{E}_{x \sim p_{data}(x)}\big[log\ D_X\big(x \times W(x,\gamma)\big)\big] + \mathbb{E}_{y \sim p_{data}(y)}\big[log\big(1 - D_X\big(F(y) \times W(x,\gamma)\big)\big], \tag{14}$$

where RWL_F represents the RWL of reconstructor loss value and y and x represent the gamma shift source and target images, respectively.

The role of the reconstructor is to restore a source image from a generated image. Thus, this function trains the weight map of x. The training direction is the opposite of the generator's training direction. Figure 4a depicts the target image, whereas Figure 4b,c depict the images of $W(y,\gamma)$ and $1 - W(y,\gamma)$ obtained from the generator, respectively. By optimizing the loss values between the luminance maps, a source image is learned from the luminance ROI in a target image. The proposed model demonstrated an improved image enhancement performance compared with existing models. However, the proposed model did not guarantee full tone compression performance. It did not perfectly

regulate the regions that increased and decreased brightness. Therefore, it is necessary to balance the bright region restoration speed with the dark region restoration speed.

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

Figure 4. Example of the target luminance weighted map: (**a**) original image, (**b**) weighted map, and (**c**) inverse weighted map.

3.4. Regional Fusion Training

The resulting images should not only improve the saturated areas but also make dark areas brighter and maintain the tone of the existing bright areas. RFT was used to locally adjust the training rate of the generator and reconstructor [18]. The training was conducted by synthesizing the weight map's ROIs. Past and present information was fused and transferred to the next epoch to adjust the local tone compression of the image. This process should be defined separately according to the gamma value. The corresponding equation is as follows:

$$RFT(G,x) = W(y,\gamma) \times G_{n-1}(x) + (1 - W(y,\gamma)) \times G_n(x), \tag{15}$$

where $G(\cdot)$ represents the generator's result and $W(\cdot)$ represents a weight map for $G(\cdot)$ that is an ROI according to brightness. $W(\cdot)$ was used to fuse the value of the previous state $n-1$ with the current state n and pass the result as the $n-1$ value of the next training state, $n-1$. In addition, the reconstructor function must be modified. The conventional GAN training method retains the shape of the original image and uses paired datasets. This function improves training performance, even when a small amount of training data is used; the ROI of the weighted map must be set separately based on the gamma value in the reconstructor part. The corresponding equation is as follows:

$$RFT(F,y) = (1 - W(x,\gamma)) \times F_{n-1}(y) + W(x,\gamma) \times F_n(y), \tag{16}$$

where F_n and F_{n-1} represent the reconstructed images of the current and previous epoch, respectively. The final objective function is defined, including RWL and RFT, as follows:

$$O(G,F,D_X,D_Y) = RWL_F(F,D_X,y,x) + RWL_G(G,D_Y,x,y) + \lambda L_{cyc}(G,F) + O_F(RFT(F,y),D_X,Y,X) + \\ O_G(RFT(G,x),D_Y,X,Y), \tag{17}$$

where λ controls the relative importance of the two objectives.

3.5. Mean Gamma Tunning

The proposed model tended to learn high-level local contrast at the HDR target level. Mean gamma tuning was performed on the input image after the module was completed to derive efficient results:

$$i' = 255 \times \left(\frac{i_{test}}{255}\right)^{\frac{logc}{logMean}}, \tag{18}$$

where i_{test} is the input image of luminance and c represents a constant value, i' represents the image finally inputted into the module, and chrominance channels (*ab*) are compensated for in the resulting image.

3.6. Architecture

Figure 5 depicts the framework of the proposed model. First, from the gamma shift dataset, y is a target image corresponding to the source image x. Then, the RWL process is used to determine the loss of local brightness during the training process. At this point, the gamma shift dataset is divided into dark and bright images and the function employed depends on the γ value. This is handled in the formula for $W(\cdot)$. Next, we use RFT to train the difference in the local tone mapping improvement rates. The results from the generator and reconstructor are passed to the next epoch and fused with the current image. In addition, since the base is CycleGAN, the function is applied not only to generators but also to reconstructors. Since it can be divided into forward and reverse training, it can be divided into Figure 5a,b. Finally, a training module is used (Figure 6); the proposed model can be used in CCTV surveillance systems. When the training is completed and the module has been generated, the resulting image is derived using the mean gamma tuning image as an input according to the learning tendency. The role of the created module is to convert only the local luminance value so that the chroma channel is separated. After that, the luminance value of the input image is converted and the separated chroma channel is recombined. The resulting local mapping identifies objects that Mask R-CNN cannot identify.
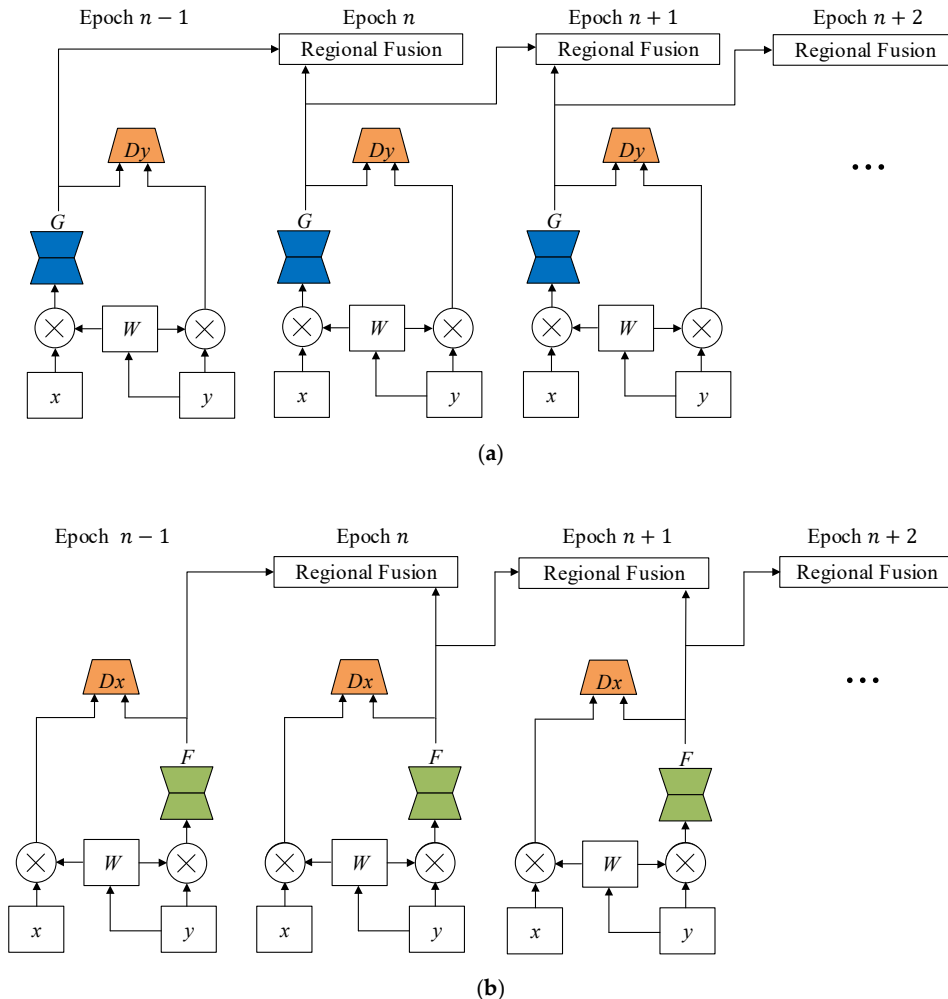
Figure 5. Architecture of the proposed training method: (**a**) generator and (**b**) reconstructor.
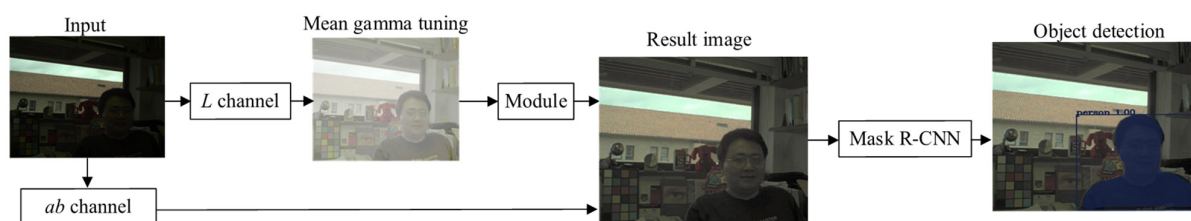
Figure 6. Mask R-CNN processing for closed-circuit television.

