



Junshu Wang<sup>1</sup>, Yue Yang<sup>2</sup>, Yuan Chen<sup>3</sup> and Yuxing Han<sup>4,5,\*</sup>

- <sup>1</sup> College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China; junshuwang@stu.scau.edu.cn
- <sup>2</sup> Pearl River Water Resources Research Institute of PRWRC, Guangzhou 510611, China; yangyue84919@pearlwater.gov.cn
- <sup>3</sup> College of Accounting, Guangdong University of Foreign Studies, Guangzhou 510006, China; 20171504386@gdufs.edu.cn
- <sup>4</sup> Matrix Mathematical Imaging Center, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China
- <sup>5</sup> Lingnan Guangdong Laboratory of Modern Agriculture, South China Agricultural University, Guangzhou 510642, China
- \* Correspondence: yuxinghan@scau.edu.cn; Tel.: +86-020-85288202

Abstract: In unmanned aerial vehicle based urban observation and monitoring, the performance of computer vision algorithms is inevitably limited by the low illumination and light pollution caused degradation, therefore, the application image enhancement is a considerable prerequisite for the performance of subsequent image processing algorithms. Therefore, we proposed a deep learning and generative adversarial network based model for UAV low illumination image enhancement, named LighterGAN. The design of LighterGAN refers to the CycleGAN model with two improvements-attention mechanism and semantic consistency loss-having been proposed to the original structure. Additionally, an unpaired dataset that was captured by urban UAV aerial photography has been used to train this unsupervised learning model. Furthermore, in order to explore the advantages of the improvements, both the performance in the illumination enhancement task and the generalization ability improvement of LighterGAN were proven in the comparative experiments combining subjective and objective evaluations. In the experiments with five cutting edge image enhancement algorithms, in the test set, LighterGAN achieved the best results in both visual perception and PIQE (perception based image quality evaluator, a MATLAB build-in function, the lower the score, the higher the image quality) score of enhanced images, scores were 4.91 and 11.75 respectively, better than EnlightenGAN the state-of-the-art. In the enhancement of low illumination sub-dataset Y (containing 2000 images), LighterGAN also achieved the lowest PIQE score of 12.37, 2.85 points lower than second place. Moreover, compared with the CycleGAN, the improvement of generalization ability was also demonstrated. In the test set generated images, LighterGAN was 6.66 percent higher than CycleGAN in subjective authenticity assessment and 3.84 lower in PIQE score, meanwhile, in the whole dataset generated images, the PIQE score of LighterGAN is 11.67, 4.86 lower than CycleGAN.

**Keywords:** UAV; unsupervised learning; LighterGAN; unpaired dataset; illumination enhancement; attention mechanism; semantic consistency loss; PIQE; generalization ability

# 1. Introduction

With the development of unmanned aerial vehicle (UAV) technologies, more important and complex tasks are performed by low or ultra-low altitude UAVs which embedded powerful functions, especially in the field of urban remote sensing [1]. Meanwhile, in the last decades, with the development of computer science and the popularization of its applications, intelligence has become the development tendency of modern cities [2]. In



Citation: Wang, J.; Yang, Y.; Chen, Y.; Han, Y. LighterGAN: An Illumination Enhancement Method for Urban UAV Imagery. *Remote Sens.* **2021**, *13*, 1371. https://doi.org/10.3390/rs13071371

Academic Editors: Hyungtae Lee, Sungmin Eum and Claudio Piciarelli

Received: 27 February 2021 Accepted: 31 March 2021 Published: 2 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). terms of urban-wise UAV applied techniques, one of the most compelling aspects is the computer vision. The images or videos collected by UAV in applications such as urban observation, monitoring, object detection, 3D reconstruction, scene understanding, digital surface model generation, etc., require various and detailed image information. Unfortunately, insufficient illumination can severely degrade the recognizability and the readability of images, meanwhile, the performance of subsequent computer vision algorithms highly relies on the quality of collected images and effective extraction of the features contained in the image [3]. For UAVs which performing night missions, the poor shooting environments caused image degradation will be more obvious due to the limitations of brightness, viewing angle, etc., meanwhile, when working in low altitude and ultra-low altitude, the unavoidable light pollution in the city will also challenge the performance of photographic equipment carried by UAVs. Therefore, the illumination enhancement of degraded UAV images is a critical preprocessing.

Because traditional image enhancement or correction relies on the prior knowledge of scholars, the prior information needs to be injected into models artificially [4]. Generally, items in urban UAV imagery are dense and complex in feature, meanwhile, uncontrollable factors such as reflected light, shadows, uneven illumination will contribute to more complex variables. Therefore, it is more appropriate to use a learning-based enhancement method based on feature extraction. The convolutional neural network structure itself has proven its ability to obtain prior information [5]. The use of neural networks to achieve image illumination enhancement has gradually become a trend. Recently, scholars such as Ren et al. [6], Lv et al. [7], Guo et al. [8] have started to use neural networks to enhance low illumination images and contributed successful networks. However, these supervised models are all trained by size-limited paired datasets, images in either low illumination subdataset, or a high illumination sub-dataset that is artificially generated. Ideally, to ensure the quality of subsequent algorithms, the items in the enhanced images should be as close to their appearance under sufficient illumination in reality. Unfortunately, most of man-made images may not realistic enough for good training, as distributions of real-world features are complex and multimodal. So, in terms of the dataset, the richer the content of the images in the dataset, the more likely it is to train a high-quality illumination enhancement model. However, even the best low-light dataset (LOL) at present (including 500 pairs of images, use exposure time and ISO speed to adjust the image brightness) may be difficult to train a model that possess strong enough robustness, while avoiding the insufficient data that results in discriminator overfitting. The strictly paired LOL dataset requires the camera to be fixed during the collection process, and the scene cannot be moved, therefore, paired data collection is difficult. In this article, an unsupervised illumination enhancement GAN, named LighterGAN, is proposed referring to CycleGAN [9]. Generally, the performance of deep learning-based models will be affected by the quality of the dataset, therefore, the unlabeled and unpaired dataset collected by us for the urban-wise UAV image enhancement is more advantageous, because it possesses the properties of low time and energy consumption which benefit the collection of a significantly large amount of data. In modeling, two mappings were trained by using two sub-datasets (low illumination and sufficient illumination). One of the mappings will be used to achieve the enhancement task, and the other is also indispensable, which will promote the training of a better model. In terms of model improvement, attention mechanism and semantic consistency loss are proposed; the former is inspired by the impressive performance of non-local neural networks [10], and the inclusion of the latter is taking into account that the original CycleGAN only calculates the image space when calculating the loss function and ignores the calculation of the feature space. After proposing the above improvements, the combination was proven experimentally and therefore determined the model structure of LighterGAN.