4. Experiment Results

4.1. Enviroment

In the experiment environment, the gamma values for bright tones were specified as 0.2, 0.4, and 0.6, whereas the dataset was expanded to gamma values of 1.2, 1.4, and 1.6 for dark tones. The dataset comprised 250 elements and a training number of 180 epochs. The resulting images were data not used for training. The equipment used in the experiment was an NVIDIA GeForce RTX 2080 Ti GPU, an Intel Core i7-9700 CPU, and a 32-GB RAM card. The size of the set of training data was 512 × 512 [27,28].

4.2. Step-by-Step Results of the Proposed Method

In this section, each step result of the proposed method is analyzed. Figure 7 displays the images resulting from the gamma shift to the RFT process in the proposed model section. Figure 7a is an input image and a dark environment, but the lighthouse correspond to a bright area. The brightness around the ground and trees is extremely dark, making it difficult to identify objects. Figure 7b displays the results of learning the gamma shift dataset using the proposed model. The image becomes brighter overall, but the brightness around the lighthouse and window also becomes oversaturated and the boundaries of the objects are very blurred, indicating that the network has difficulty in learning the local brightness of the image. We accordingly use the RWL to induce local tone mapping learning. Results, such as those shown in Figure 7c, which would have learned the gamma shift data and RWL process, are observed. The brightness of the light part remains unchanged, whereas the brightness of the dark part of the ground and the wood part increases.

However, from the tonal compression perspective, the brightness of the dark area should be slightly increased. The training module must allow the local training rate of the images to be regulated, so we used the RFT method as the next process. Figure 7d is the result of adding RFT, which is the final training process. The brightness of the lighthouse is significantly improved while maintaining its brightness due to RFT—a process that reduces the learning speed of the light region of the image while increasing the learning speed of the dark region.
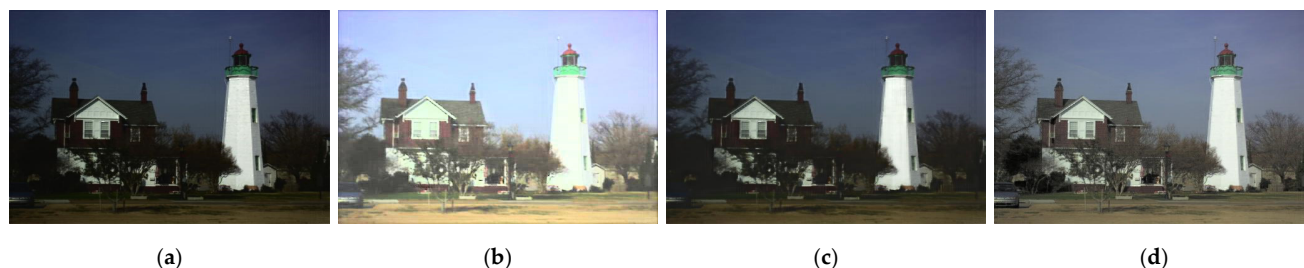

| (a) | (b) | (c) | (d) |

Figure 7. Part of each method's result: (**a**) input, (**b**) gamma shift augmentation, (**c**) gamma shift with regional weighted loss (RWL) result, and (**d**) gamma shift and RWL with regional fusion training result.

4.3. Comparison of Gamma Shift and Brigntness Augmentations

DL researchers employ data augmentation for efficient learning. The proposed model also used brightness augmentation when learning images. However, the brightness augmentation approach results in significant differences from ground truth, which is the aim of this study. Figure 8a shows the application of the typical brightness augmentation value to the proposed method instead of the gamma shift training dataset, whereas Figure 8b shows the proposed method with the gamma shift dataset. As a result, we showed that the contrast improvement and tone mapping performance were similar to that shown in Figure 8c.

(**a**)

(**b**)

(**c**)
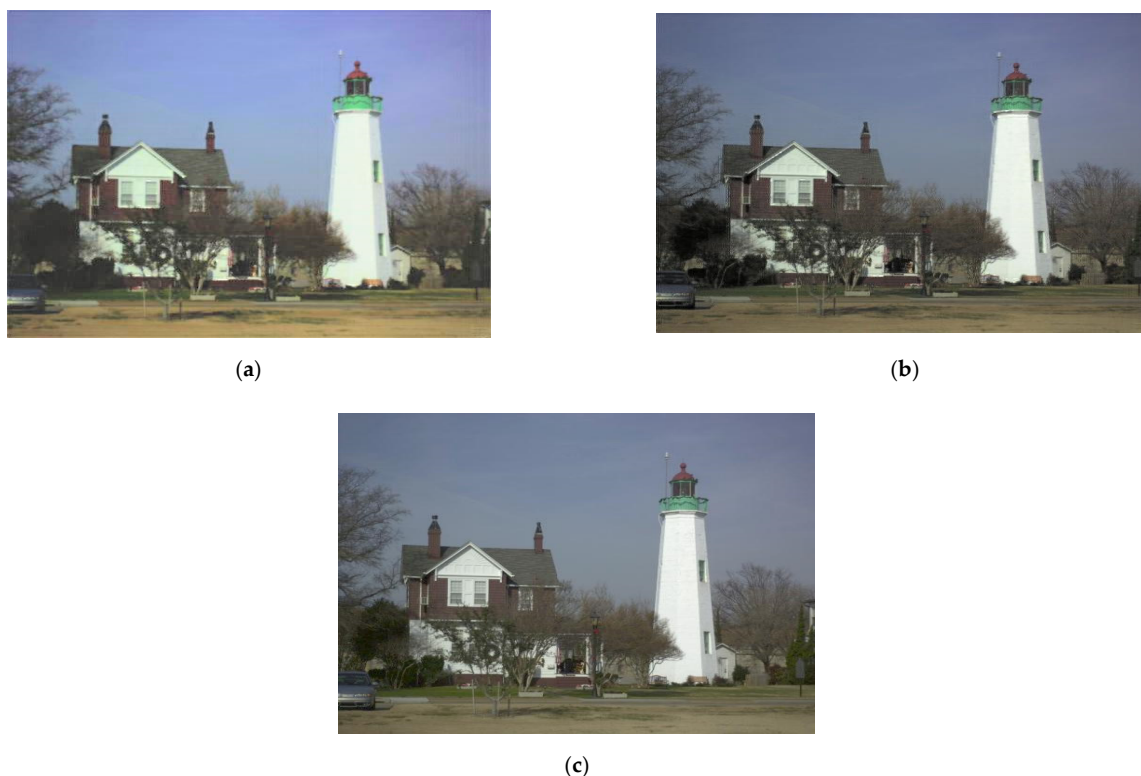
Figure 8. Comparison of general brightness augmentation and the proposed model's gamma shift augmentation: (**a**) general brightness augmentation, (**b**) gamma shift augmentation, and (**c**) ground truth.

4.4. Comparisons with Conventional Methods

We compared the results of the proposed method with those of five conventional methods, CycleGAN, Lum CycleGAN, EV GAN, HDRCNN, and ExpandNet. Figures 9 and 10 are foggy images and Figures 11 and 12 are high-contrast images. Figures 13 and 14 are the dark surrounding images inside the building and Figures 15 and 16 are the bright surrounding images outdoors.

Figure 9 is a satellite image of the sea and continent. In Figure 9c, CycleGAN shows color distortion and noise in the whole image. In Figure 9d, Lum CycleGAN shows that the sea and continent are darkened. In Figure 9e,f, EV GAN and HDRCNN show low contrast and desaturation, which makes it difficult to identify objects. In Figure 9g, ExpandNet shows the blocking noise. In the proposed method shown in Figure 9h, the fog is removed and the contrast is higher than that seen in other methods. The result is also similar to the target image of Figure 9b. Figure 10 is a satellite image of the port. The result of CycleGAN in Figure 10c shows that the sea and port colors are distorted to red and noise appears around the port. In Lum CycleGAN of Figure 10d, it is mostly dark except

for the port. The results of EV GAN and HDRCNN show a very low contrast in Figure 10e,f. The result of ExpandNet in Figure 10g is so dark that most of the information is lost. The proposed method shown in Figure 10h is restored so that ports and seas are well identified.

Figures 11 and 12 are high-contrast images. In Figure 11c, the color of the illuminated area is distorted. There is no change in Figure 11d. Figure 11e is brightened overall, but the clothes of the people are difficult to identify. Figure 11f has a low saturation. Figure 11g shows that the people in the shadows have lost their information. Figure 11h shows that all people and backgrounds are identifiable. In Figure 12c,d, the brightness is not changed. Additionally, the color of outside the window is distorted in Figure 12c. Figure 12e is brighter than the input image, but objects behind the person are dark. Figure 12f shows a low contrast. In Figure 12g, there is no improvement in brightness. In Figure 12h, the result of proposed method well enhances the brightness and contrast.

Figure 13 shows an image of dark surroundings on the inside of a building, and the goal of each model is to maintain the brightness of the bright ceiling and windows. In Figure 13c, the color of the dark area has become gray. There is little change in Figure 13d. Figure 13e is relatively bright, but there is no change in the dark area in the upper right corner of the image. Figure 13f is over brightened. Figure 13g features color distortion. In Figure 13h, the dark areas are brightened while maintaining the brightness of the windows and ceiling. Figure 14 shows a dark surrounding image of the inside of the building. The image has a chair and a stained glass. The goal of each model is to restore the brightness of the chairs, stained glass, and the inner walls of the building. In Figure 14c, the color of the stained glass becomes red. There is little brightness change in Figure 14d. Figure 14e shows the chair, but the left and right sides of the image are still dark. In Figure 14f, the brightness is increased, but the noise in the dark area is also increased. The brightness in Figure 14g is unchanged. In Figure 14h, the image generated by the proposed method shows that the chair is identifiable and the structure of the inner wall of the building is observable.

Figure 15 show an outdoor image of bright surroundings. The image shows a young girl and her reflection in the car's window and the background. The goal of each model is to brighten the shadows of Figure 15b and the girl's face, restore the reflection in the window, and maintain the brightness of the background. Figure 15c shows the color distortion of the vehicle and the girl's face turns gray. There is little change in Figure 15d. Figure 15e shows an oversaturated background. Figure 15f shows a low contrast and color distortion. Figure 15g is darker than the input image. In Figure 15h, the proposed method preserves the color of the input image well and has a high contrast. Figure 16 is an image of bright surroundings. The goal of each model is to restore detail in the trees and restore the brightness inside the white pillars of the building and the ground. Figure 16c shows the color distortion of the pillars of the building. Most of the tree detail information is lost. There is no improvement in brightness in Figure 16d. Figure 16e is similar to the ground truth, but some noise is observed. Figure 16f is saturated. Figure 16g is darker than the input image. In Figure 16h, the proposed method has less color distortion and noise and a better local contrast than the image produced using conventional methods.

Unlike Lum CycleGAN, EV GAN, and the proposed models, it seemed that CycleGAN, which learned RGB, trains all channels, even with the same number of training epochs, resulting in underfitting. In Lum CycleGAN, the color distortion problem is resolved and fog images are resolved to some extent, but no learning of local tone mapping is performed, indicating that the training network did not learn about chroma and required more training. The reason for this problem was that EV GAN, similar to the proposed model, also proceeded with learning for local tone mapping but used a limited set of training datasets, which illuminated dark images. Meanwhile, in the proposed model, we use the RWL and RFT of EV GAN differently depending on the gamma value of the training image to perform local tone mapping with a wide range of luminance. Compared with state-of-the-art technology, the results of ExpandNet are barely changed. The overall

results are noisy and have a low contrast. The results of HDRCNN show very low saturation and contrast.
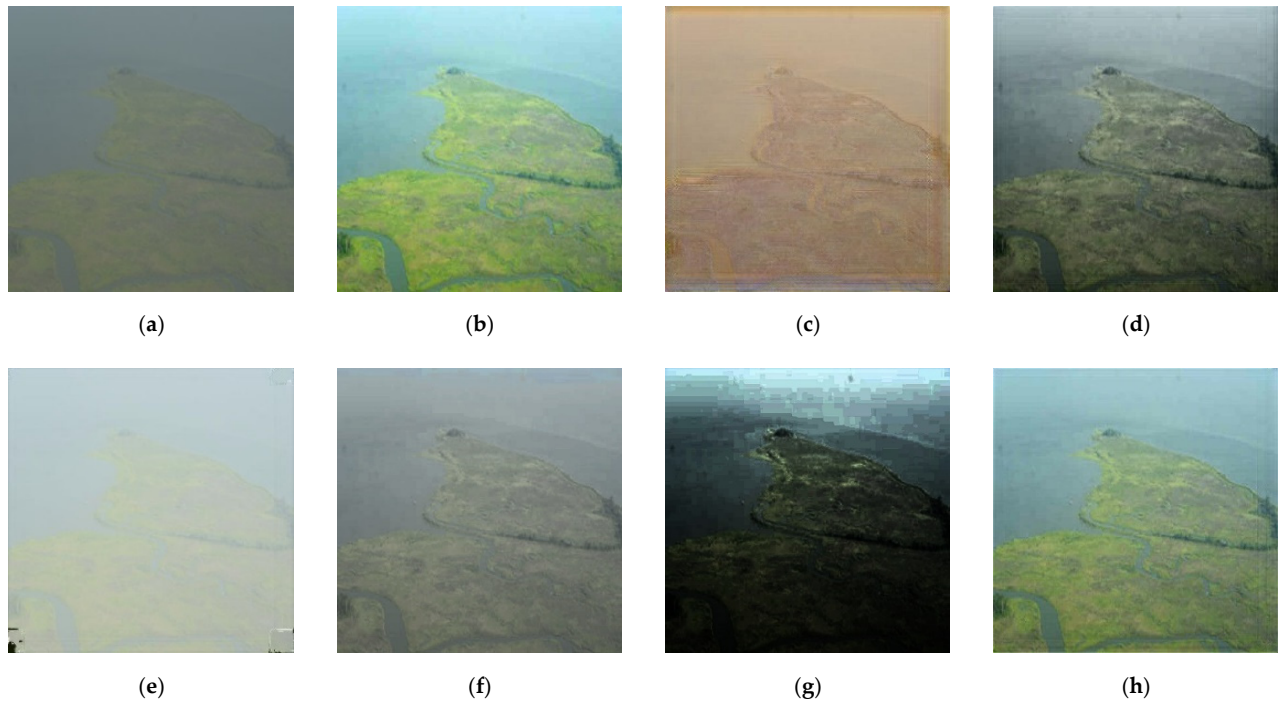