The main contributions of this paper are that, in terms of illumination enhancement, in order to demonstrate performance of LighterGAN more intuitively, we compare it with five other algorithms (LIME [11], DUAL [12], Retinex [13], CycleGAN [9] and Enlighten-

GAN [14]), and apply two evaluation criterion, visual subjective perception and objective no-reference image quality assessment. In the test set experiment, our algorithm achieves the best results in both evaluations. Subjectively, LighterGAN is 0.42 higher than state-ofthe-art EnlightenGAN which ranked the second place, meanwhile, the PIQE (perception based image quality evaluator) [15] score was 2.73 points lower than the second place. In the enhancement experiment of sub-dataset Y, LighterGAN also achieves the lowest PIQE score of 12.37, 2.85 points lower than second place. In terms of model structure improvement and generalization ability enhancement of LighterGAN, as an improved model based on CycleGAN, the promotion of its two improvements, attention mechanism and semantic consistency loss, has also been demonstrated by the combination of subjective and objective evaluation. Because the two sub-datasets used for training are all collected in reality, the authenticity of the generated image can be used as a criterion for evaluating the generalization ability of the model. After using the two mappings to generate test images, LighterGAN is 10 percent higher than CycleGAN in subjective authenticity assessment and 3.84 lower in PIQE score. Moreover, after using the two mappings to generate images for the whole dataset, PIQE score of LighterGAN is 11.67, 4.86 lower than CycleGAN. As the improved model is proven to be able to generate images with higher authenticity and higher quality, the benefits of improvements can be demonstrated.

#### 2. Related Works

## 2.1. Limitations of Size-Limited Paired Dataset in Urban-Wise Image Illumination Enhancement

In terms of implementation details of LigherGAN, a low illumination UAV image will be mapped to the feature space by the CNN based encoder and become feature maps, this process could be described as either feature extraction or down-sampling. Whereafter, in the up-sampling, the feature maps will be translated into illumination enhanced image by the decoder, items in the generated image will be characterized by sufficient illumination. In order to enhance an image under the premise of ensuring authenticity, theoretically, the larger the number of images in a dataset, the greater the probability of implementation. However, the image semantic feature information provided by the artificially generated paired dataset is not sufficient. In the urban-wise low illumination images, many scenes include interference light, affecting image readability, which is caused by tamps, orange light reflected by water and building surface, etc. but these scenes will not appear under the case of sufficient illumination environment. By using fixed-parameters, enhanced images will also result in the neural network mapping being inflexible and unable to cope with the diversified application environments where the illumination varies with the environment —even the images in an artificially generated paired dataset can be created simply by using the brightness linear adjustment algorithms. On the other hand, the GAN based model needs enough data to support the convergence in training [16]. In the case of training when only a small amount of data is available, it is difficult to achieve a satisfactory Nash equilibrium [17], and it is easy to fall into model collapse, in which case images generated by the model are not easily recognized by the discriminator, like meaningless noise or even things do not exist in reality. As such, for our study, an unpaired dataset is used. In the design of the model, we consider that when image illumination has enhanced, items in the image is expected to show some necessary sufficient illumination related characteristics, e.g., road and surface of the water will not reflect the light of a streetlamp significantly, resulting in the feature distributions changing. However, the transformation of these features is based on the model being trained on a more reasonable dataset, therefore, our unlabeled and unpaired dataset is more advantageous.

## 2.2. Image Quality Assessment (NR-IQA)

IQA (image quality assessment), the necessary evaluation for image enhancement algorithms, is divided into subjective assessment and objective assessment. The former evaluates image quality from subjective visual perception, while the latter uses mathematical models to give quantitative values with the goal of making objective judgments consistent with the subjective perception of visual perception. Subjective evaluation will inevitably be affected by display equipment, lighting conditions, visual ability and emotion of observers, etc., and it is difficult to handle a large number of tasks. Therefore, automatic and accurate prediction of the objective quality evaluation is necessary. IQA can be divided into 3 types according to the amount of information provided by the original reference image: FR-IQA (full reference-IQA), RR-IQA (reduced reference-IQA) and NR-IQA (no reference-IQA). FR-IQA has both the original image and the enhanced image, the core of which is to compare the information amount or feature similarity. The information possessed by NR-IQA is only the enhanced image itself without any other reference. RR-IQA, the type of method between FR-IQA and NR-IQA, possesses partial features extracted from the original image [18].

For FR-IQA, the most authoritative algorithms are PSNR (peak signal to noise ratio) and SSIM (structural similarity index) [19], both of which are based on the premise that there is a mapping relationship between the original and enhanced images. In this article, unsupervised learning and unpaired dataset are used, which means that each enhanced image does not have a unique correct solution. Therefore, PSNR and SSIM are not suitable for evaluating the quality of images enhanced by the LighterGAN.

Because people are fully capable of judging the quality of a specific image without any reference, NR-IQA with the same preset conditions is more practical in the evaluation of the performance of LighterGAN. In this article, perception based image quality evaluator (PIQE), a representative evaluation method, is applied. As a built-in function in the image processing toolbox of MATLAB, the quality evaluation ability and impartiality of PIQE have been proven, and it has been widely used in the evaluation tasks of images generated by deep learning model, especially in generative adversarial networks, such as [20–22] et al. For urban-wise UAV applications, high authenticity and avoidance of distortion and degradation are required by the subsequent intelligence image algorithms. Meanwhile, in each patch of urban UAV aerial photography captured image, items are greatly complex and various. Therefore, as a block-wise algorithm which uses the method of dividing the image into non-overlapping blocks for subsequent calculations, the PIQE is obviously appropriate. More specifically, there are three masks in PIQE, the activityMask can detect compression artifacts and noise in the input image and estimate the spatial quality, the noticeableArtifactsMask focuses on the blocking artifacts and sudden distortions in the activityMask, the noiseMask can locate and estimate the blocks contain Gaussian noise. In addition, the PIQE image quality level is based on the LIVE Image Quality Assessment Database Release 2 [23]. For an enhancement algorithm based on convolutional autoencoder, the performance of LighterGAN highly relies on the quality of image reconstruction, meanwhile the model generalization ability can also be evaluated by the quality of the generated image [24]. Considering comprehensively, in this article the PIQE score is used in image enhancement quality objective assessment and the quantitative evaluation of model generalization ability.