Figure 9. Foggy image #1: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.
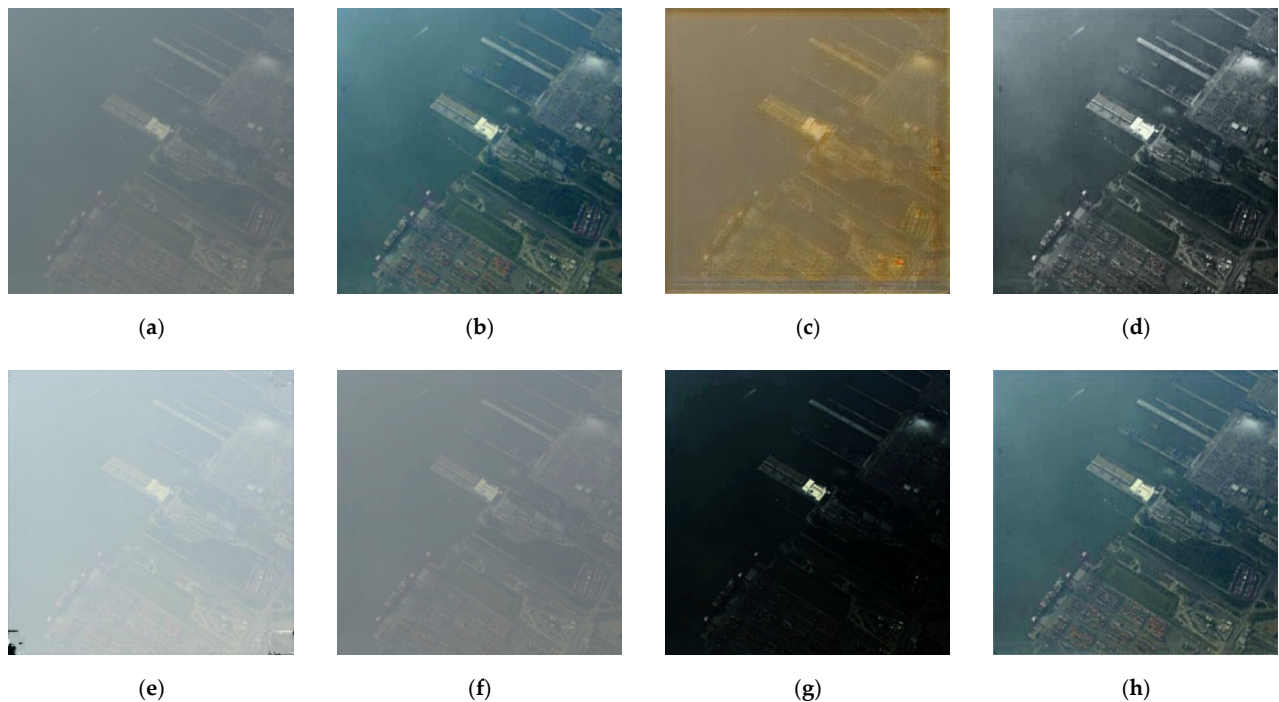
Figure 10. Foggy image #2: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.
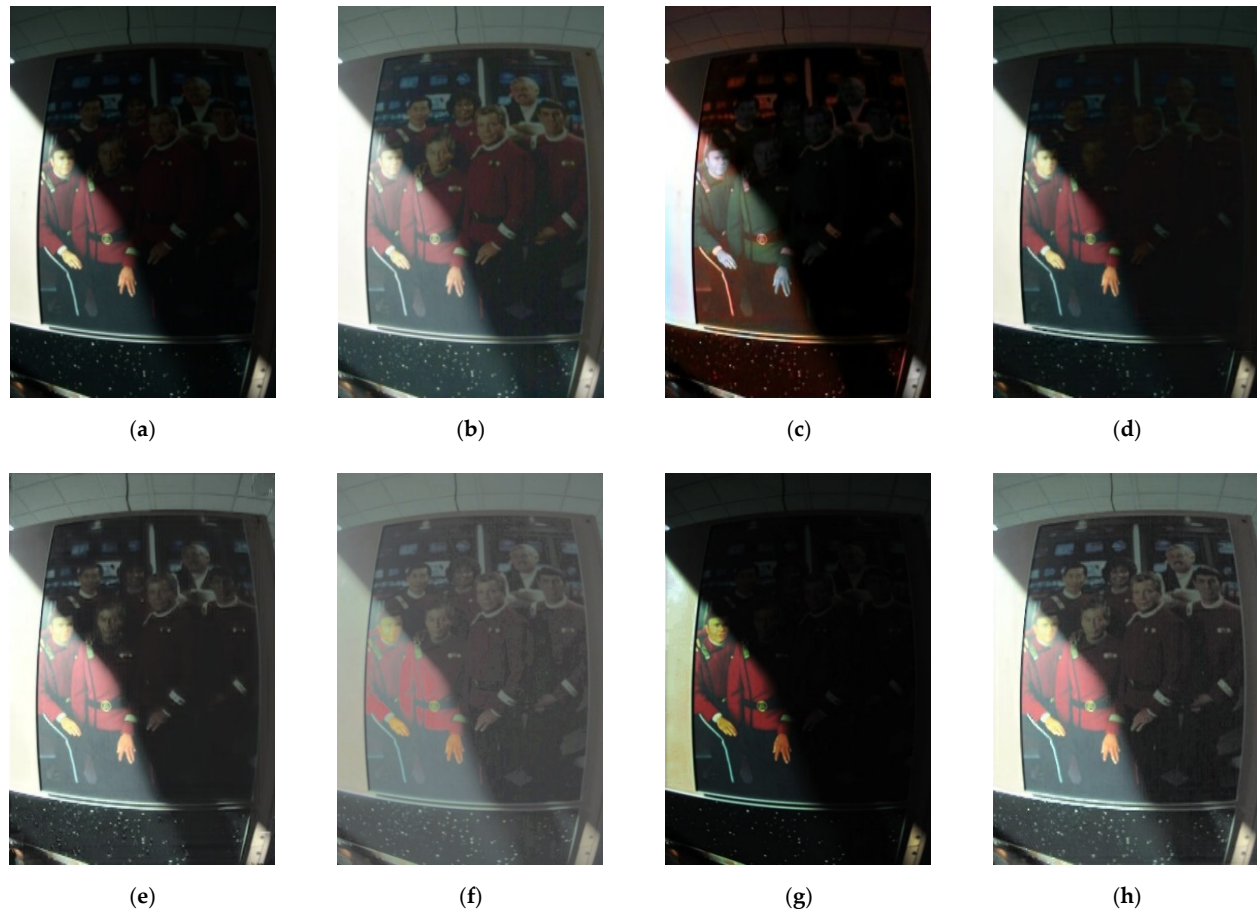
Figure 11. High-contrast image #1: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.

To compare the quantitative performance results, we used four evaluation methods—namely, a perception-based image quality evaluator (PIQE) [29]multiscale structural similarity (MS-SSIM) [30], HDR-VDP2 quality, and HDR-VDP2 visibility [31]. The PIQE was computationally less efficient but provided a global quality score and local quality measure. The MS-SSIM and HDR-VDP2 methods, based on multiscale structural similarity, were used to assess the HDR image visibility and quality. Each comparison score is shown in Figure 17. Generally, all reference-free quality metrics were superior to full-reference metrics in terms of their consistency with human subjective quality scores. In Figure 17, the average score of MS-SSIM and HDR-VDP2 is the highest in the proposed method. However, in PIQE, HDRCNN's score shows the best performance. The lower the PIQE score is, the better the performance is, whereas for the other evaluation methods, the higher the score is, the better the performance is. The results show that the proposed model performed well overall, except in terms of PIQE. The proposed model showed a visual superiority to the other existing methods, but its PIQE score was higher than those of HDRCNN and ExpandNet. The PIQE measured the quality of images through block-wise estimation. The fine noise generated by GAN [32] was included in the proposed model, which could explain the high PIQE score.
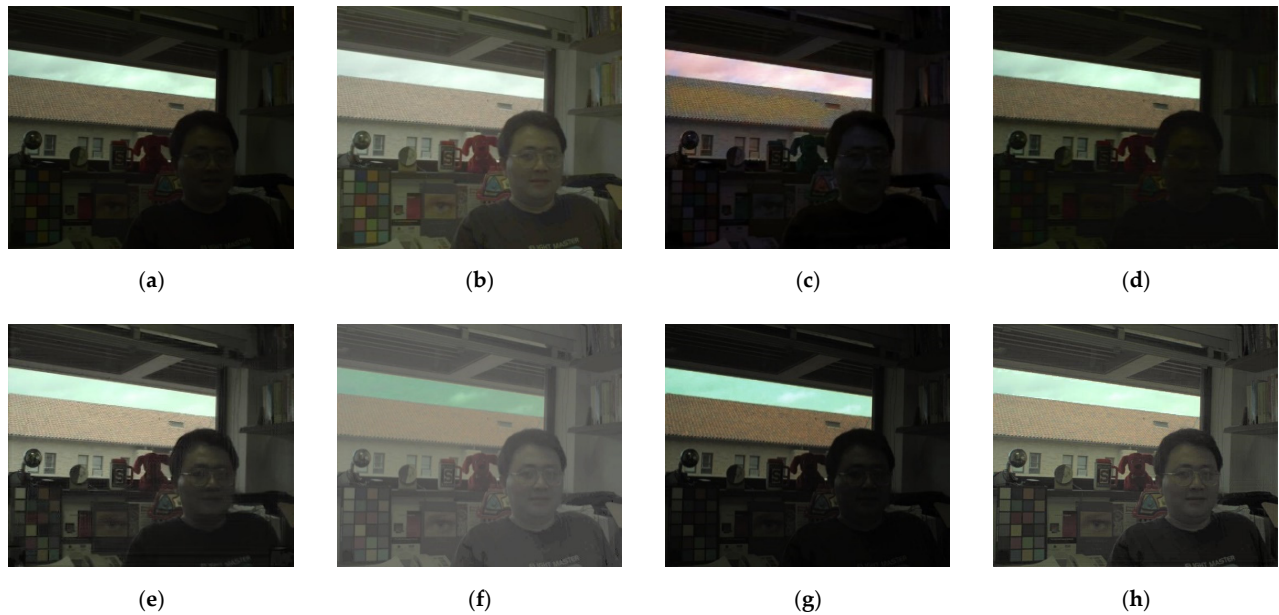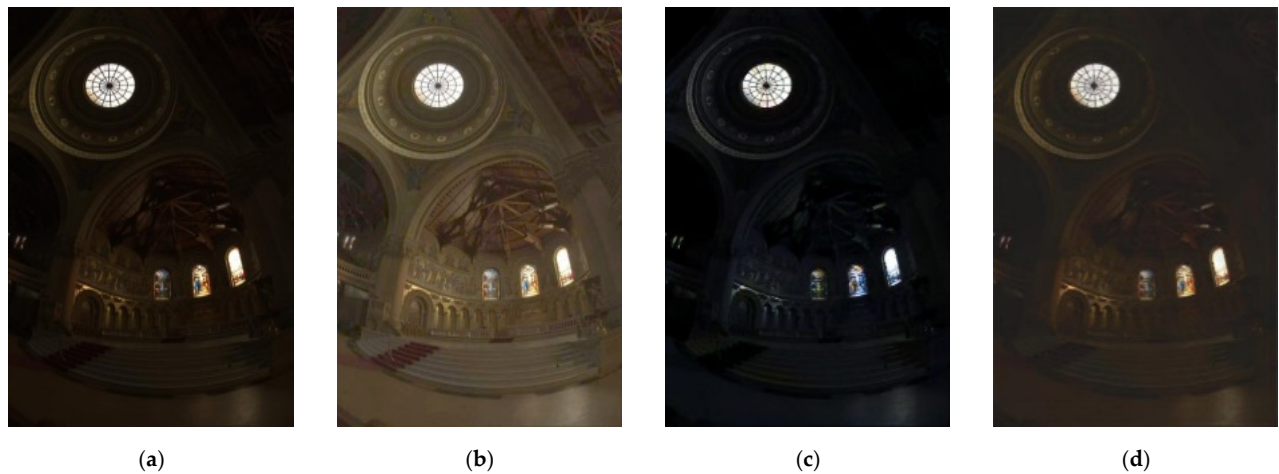
Figure 12. High-contrast image #2: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.

Table 1 represents the computational time needed for each method. All methods were run on the same environment and image resolution (512 × 512). The results show that the proposed methods, Lum CycleGAN and EV GAN, have similar computational times because they use a gray image. HDRCNN has the slowest processing speed and ExpandNet has the fastest.
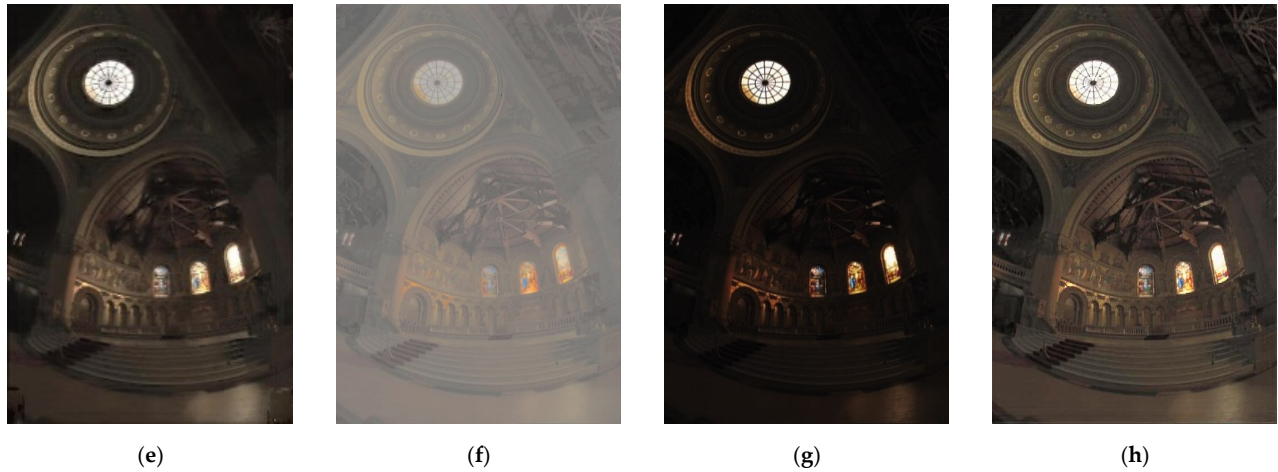
Figure 13. Dark surrounding image #1: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.
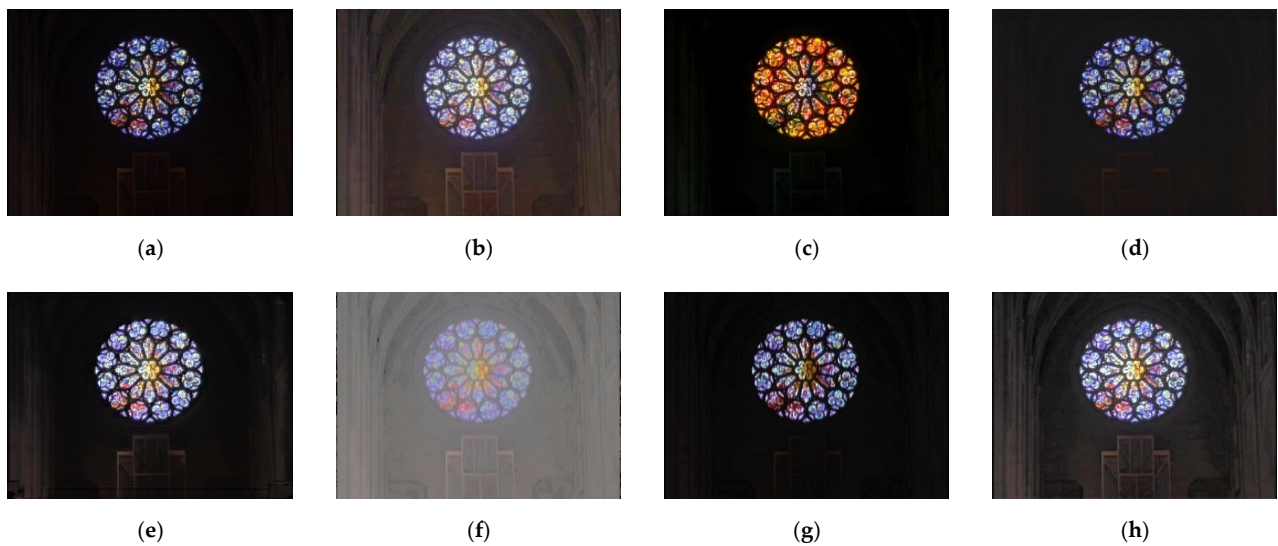
Figure 14. Dark surrounding image #2: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.
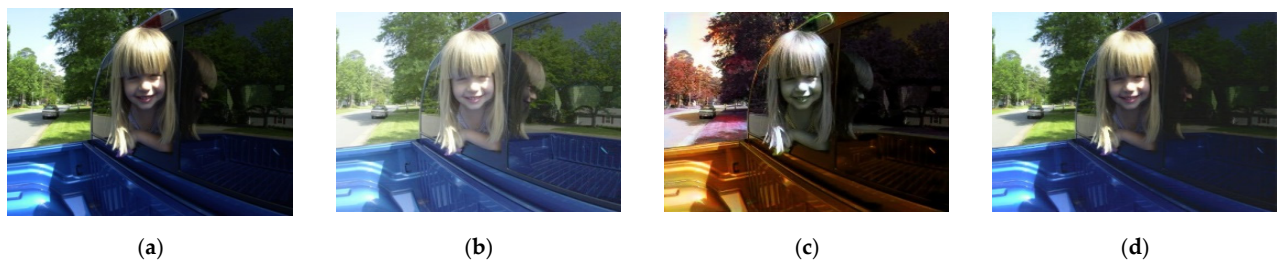
Figure 15. Bright surrounding image #1: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.
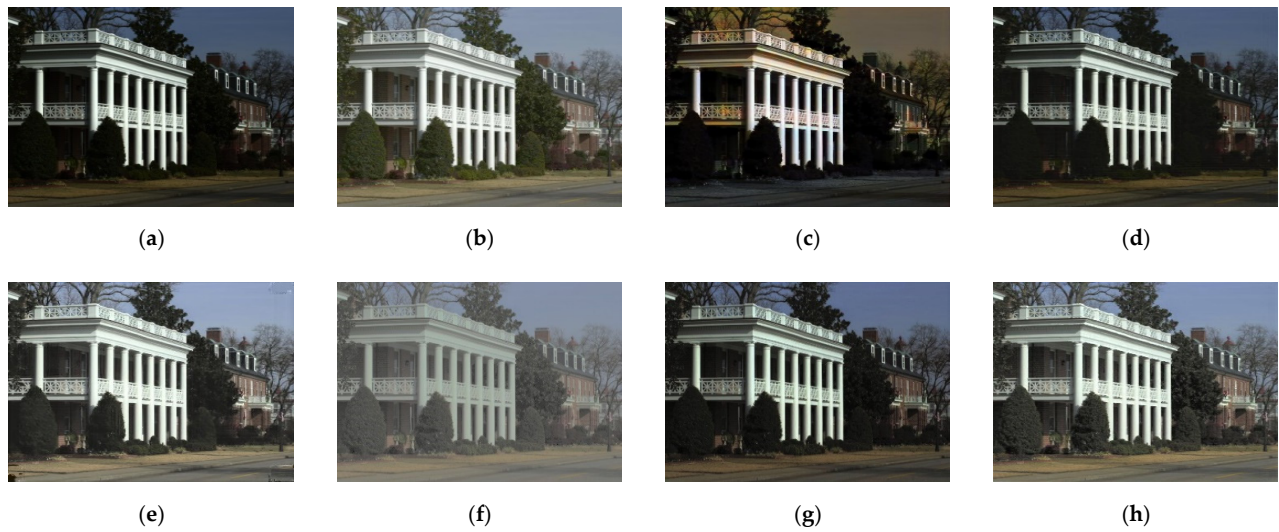
Figure 16. Bright surrounding image #2: (**a**) input, (**b**) target, (**c**) CycleGAN, (**d**) Lum CycleGAN, (**e**) EV GAN, (**f**) HDRCNN, (**g**) ExpandNet, and (**h**) the proposed model.
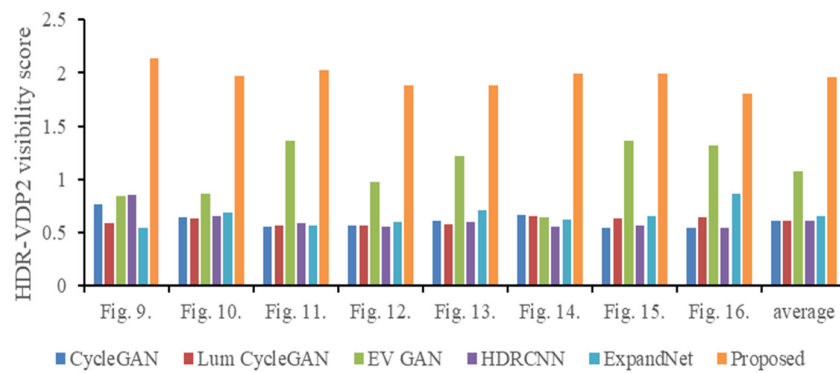
Table 1. Computational time needed for each method.