#### 3. Materials and Methods

#### 3.1. Generative Adversarial Network

Generally, there are two types of deep learning models in terms of tasks: the generative model and the discriminative model [25]. The discriminative model allowed a machine to learn the decision function or conditional probability distribution to output the prediction results. The generative model allowed the machine to learn the joint probability density function and made the conditional probability distribution as the output; GAN is the combination of two models. In 2014, the generative adversarial network was designed by Goodfellow, which allowed researchers to mix the spurious images with the genuine by using unsupervised learning [26]. In GAN, the generator learns feature distributions of real images to make the generated images more realistic to deceive the discriminator. The discriminator has to become more intelligent to discriminate the authenticity between both

types of images with model training. After reaching the Nash equilibrium, the generated images achieve high quality and are hard to be identified.

In recent years, with the development of the GAN family, more effective GAN such as WGAN [27], DCGAN [28], and CycleGAN have been designed to improve the performance in the fields of dataset augmentation, image enhancement, image denoising, image classification, super-resolution, semantic segmentation, etc. [29]. Among these applications, image translation is the most eye-catching one in recent years, as GAN has shown its advantages in the feature transformation field. Initial image translation started with a CGAN-based supervised image translation model, the Pix2Pix, which was proposed by Isola [30]. 'Supervised' means it uses a paired dataset for training, where each image has a mapping with a certain image in another subset, shown in Figure 1a. It uses an end-to-end architecture based on GAN, where the input is the original image x, and the output is the translated image T(x). The original images and the translated images are combined to train the discriminator. Paired datasets are costly in certain cases and have to be well-labeled; therefore, more and more unsupervised image translation models have been proposed recently [31], the most classic one is the CycleGAN. CycleGAN uses an unpaired dataset for training and has a cycle consistency loss to replace the previous reconstruction loss in the achievement of image translation. In this article, the principle of image illumination enhancement of LigherGAN is also based on image translation, the enhancement will be achieved by translating images with low illumination features into sufficient illumination features after LigherGAN has been well-trained by the unpaired dataset, shown in Figure 1b.



Figure 1. Examples of different dataset types, (a) paired dataset; (b) unpaired dataset.

#### 3.2. Unpaired Dataset

As above-mentioned, in the illumination enhancement task, especially when researchers want to collect images in reality instead of adjusting brightness through algorithms, paired data is hard to collect [32]. So, in this paper, unpaired dataset is applied, two UAV imagery sub-datasets are collected, sufficient illumination sub-dataset  $X = \{x_i\}$   $\begin{cases} 2954 \\ i = 1 \end{cases}$  and low illumination sub-dataset  $Y = \{y_j\}$   $\begin{cases} 2000 \\ j = 1 \end{cases}$ . It is worth noting that there is no labeling information in both sub-datasets. Objects found in images included nearly everything in common urban areas, such as buildings, plants, streets, people, cars, grass, water, etc. The images in the dataset are collected by ultra-low altitude (flight altitude

is 0–100 m) or low altitude (flight altitude is 100–500 m) UAV, the spectral range of the sensor is between 450 and 650 nm, and the radiometric depth is 8-bits, meanwhile, the dataset also contains images collected from websites [33–35].

Considering the authenticity of the enhanced image, ideally, the neural network may have sufficient comprehensive ability concerning what the objects in the image should be like in the reality and, using intelligible information to perform the illumination enhancement with image translation. Furthermore, the underlying relationship between two kinds of scenes (low illumination and sufficient illumination) has to be found out carefully. In LighterGAN, a wiser discriminator is needed to judge the authenticity of the enhanced image, as well as a more cunning generator to make an enhanced image as realistic as possible. Hence, details of an image have to give the neural networks as much information as possible. Aiming to guarantee the diversity of dataset content, the multiformity of scene depth in the collecting of dataset is considered. There are three image types, close shot, medium shot, and panoramic shot. Furthermore, for the low illumination images collecting, various exposure time has set from 0 to 2 s, to make sure that the image semantics are identifiable.

In the design of both sub-datasets, because the quantity of images in our dataset is significantly larger than other illumination enhancement datasets, traditional data augmentation methods such as image reversal and image stretch are not applied. This also guarantees the authenticity of the enhancement to a certain extent.

#### 3.3. Normalizations

To prevent the influence of affine transformation, geometric transformation, and also speed up the gradient descent, normalization methods have to be used. There are two common normalizations, instance normalization and batch normalization. Each image has several channels in a convolution layer of a convolutional neural network (CNN), batch normalization refers to the same channel for each of feature maps to perform the normalization operation together. Instance normalization means that a single channel of a single image is individually subjected to achieve normalization operation.

Batch normalization is suitable for discriminative models, such as image classification [36] because it focuses on normalizing each batch to ensure the consistency of the data distribution. The result of the discriminant model depends on the overall distribution of the data; however, batch normalization is not suitable for all applications for two reasons:

- It is sensitive to the batch-size. Because mean and variance are calculated on one batch, if the batch size is too small, the calculated mean and variance will not be enough to represent the entire data distribution.
- The statistics involved in the calculation of the normalization output of a particular sample will be affected by other samples in the batch.

Instance normalization is suitable for generative models, particularly image translation which will be used in our image enhancement algorithm [37]. Because the results of image generation mainly depend on an image instance, normalizing the entire batch is not suitable for image translation tasks. Using instance normalization in image translation could not only accelerate model convergence but also maintain independence between each image instance.

$$\mu_{nc}(x) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{nchw}$$
(1)

$$\sigma_{nc}(x)^{2} = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (x_{nchw} - \mu_{nc}(x))^{2}$$
(2)

$$y_{nchw} = \frac{x_{nchw} - \mu_{nc}(x)}{\sqrt{\sigma_{nc}(x)^2 + \epsilon}}$$
(3)

where *x* is the sample image, *n* is the number of samples, *c* represents the channel, *h* is height, *w* is weight,  $\mu_{nc}$  is the means of *x* in one channel,  $\sigma_{nc}(x)^2$  is the variance of a certain channel,  $\epsilon$  is a constant added to increase training stability, and  $y_{nchw}$  is the normalization result. For the reasons above, instance normalization was used in the 2nd, 3rd, and 4th convolutional layers in the discriminator, the 1st and 2nd deconvolutional layers in the decoder, and also used in the residual block.

#### 3.4. Network Structure

After the creation of the generative adversarial network, the original structure has been modified by many scholars and used to implement their own applications. Generally, when the GAN structure reaches a reasonable dynamic equilibrium, a stricter discriminator will compel the generator to provide generated images close to the real, making it difficult for the discriminator to make a judgment. The image enhancement is the basis of subsequent algorithms, and the authenticity of the enhanced image will also directly affect the performance of subsequent image processing algorithms. So, the advantages of better-performing generators and discriminators are obvious.

Generally, the factors that need to be referenced in the initialization of a neural network structure are:

- Size of the dataset;
- Type of data;
- Quality of data;
- The function of the neural network;
- Strength of computing resources;
- Existing network structure with reasonable or excellent performance that could be modified for aiming application;

Experimentally, the performance of the initial network structure has evaluated. When a reasonable structure has found, it optimized manually to fine-tune the depth of channels, transfer functions, normalization methods, etc.

### 3.4.1. Markovian Discriminator (PatchGAN)

Generally, a discriminator of a GAN will output a certain number (0 for fake, 1 for real) for an image; however, features in an image depend on pixels and their neighboring area, i.e., pixels far away have less relevance [38]. For an image enhancement model pursuing authenticity, it has to make a certain judgment for each part of the image, so a patchbased discriminator is applied. This discriminator takes advantage of a convolutive neural network and can process each image block independently in the same way; therefore, each result in the judgment matrix that is output by the discriminator is the result of performing individual judgment of each patch—this design makes a more reasonable judgment.

In the structure of discriminator, there are 3 down-sampling convolutional layers with the rectified linear units (ReLU) transferring a  $256 \times 256 \times 3$  input image *x* into a  $32 \times 32 \times 1$  feature map. For the high-frequency information, only the local patches are the focus of the image. The instance normalization is used to map data to the range  $\{0 \sim 1\}$ , which optimizes the efficiency of model training while avoiding gradient vanishing and exploding. At the last CNN layer, a  $1 \times 1$  convolution is used to achieve dimensionality reduction [39], from  $32 \times 32 \times 512$  to  $32 \times 32 \times 1$ , and achieve cross-channel information interaction and nonlinearity increase. An attention model is added at the 2nd and 3rd convolutional layers.

## 3.4.2. Autoencoder

To perform feature extraction of an image and efficient illumination enhancement mapping, hourglass-like convolutional autoencoder architecture (shown in Figure 2) is applied. As a typical unsupervised learning model, the autoencoder contains two major parts: an encoder and a decoder [40]. The encoder maps the input low-illuminance



sample x to the feature space z, the decoder maps z (the abstract features) to the sufficient illumination feature space to obtain the sample y (illumination enhanced image).

Figure 2. Illumination enhancement mapping and the structure of LighterGAN.

Convolutional neural network and deconvolutional neural network are used in the structure of the encoder and the decoder, respectively. A CNN (convolutional neural network) is a multilayer Hubel–Wiesel structure constructed by imitating the information processing process of the human optic nerve system [41]. One of the advantages of CNN is that it can avoid the complicated pre-processing of images. A CNN can also simulate the multi-level structure of the visual signal processing system in a human brain to establish a mapping from low-level features to high-level semantics, so as to achieve data hierarchical feature extraction.

When the classical neural network model is adopted (fully connected neural networks), the computer needs to read the whole image as the input of the neural network model. When the size of the image is large, its parameter scale and training computation will become huge. It is worth noting that human cognition of the world is generally to recognize certain parts first and then gradually recognize the whole [42]. Similarly, in the image spatial connection, pixels within a local range are closely related to each other, while those at a distance are weakly related. Each neuron in a neural network does not need to perceive the global image. It only needs to perceive one certain part and the integration of the local information can obtain global information. This is the receptive field mechanism of convolutional neural networks.

In the autoencoder structure, ReLU (the rectified linear unit) is used as the activation function to perform nonlinear mapping on the output of the convolutional layer [43], it can improve the neural network's perception of item features in sufficiently illuminated natural light images. In this study, one encoder was trained for both types (low/sufficient illumination) of images and decoders were trained for Mapping *A* and Mapping *B*, respectively. Through this end-to-end structure, low-to- sufficient and sufficient -to-low image illumination feature transform will be realized.

### 3.4.3. Encoder

The encoder is designed to generate multi-channel feature maps for a low illumination image and increases the nonlinear relationship between pixels and regions and realizes the extraction of image semantics. The input image is  $256 \times 256 \times 3$ , and after a padding process, the input image will be transformed to  $64 \times 64 \times 256$  by using three convolution

layers. To ensure the integrity of the image information and increase the depth of the neural network, 9 residual blocks are added after the convolution layers. This skip-connecting structure solves the problem of information loss caused by the increase of network depth and also helps to solve the problem of gradient vanishing and gradient exploding. This structure makes our model deeper in layers while ensuring reasonable performance. Finally, the input image was mapped to a high-dimensional space by the encoder, where the output feature map was  $64 \times 64 \times 256$ .

## 3.4.4. Decoder

After using the encoder to map the input low-light image to the high-dimensional feature space, the decoder will reformat the  $64 \times 64 \times 256$  feature maps produced by the encoder into an enhanced sufficient illumination image. This up-sampling operation is completed by the deconvolution layer. In the operation, each point in the feature map (low resolution, high-level descriptions) will be used to reconstruct a region in the output image—it can be described as translation. Two deconvolution layers are designed to fix the output size ( $256 \times 256 \times 3$ ). An attention module is used before the first deconvolution layer to make key semantic translation more prominent. Instance normalization is also used to improve translation results.

#### 3.4.5. Attention Mechanism

In LighterGAN, the input for mapping is an image whose information is abstract and complex; therefore, a suitable mechanism is needed to allocate computing resources to something more important. There are 4 types of attention mechanisms: item-wise soft attention, item-wise hard attention, location-wise soft attention, and location-wise hard attention [44]. Item-wise attention focuses on list input (e.g., NLP (natural language processing)), whereas the location-wise attention is designed for feature map input. Location-wise attention is designed by mimicking human visual system characteristics (scan image globally to obtain the target area that needs to be focused, and then invest more attention in this area to obtain more details), Similarly, location-wise soft attention transforms the entire feature map so that the part of interest can be highlighted. Compared with the location-wise hard attention (non-differentiable), which discretely selects a sub-region from the input feature map as the final feature, the location-wise soft attention (differentiable) [45] can calculate the gradient through the neural network and learn the weight of attention through forward and backward propagation, which is more suitable for our requirements.