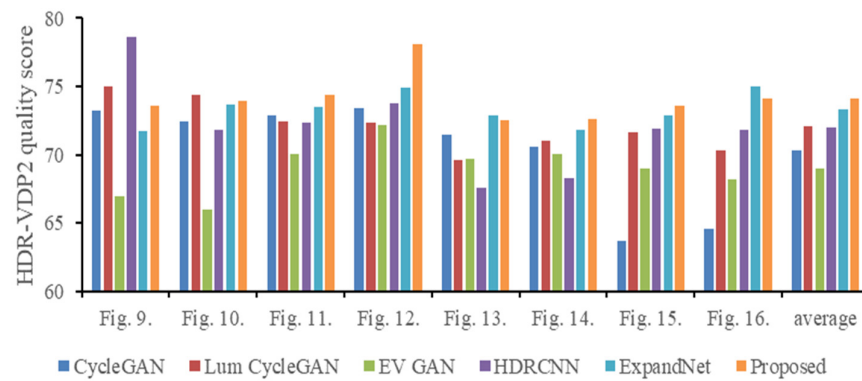| Method	Time(s)
CycleGAN	0.398
Lum CycleGAN	0.334
EV GAN	0.339
HDRCNN	1.006
ExpandNet	0.267
Proposed model	0.339
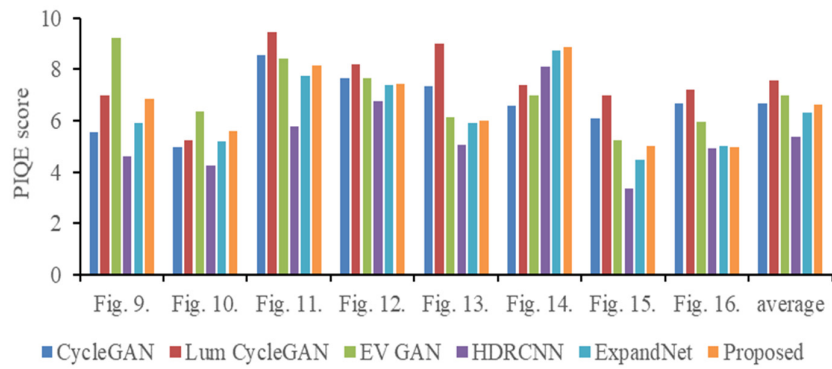
(**a**)
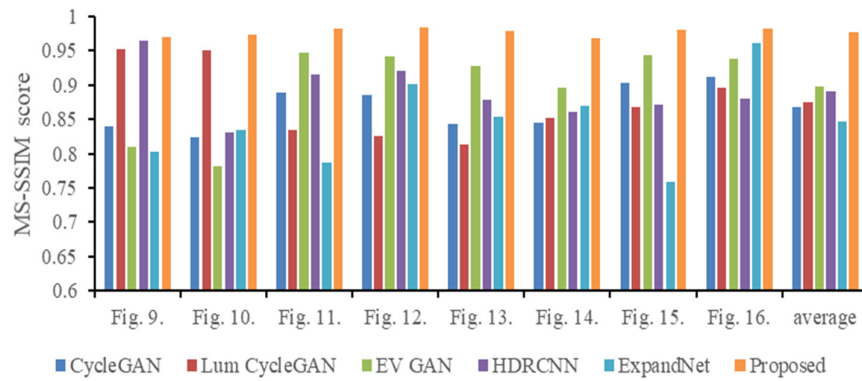
(**b**)

(**c**)

(**d**)

Figure 17. Comparison of metrics: (**a**) perception-based image quality evaluator, (**b**) multiscale structural similarity, (**c**) HDR-VDP2 quality, and (**d**) HDR-VDP2 visibility.

4.5. Object Detection Using Mask R-CNN

Figure 18 depicts an image acquired using a cell phone camera. The image obtained using the proposed model increased the brightness of the objects of the dim surroundings of the image to an identifiable level. Following this, if objects of each of the two images were identified using Mask R-CNN, pedestrians could not be identified in the original images.
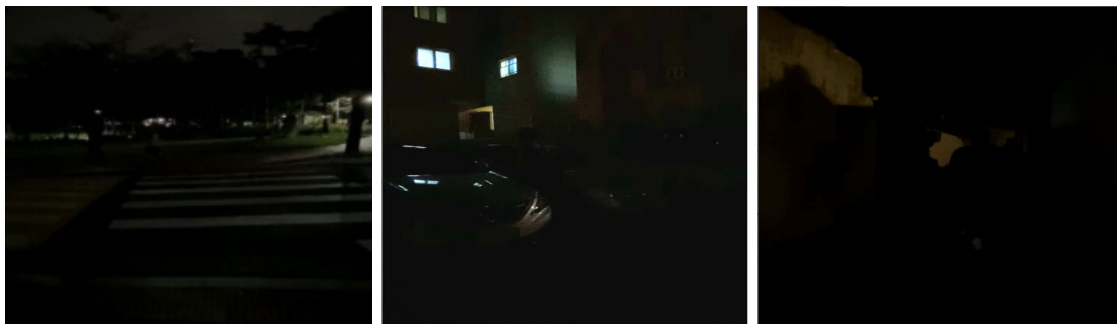
(**a**)
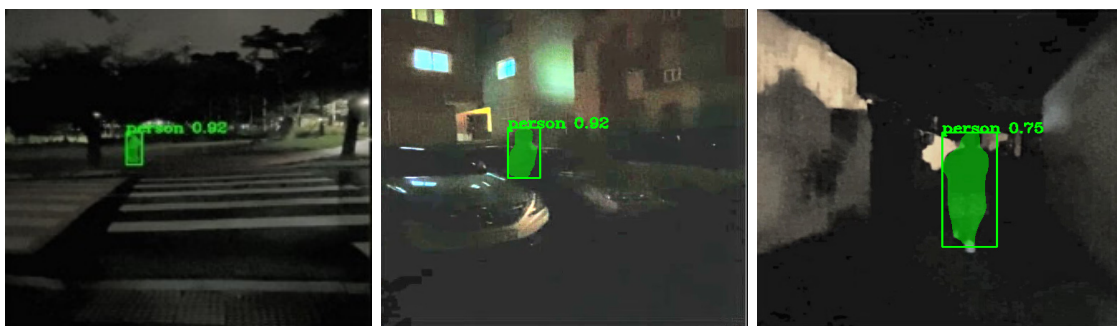
(**b**)