The weight of the feature map has been re-adjusted, and the important part became more prominent by letting the deep neural network learn the areas that need attention in the picture via network training. This design of the attention mechanism is inspired by the non-local neural network [10]. The attention mechanism is implemented by an additional neural network connected to the original neural network. The entire model is still end-to-end so the attention module can be trained simultaneously with the original model. The principle formula is shown as:

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j)$$
(4)

where *i* is the index of an output position (in space, time, or spacetime) whose response is computed and *j* is the index that enumerates all possible positions. *x* is the input signal (image, sequence, video; often their features) and *y* is the output signal of the same size as *x*. A pairwise function *f* computes a scalar (representing relationship such as affinity) between *i* and all of *j*. The unary function *g* computes a representation of the input signal at the position *j*. The response is normalized by a factor C(x).

For the presentation of the similarity between  $x_i$  and  $x_j$ , the embedding Gaussian function is used:

$$f(x_i, x_j) = e^{\theta(x_i)^T \varphi(x_j)}$$
(5)

In this formula,  $\theta(x_i) = W_{\theta}x_i$  and  $\varphi(x_j) = W_{\varphi}x_j$  are two embeddings,  $W_{\theta}$  and  $W_{\varphi}$  are weight matrices that to be learned. Then the nonlocal block could be defined as:

$$z_i = W_z y_i + x_i \tag{6}$$

where  $y_i$  is given in Formula (4),  $W_z$  is initialized as zero. Significantly,  $x_i$  is used as a residual connection, to keep information transmission. In this paper, the attention block is shown in Figure 3. In the block, to achieve the attentional behavior, first, 3 feature maps copies were created, f, g, and h (shape of their tensors are shown in Figure 3), f and g reduced the number of channels to one-eighth of the original to reduce the computation cost, then f multiplied the g transpose, then the SoftMax operation is used to generate the output attention map. Then, the map is multiplied with h. In the end, the input feature map is added to generate the attention block output, which obtained the shape as same as the input.



Figure 3. Attention module.

In order to better explore the improvement of attention mechanism module on enhanced image visual perception and no-reference image quality assessment, a comparative experiment is designed in Section 4.

## 3.5. Model Outline

In the outline of LighterGAN, with the purpose of extracting particular characteristics in both unpaired sub-datasets and to know how these characteristics could be used to enhance the brightness of urban-wise low illumination images while ensuring authenticity. Two major mappings inside two circular structures are designed, shown in Figures 4 and 5:



**Figure 4.** Discriminators and loss: (a)  $D_A$ , discriminator between x and B(y); (b)  $D_B$ , discriminator between y and A(x).



**Figure 5.** Logic outlines of two cycles with mappings and loss: (**a**) sufficient illumination cycle and loss; (**b**) low illumination cycle and loss.

- Mapping A: sufficient illumination→low illumination;
- Mapping *B*: low illumination  $\rightarrow$  sufficient illumination;
- Sufficient illumination cycle:  $x \to A(x) \to B(A(x)) = \hat{x} \approx x$ ;
- Low illumination cycle:  $y \to B(y) \to A(B(y)) = \hat{y} \approx y$ .

*x* comes from sufficient illumination sub-dataset *X* and *y* comes from low illumination sub-dataset *Y*, A(x) and B(y) are the translated low/sufficient illumination image produced by the Mapping *A* and Mapping *B*, both are used in the training of sufficient illumination image discriminator  $D_A$  and low illumination image discriminator  $D_B$ . Then, A(x) and B(y) are used in the Mapping *B* and Mapping *A*, respectively, the image generated in this step could be described as  $\hat{x} = B(A(x))$  and  $\hat{y} = A(B(y))$ , which is then produced by the decoder for the calculation of cycle-consistency loss.

In our task, Mapping *B* will be used to achieve illumination enhancement, discriminator  $D_A$  will judge the authenticity. However, it is worth emphasizing that, as a part of the circular structure, the prerequisite for the excellent performance of Mapping *B* is that each part of the entire structure is well-performing, and it cannot be separated from the overall structure. Therefore, in the training phase, careful training was conducted on Mapping *A* and discriminator  $D_B$ . In the experimental section, in order to prove the generalization ability of the LighterGAN, Mapping *A* will be used to generate low illumination images for visual subjective evaluation and the evaluation of NR-IQA.

## 3.6. Adversarial Loss

The role of the discriminator  $D_A$  was to make judgement between the sufficient illumination image B(y) enhanced by Mapping B and the real sufficient illumination images in sub-dataset X, discriminator  $D_B$  aims to make judgement between the low illumination image A(x) which is generated by Mapping A and the real low illumination images in sub-dataset Y, the adversarial loss which content discriminator can be described as:

$$Loss_{D_{A}}(D_{A}, B) = \mathbb{E}_{x \sim p_{data(x)}}[\| D_{A}(x) - M_{ones} \|_{2}] + \mathbb{E}_{y \sim p_{data(y)}}[\| D_{A}(B(y)) - M_{zeros} \|_{2}]$$
(7)

$$Loss_{D_B}(D_B, A) = \mathbb{E}_{y \sim p_{data(y)}}[\| D_B(y) - M_{ones} \|_2] + \mathbb{E}_{x \sim p_{data(x)}}[\| D_B(A(x)) - M_{zeros} \|_2]$$
(8)

$$Loss_{D_{sum}}(A, B, D_A, D_B) = (Loss_{D_A}(D_A, B) + Loss_{D_B}(D_B, A))/2$$
(9)

where the data distributions are denoted as  $x \sim pdata(x)$  and  $y \sim pdata(y)$ . In the training of the neural networks, to distinguish real one from generated one, the discriminator is expected to approach infinitely close to  $M_{ones}$  when recognizing a real image, and to approach infinitely close to  $M_{zeros}$  when recognizing a generated image. Furthermore, a more stringent discriminator also benefits from the performance of the generator (to produce the enhanced images realistically).

In the loss function, the *L*2 distance between the discriminator output and the  $32 \times 32 \times 1$  matrix  $M_{ones}$  was calculated (each value in the matrix is 1) for the real image, while also calculating the *L*2 distance between the  $32 \times 32 \times 1$  matrix  $M_{zeros}$  for translated images. The total loss function of the discriminator is represented by the addition of the loss functions above. It is expected that the value of the loss function will continually minimize and converge to a reasonable value approaching 0 by the end of the training.

#### 3.7. Cycle-Consistency Loss

First, the original cycle-consistency loss is calculated referencing to the CycleGAN. Two types of cycle (sufficient illumination and Low illumination) generated images  $\hat{x} = B(A(x))$  and  $\hat{y} = A(B(y))$ , which are used to calculate the *L*1 distance between two kinds of real images. In the sufficient illumination cycle, the  $Loss_{cycx}(X, B(A(x)))$  is calculated, while  $Loss_{cycy}(Y, A(B(y)))$  in a low illumination cycle. Then, the cycle-consistency loss  $Loss_{cyc}(A, B)$  could be described as the addition of the two losses above.

$$Loss_{cycx}(X, B(A(x))) = \mathbb{E}_{x \sim p_{data(x)}} [\| B(A(x) - X \|_{1}]$$
(10)

$$Loss_{cycy}(Y, A(B(y))) = \mathbb{E}_{y \sim p_{data(y)}}[|| A(B(y) - Y ||_{1}]$$
(11)

$$Loss_{cyc}(A,B) = Loss_{cycx}(X,B(A(x))) + Loss_{cycy}(Y,A(B(y)))$$
(12)

#### 3.8. Semantic Consistency Loss and Model Overall Loss

In LighterGAN, it has to achieve semantic consistency in the enhancement process, which means translating images while ensuring that the scenes in the enhanced images still maintain their original characteristics in reality (e.g., a car in a low illumination image will not be wished to be translated into a boat in the sufficient illumination). Therefore, in LighterGAN, we propose semantic consistency loss  $Loss_c(A, B)$ . As known, a CNN can extract image semantics and this work has been achieved by the down-sampling convolutional layers in the encoder. So, the *L*2 distance between two kinds of encoder results  $Loss_{cx}(X, B(A(x)))$  was calculated for x and  $\hat{x}$ ,  $Loss_{cy}(Y, A(B(y)))$  for y and  $\hat{y}$ , then, semantic-consistency loss  $Loss_c(A, B)$  could be described as the addition of the two losses above.

$$Loss_{cx}(X, B(A(x))) = \mathbb{E}_{x \sim p_{data(x)}} [\| E(B(A(x)) - E(X) \|_{2}]$$
(13)

$$Loss_{cy}(Y, A(B(y))) = \mathbb{E}_{y \sim p_{data(y)}}[\| E(A(B(y)) - E(Y) \|_{2}]$$
(14)

$$Loss_{c}(A, B) = Loss_{cx}(X, B(A(x))) + Loss_{cy}(Y, A(B(y)))$$
(15)

In order to demonstrate the improvement of the illumination enhanced image quality by adding semantic consistency loss function, as well as proving the improvement of attention mechanism, we used no-reference image quality assessment (NR-IQA) to compare the images before and after adding semantic consistency loss in the comparative experiment in Section 4. In the training of the generator, the loss function  $Loss_{cd}(A, B)$  is also added to determine whether the images were realistic enough. With this purpose,  $D_B(A(x))$  and  $D_A(B(y))$  are used to calculate the *L*2 distance with matrix  $M_{ones}$  respectively, this loss could be described as follows:

$$Loss_{cd}(A,B) = \mathbb{E}_{x \sim p_{data(x)}}[\| D_B(A(x)) - M_{ones} \|_2] + \mathbb{E}_{y \sim p_{data(y)}}[\| D_A(B(y)) - M_{ones} \|_2]$$
(16)

Then,  $Loss_{G_{sum}}(A, B, D_A, D_B)$ , the overall loss of this model, could be expressed as the addition of three losses motioned above.

$$Loss_{G_{sum}}(A, B, D_A, D_B) = Loss_{D_{sum}}(A, B, D_A, D_B) + Loss_c(A, B) + Loss_{cd}(A, B)$$
(17)

# 3.9. Training Details

In the training strategy and environment, the most significant part has to be the selection of optimization methods and the learning rate. An excessively low rate will consume a large amount of time and an excessively high learning rate would lead to high fluctuation in the convergence curve, even for nonconvergence [46]. We applied adaptive moment estimation by using AdamOptimizer [47], with an initial learning rate set to  $\lambda = 0.0002$  and the exponential decay rate for the 1st moment estimation set to 0.5. The model was built and trained by using TensorFlow v1.8 [48], the training equipment was a single-core GeForce RTX2080ti with 11 Gb of graphics memory, the environment was ubuntu 16.04 with Intel Xeon E5-2678 v3. The number of epochs was set as 140.

# 4. Experiments

## 4.1. Visual Subjective Evaluation and NR-IQA

As mentioned in Section 2, the enhancement of low illumination UAV aerial photography captured image should not only focus on the enhancement performance, but also on image authenticity and quality. In this comparative experiment, to evaluate the quality of images which were enhancement by different enhancement algorithms, 44 images were randomly selected from the test set, and then, a combination of subjective and objective evaluation was used. In terms of subjective evaluation, statistical analysis of the human visual perception evaluation was used. In terms of objective evaluation, because traditional methods such as PSNR and SSIM were not suitable for the evaluation of LighterGAN, the score of perception based image quality evaluator was used as an objective evaluation standard. In both evaluation methods, the performance of LighterGAN was compared with LIME, DUAL, Retinex, the CycleGAN (trained by same dataset as LighterGAN) and the EnlightenGAN. LIME and DUAL are two traditional computer vision methods that have performed well recently, the former is based on the illumination map, the latter is based on exposure correction. Retinex is an instructive classic method and the EnlightenGAN is a cutting-edge GAN based deep learning illumination enhancement method. CycleGAN is the reference model, moreover, in image enhancement performance and generalization capabilities estimation, CycleGAN can be used to reflect the benefits of the improvements which are proposed in this article.

In detail, CycleGAN was trained on the same dataset. We trained the model and generated the images by using the sequential and simultaneous adoption of improvements. In reproductions of LIME, DUAL, Retinex and EnlightenGAN, the default settings suggested by the authors were used. In order to observe the performance of image enhancement more intuitively, partial comparison experiment results are shown in Figure 6.



**Figure 6.** Illumination enhanced images display: (**a**) original urban-wise low illumination images; (**b**) images enhanced by DUAL; (**c**) images enhanced by LIME; (**d**) images enhanced by Retinex; (**e**) images enhanced by EnlightenGAN; (**f**) images enhanced by CycleGAN; (**g**) images enhanced by LighterGAN.

In the statistical analysis of the human subjective evaluation, 135 participants ranked the results of 6 different image enhancement algorithms in a total of 44 ranking questions based on brightness and authenticity, furthermore, we scored the different algorithms in each question by weighting the ranks, as shown in the following formula:

$$Score = \frac{\sum_{i=1}^{6} W_i * Num_i}{N}, N = 135$$
(18)

Here *i* was the ranking,  $W_i$  was the weight corresponding to the ranks,  $W_i \in \{1, 2, 3, 4, 5, 6\}$ ,  $Num_i$  was the number of persons who given this rank, N was the number of participants. For 44 randomly selected urban-wise low illumination images, scores of each enhancement algorithms in 44 ranking question were calculated according to the criterion that the higher the rank, the greater the weight. With this method, the average score of each algorithm in visual subjective evaluation were calculated, the results were shown in Table 1.