Figure 18. Detection of Mask R-CNN: (**a**) detection of original images and (**b**) detection of proposed results.

4.6. Parameter Optimization

Figures 19–21 show the results of the generator optimal layer experiment. In the generator structure, known as U-Net, an equal number of encoders and decoders are used, and when the result of the encoder is skipped and connected to the decoder the resulting image is partially clear. In this experiment, to find the optimal number of U-Net layers, the number of encoder layers was increased from 4 to 7 and the gamma step was tested from 4 to 10 for optimal data proliferation to quantitatively and qualitatively determine the optimal method. As a qualitative model, the qualities of four layers and six steps were generally good. Figures 19 and 20 show the experiment results. Figure 21 displays the plotted quantitative figures. In Figure 19, the experiment was conducted with the foggy image with no contrast difference in the input image. When U-Net's encoder and decoder are composed of four layers and the six gamma steps, the results show the best image quality. In Figure 20, an experiment was performed with images with dark surroundings with a clear contrast to the input image. If the encoder and decoder of U-NET are composed of four layers and the gamma step is six, good results are common. Figure 21 shows quantitative statistics for the images. In Figure 21a, the blue line shows the lowest value at six steps on the horizontal axis, indicating the best performance. In Figure 21b–d, the blue line shows the highest value at six steps on the horizontal axis, indicating the best performance. This experiment revealed that the training performance was not proportional to the training depth. In addition, the resulting image was output differently depending on the gamma step of the data, indicating that the random data augmentation technique widely used in DL experiments is somewhat inappropriate.
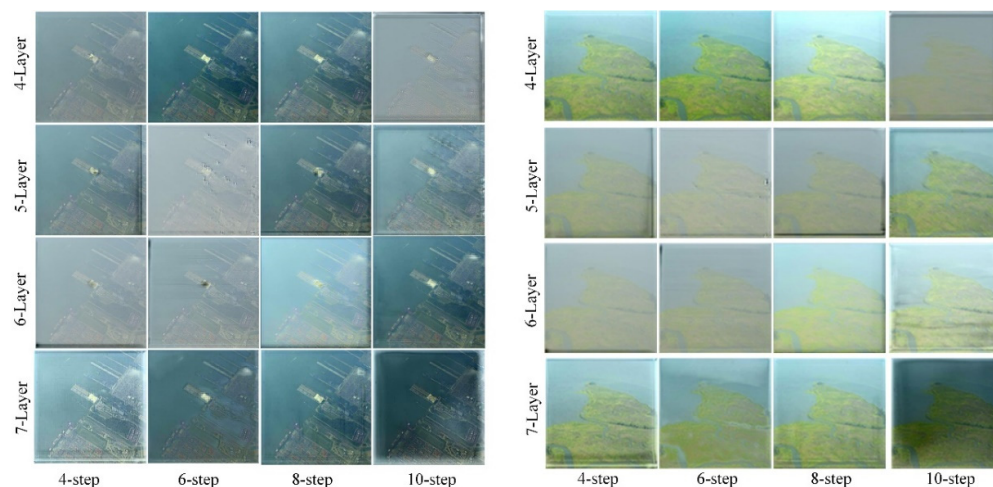

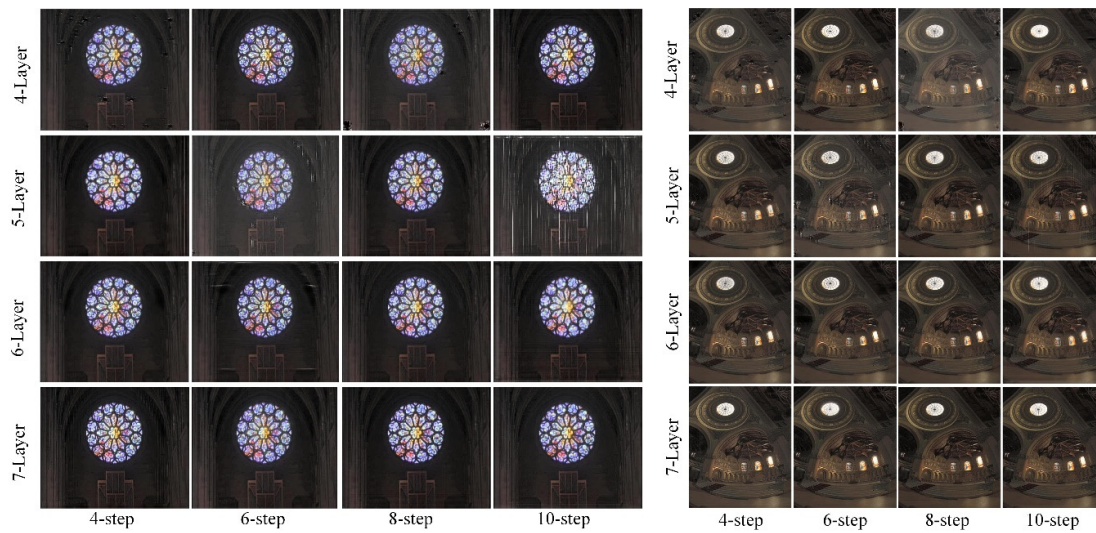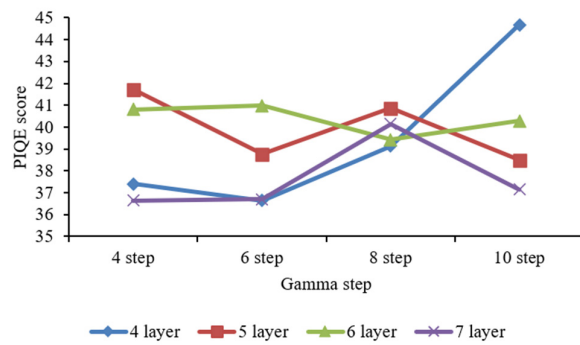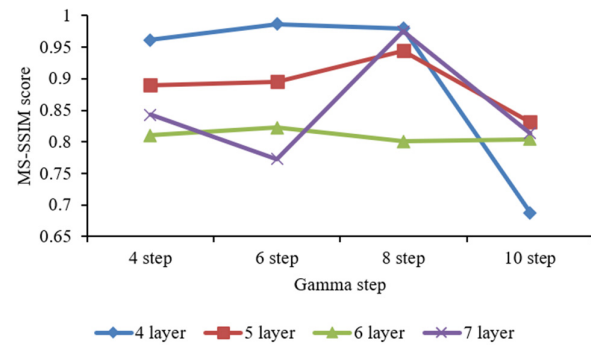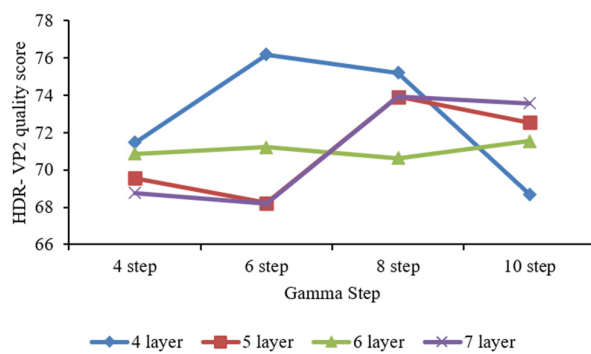**Figure 19.** Results of foggy images according to the number of layers and steps.