Table 1. Average score of 6 different enhancement algorithms.

Algorithm	LighterGAN	N EnlightenG	AllycleGAN	Retinex	LIME	DUAL
Score	4.91	4.49	4.09	2.46	2.51	2.55

In the NR-IQA evaluation, to demonstrate quantitative comparisons of mentioned enhancement algorithms, the PIQE was used to score the 44 test images, the results were shown in Table 2.

Algorithm	LighterGAN	EnlightenGAN	CycleGAN	I Retinex	LIME	DUAL
PIQE	11.75	14.48	15.83	16.80	17.60	17.03

**Table 2.** Average perception based image quality evaluator (PIQE) scores of 44 test images in 6 different enhancement algorithms.

In this evaluation, it is worth emphasizing that, the lower the PIQE score, the better the image quality. It could be found that among the mentioned algorithms, the rank of PIQE scores and the rank of visual evaluation results were almost the same, especially in the ranking of the three algorithms with lower scores (LighterGAN, EnlightenGAN and CycleGAN). In the test set images comparative experiments, LighterGAN has achieved the best results in visual perception subjective judgment and NR-IQA objective quantitative judgment, furthermore, the results could illustrate that PIQE could objectively evaluate the performance of the illumination enhancement algorithms.

## 4.2. NR-IQA Evaluation of Sub-Dataset Y

Because LighterGAN used unsupervised learning and an unpaired dataset, meanwhile, it could be found from the test images that the model did not show the phenomenon of model collapse and overfitting, there were no unique correct results for the enhancements in the low illumination sub-dataset *Y*, moreover, the result of each low illumination image in the sub-dataset was in an unknown state before being enhanced. Therefore, in order to evaluate the performance of image enhancement more comprehensively and objectively, we have further expanded the scale of the number of images and designed a comparative experiment for all images of the sub-dataset *Y* for this purpose.

In Section 4.1, the objectivity of PIQE to the quantitative evaluation of the enhancement algorithm has been proven through the test set images subjective and objective comparison experiments results. Therefore, PIQE was also used to evaluate the improvement benefit from the attention mechanism in illumination enhancement. In this comparative experiment, LighterGAN without attention mechanism was trained and joined the quantitative evaluation.

From the PIQE evaluation of all 2000 images in the sub-dataset *Y*, it could be found that the average PIQE score of LighterGAN was 12.37, it still achieved the best result, the following two were still EnlightenGAN and CycleGAN. Comparing with the comparative experimental results using the test set, it could be found that three deep learning based algorithms (LighterGAN, EnlightenGAN and CycleGAN) still achieved better results, and the rank of these three algorithms were also consistent with the previous experiment. Moreover, from the results, it was worth to highlight that the attention mechanism has indeed achieved an improvement in image quality, in Table 3, compared with the LighterGAN without attention, the average PIQE score of LighterGAN has been reduced by 0.44, and the improvements brought by semantic consistency loss was also show by the difference between the scores of LighterGAN without attention and CycleGAN in the table.

Algorithm	LighterGAN	LighterGAN without Attention	EnlightenGAN	CycleGAN	Retinex	LIME	DUAL
PIQE	12.37	12.81	15.22	17.37	18.10	19.41	19.30

**Table 3.** Average PIQE scores of sub-dataset *Y* enhancement.

#### 4.3. Comparison of Generalization Ability with CycleGAN

As the predictive ability of a model for unknown data, generalization ability was an important criterion for evaluating deep learning models. Therefore, as an improved CycleGAN model, the rationality of the improved strategies in LighterGAN should be evaluated by using generalization ability. Because LighterGAN used an unpaired dataset, in the model training, the enhanced image was not provided with the only correct image. On the contrary, the model needed to provide a conjectural result after fully learning the features in the two sub-datasets which were collected from real scenarios (sufficient illumination subset *X* and low illumination subset *Y*). Therefore, in terms of the visual subjective evaluation, the authenticity of the enhanced image could be used as the standard, in addition, in terms of objective evaluation, PIQE was also used as NR-IQA method.

In the generalization ability contrast experiment, 99 images (including 55 sufficient illumination images and 44 low illumination images) in the test set were used. The same as the operation of CycleGAN, Mapping *A* and Mapping *B* in the cycle structure of Lighter-GAN were used to generate images respectively. In order to observe the performance of image enhancement more intuitively, partial comparison experiment results are shown in Figure 7. In the visual subjective evaluation, the average probability of the images generated by the two models being selected as the more authentic one was calculated and recorded separately. In the objective evaluation, PIQE was used to calculate the scores of the two models in 99 test set images and all images in the unpaired dataset (contained sub-dataset *X* and sub-dataset *Y*) and list them in Table 4.



**Figure 7.** Generated images display, (**a**,**d**) original images; (**b**,**e**) images generated by Mapping *A* in CycleGAN; (**c**,**f**) images generated by Mapping *A* in LighterGAN.

Table 4.	Evaluation	results	between	Lighter	GAN	and (	CycleG.	AN.
----------	------------	---------	---------	---------	-----	-------	---------	-----

	LighterGAN	CycleGAN
Average selected probability	53.33	46.67
PIQE (test set)	11.99	15.83
PIQE (whole dataset)	11.67	16.35

In the experimental results, LighterGAN was 6.66 percentage points higher than CycleGAN in subjective evaluation. In the objective evaluation, the PIQE scores of test set images and full dataset images were 3.81 and 4.68 less than CycleGAN respectively.

# 5. Discussions

As an improvement to the original CycleGAN structure, the performance of the attention mechanism in improving the quality of illumination-enhanced images has been quantified from the experimental results using PIQE scores. In order to further explain the reasons for the decrease in PIQE scores in Table 4, representative images were selected and listed in Figure 8.





**Figure 8.** Attention processes and results: (**a**) the original images; (**b**) without attention mechanism; (**c**) with attention mechanism.

In terms of prediction performance, as a GAN based image translation model, authenticity and quality of the image after translation is benefits from the predictive ability of the model. For LighterGAN, the predictive ability can be reflected in the objective evaluation score of the image and visual intuitive experience. From the perspective of PIQE, shown in Table 3, the score of LighterGAN decreased by 0.44 benefit from the attention mechanism. Although the gap of score was not significant, it has improved the visual performance of image quality indeed, especially in the edge area near the scene. In examples in Figure 8, it could be found that after adding attention mechanism, the puzzling ripple phenomenon which should not exist in reality was alleviated to a large extent, therefore, the authenticity of enhanced image could be considered as being improved.