Figure 20. Results of dark surrounding images according to the number of layers and steps.
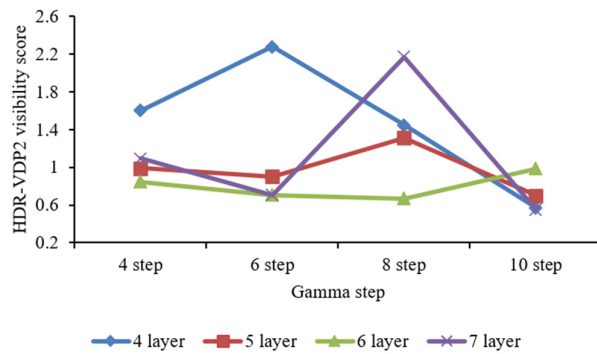

(**a**)

(**b**)

(**c**)

(**d**)

Figure 21. Metric scores for parameter optimization: (**a**) perception-based image quality evaluator, (**b**) multiscale structural similarity, (**c**) HDR-VDP2 quality, and (**d**) HDR-VDP2 visibility.

5. Conclusions

When using the CycleGAN model to generate HDR tone mapping images, the color of the resulting image is distorted and the global tones only change slightly. Therefore, the use of a GAN training optimization model for converting LDR images into HDR images is proposed in this study. First, a gamma shift training method with an extended luminance range is proposed. Second, a weighted loss map trains for tone compression in the local area of images. Third, an RFT method is used to restore the speed of local tone training by balancing the training method with the regional weight map. As the generated module performs well in bright images, mean gamma tuning is applied to the luminance channel of the test image, which is then fed into the module. The proposed model was tested on foggy images, images with dark surroundings, images with bright surroundings, and high-contrast images, and the proposed model outperformed conventional models in the comparative analysis. The HDR-VDP2, MS-SSIM, and PIQE methods were used as evaluation indices. All the evaluation indices showed that the proposed model had an excellent performance on average. The proposed model complements the performance of Mask R-CNN, a prominent object detection model, even in a real night environment. The proposed model successfully detected an object that Mask R-CNN could not detect. Furthermore, the proposed model restored the tone of the surrounding background to an appropriate level. Thus, it has potential applications in commercial CCTV surveillance systems and security industries.

Author Contributions: Conceptualization, S.-H.L.; Methodology, S.-H.L.; Software, S.-W.J.; Supervision, S.-H.L.; Writing—original draft, S.-W.J.; Writing—review and editing, H.-J.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Debevec, P.E.; Malik, J. Recovering high dynamic range radiance maps from photographs. In Proceedings of the ACM SIGGRAPH 2008 Classes on-SIGGRAPH 2008, Los Angeles, CA, USA, 11–15 August 2008; p. 31, doi:10.1145/1401132.1401174.
2. Kwon, H.-J.; Lee, S.-H.; Lee, G.-Y.; Sohng, K.-I. Radiance map construction based on spatial and intensity correlations between LE and SE images for HDR imaging. *J. Vis. Commun. Image Represent.* **2016**, *38*, 695–703, doi:10.1016/j.jvcir.2016.04.022.
3. Banterle, F.; Ledda, P.; Debattista, K.; Chalmers, A. Inverse tone mapping. In Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, Kuala Lumpur, Malaysia, 29 November–2 December 2006; pp. 349–356, doi:10.1145/1174429.1174489.
4. Rempel, A.G.; Trentacoste, M.; Seetzen, H.; Young, H.D.; Heidrich, W.; Whitehead, L.; Ward, G. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph.* **2007**, *26*, 39, doi:10.1145/1276377.1276426.
5. Huo, Y.; Yang, F.; Dong, L.; Brost, V. Physiological inverse tone mapping based on retina response. *Vis. Comput.* **2013**, *30*, 507–517, doi:10.1007/s00371-013-0875-4.
6. Endo, Y.; Kanamori, Y.; Mitani, J. Deep reverse tone mapping. *ACM Trans. Graph.* **2017**, *36*, 1–10, doi:10.1145/3130800.3130834.
7. Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R.K.; Unger, J. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* **2017**, *36*, 178, doi:10.1145/3130800.3130816.
8. Marnerides, D.; Bashford-Rogers, T.; Hatchett, J.; Debattista, K. ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. *Comput. Graph. Forum* **2018**, *37*, 37–49, doi:10.1111/cgf.13340.
9. Liu, Y.-L.; Lai, W.-S.; Chen, Y.-S.; Kao, Y.-L.; Yang, M.-H.; Chuang, Y.-Y.; Huang, J.-B. Single-Image HDR reconstruction by learning to reverse the camera pipeline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1648–1657, doi:10.1109/CVPR42600.2020.00172.

10. Li, H.; Ma, K.; Yong, H.; Zhang, L. Fast Multi-Scale Structural Patch Decomposition for Multi-Exposure Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 5805–5816, doi:10.1109/tip.2020.2987133.

11. Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.* **2018**, *27*, 2049–2062, doi:10.1109/tip.2018.2794218.

12. Wang, R.; Zhang, Q.; Fu, C.-W.; Shen, X.; Zheng, W.-S.; Jia, J. Underexposed photo enhancement using deep illumination estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 6842–6850, doi:10.1109/CVPR.2019.00701.

13. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551, doi:10.1162/neco.1989.1.4.541.

14. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976, doi:10.1109/CVPR.2017.632.

15. 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 (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251, doi:10.1109/ICCV.2017.244.

16. 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, doi:10.1016/j.jvcir.2007.06.003.

17. Jung, S.-W.; Son, D.-M.; Kwon, H.-J.; Lee, S.-H. Regional weighted generative adversarial network for LDR to HDR image conversion. In Proceedings of the International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 19–21 February 2020; pp. 697–700, doi:10.1109/ICAIIC48513.2020.9064984.

18. Jung, S.-W.; Kwon, H.-J.; Son, D.-M.; Lee, S.-H. Generative Adversarial Network Using Weighted Loss Map and Regional Fusion Training for LDR-to-HDR Image Conversion. *IEICE Trans. Inf. Syst.* **2020**, *103*, 2398–2402, doi:10.1587/transinf.2020edl0001.

19. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397, doi:10.1109/TPAMI.2018.2844175.

20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90, doi:10.1145/3065386.

21. Wollmer, M.; Blaschke, C.; Schindl, T.; Schuller, B.; Farber, B.; Mayer, S.; Trefflich, B. Online Driver Distraction Detection Using Long Short-Term Memory. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 574–582, doi:10.1109/tits.2011.2119483.

22. Chen, Y.-S.; Wang, Y.-C.; Kao, M.-H.; Chuang, Y.-Y. Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6306–6314, doi:10.1109/CVPR.2018.00660.

23. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Trans. Graph.* **2016**, *35*, 110, doi:10.1145/2897824.2925974.

24. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends® Mach. Learn.* **2009**, *2*, 1–127, doi:10.1561/2200000006.

25. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; doi:10.1109/CVPR.2018.00262.

26. Wu, S.; Zhong, S.-H.; Liu, Y. Deep residual learning for image steganalysis. *Multimed. Tools Appl.* **2017**, *77*, 10437–10453, doi:10.1007/s11042-017-4440-4.

27. Retinex Images. Available online: https://dragon.larc.nasa.gov (accessed on 11 November 2019).

28. Fairchild, M.D. HDR Dataset. Available online: http://rit-mcsl.org/fairchild//HDR.html (accessed on 6 June 2019).

29. Moorthy, A.K.; Bovik, A. Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. *IEEE Trans. Image Process.* **2011**, *20*, 3350–3364, doi:10.1109/tip.2011.2147325.

30. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multi-scale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; doi:10.1109/acssc.2003.1292216.

31. Mantiuk, R.; Kim, K.J.; Rempel, A.G.; Heidrich, W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.* **2011**, *30*, 40, doi:10.1145/2010324.1964935.

32. Chen, Z.; Zeng, Z.; Shen, H.; Zheng, X.; Dai, P.; Ouyang, P. DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images. *Biomed. Signal Process. Control* **2019**, *55*, 101632, doi:10.1016/j.bspc.2019.101632.