Compared with the attention mechanism which has little effect on PIQE scores, it could be seen from Table 3 that LighterGAN (without attention) is 4.56 lower than CycleGAN after adding semantic-consistency loss, which could be objectively identified as the most critical factor to improve image quality. Therefore, the loss calculation in the feature space is more beneficial to improve the quality of the LighterGAN generated image and the generalization ability of the model.

In the urban-wise UAV low illumination images, there are many interferences caused by lamps and reflected light from water or road surface which are affecting the image readability. From the results of the visual perception, it could be observed that LighterGAN can more effectively alleviate the influence of the light which reduces the readability of the image, this advantage could improve the potential of the image in subsequent processing. In terms of generalization ability, the advantage of LighterGAN was proved by both subjective evaluation and objective evaluation, which can be considered as an improvement brought by the combination of attention mechanism and semantic consistency loss.

In terms of disadvantages, LighterGAN could currently only process UAV images with limited resolution and due to the limitation of processing efficiency, the images cannot be processed real-timely. With the rapid development of image sensors carried by UAVs, real-time image enhancement technology for ultra-high resolution images will be a challenge. In the future, we plan to further optimize the network to achieve real-time image processing with higher resolution and to try to use the circular structure to achieve data augmentation.

## 6. Conclusions

Images captured by UAV aerial photography can be used by various computer vision algorithms to accomplish intelligent and complex urban observation or monitoring tasks, the premise of their performance is the satisfaction of image readability. In this article, we proposed LighterGAN, an image illumination enhancement method based on generative adversarial network, which could be used to relieve the frequently encountered image degeneration (the insufficient illumination and light pollution). Compared with five classic or deep learning based image enhancement algorithms, LighterGAN was found to have the best performance for visual perception and PIQE score. In the comparative experiment using the test set and the sub-dataset Y, the illumination enhancement performance of LighterGAN has exceeded the state-of-the-art EnlightenGAN and CycleGAN in both subjective and objective evaluations. Moreover, in the performance evaluation of structural improvements, the combination of authenticity subjective evaluation and objective PIQE assessments also demonstrates the enhanced generalization ability of LighterGAN over CycleGAN by using the images generated by two mappings.

**Author Contributions:** Conceptualization, J.W. and Y.H.; methodology, J.W. and Y.H.; software, J.W. and Y.Y.; formal analysis, Y.C.; writing—original draft preparation, Y.C. and Y.H.; writing—review and editing, Y.C. and Y.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by Shenzhen International Collaborative Research Project under Grant GJHZ20180929151604875. (Corresponding author: Yuxing Han). This research is also funded by a grant from the National Natural Science Foundation of Guangdong Province, China (No. 2018B030306026) and has been financially supported by the National Key Research and Development Program of China (No. 2018YFC1508200).

**Acknowledgments:** The authors would like to thank the editors and the reviewers for their valuable suggestions.

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

#### References

- Noor, N.M.; Abdullah, A.; Hashim, M. Remote sensing UAV/drones and its applications for urban areas: A review. In Proceedings of the IOP Conference Series: Earth and Environmental Science, 9th IGRSM International Conference and Exhibition on Geospatial & Remote Sensing (IGRSM), Kuala Lumpur, Malaysia, 24–25 April 2018.
- 2. Silva, B.N.; Khan, M.; Han, K. Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustain. Cities Soc.* **2018**, *38*, 697–713. [CrossRef]
- Wang, W.; Wei, C.; Yang, W.; Liu, J. GLADNet: Low-Light Enhancement Network with Global Awareness. In Proceedings of the Automatic Face & Gesture Recognition (FG) & 2018 13th IEEE International Conference, Xi'an, China, 15–19 May 2018.
- Yang, H.; Chen, P.; Huang, C.; Zhuang, Y.; Shiau, Y. Low complexity underwater image enhancement based on dark channel prior. In Proceedings of the 2011 Second International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA), Shenzhen, China, 16–18 December 2011; pp. 17–20.
- 5. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep Image Prior. *arXiv* 2017, arXiv:1711.10925v3. [CrossRef]
- 6. Ren, W.; Liu, S.; Ma, L.; Xu, Q.; Xu, X.; Cao, X.; Du, J.; Yang, M.H. Low-light image enhancement via a deep hybrid network. *IEEE Trans. Image Process.* **2019**, *28*, 4364–4375.
- 7. Lv, F.; Lu, F. Attention-guided low-light image enhancement. *arXiv* 2019, arXiv:1908.00682.
- 8. Guo, Y.; Ke, X.; Ma, J.; Zhang, J. A pipeline neural network for low-light image enhancement. *IEEE Access* 2019, 7, 13737–13744. [CrossRef]
- 9. Zhu, J.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2242–2251.
- Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
- Guo, X.J.; Li, Y.; Ling, H.B. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. Image Process.* 2017, 26, 982–993. [CrossRef]
- 12. Zhang, Q.; Nie, Y.; Zheng, W.-S. Dual illumination estimation for robust exposure correction. *Comput. Graph. Forum* 2019, *38*, 243–252. [CrossRef]
- Parthasarathy, S.; Sankaran, P. An automated multi Scale Retinex with Color Restoration for image enhancement. In Proceedings of the 2012 National Conference on Communications (NCC), Kharagpur, India, 3–5 February 2012; pp. 1–5.
- 14. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *arXiv* 2019, arXiv:1906.06972.
- 15. Perception Based Image Quality Evaluator (PIQE) No-Reference Image Quality Score. Available online: https://www.mathworks. com/help/images/ref/piqe.html (accessed on 25 July 2020).

- 16. Zhao, S.; Liu, Z.; Lin, J.; Zhu, J.Y.; Han, S. Differentiable augmentation for data-efficient gan training. arXiv 2019, arXiv:2006.1073.
- 17. Kleinberg, R.; Ligett, K.; Piliouras, G.; Tardos, É. Beyond the Nash Equilibrium Barrier. In Proceedings of the Symposium on Innovations in Computer Science (ICS), Beijing, China, 7–9 January 2011; pp. 125–140.
- 18. Mohammadi, P.; Ebrahimi-Moghadam, A.; Shirani, S. Subjective and objective quality assessment of image: A survey. *arXiv* 2014, arXiv:1406.7799.
- Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
- Ma, Y.; Liu, Y.; Cheng, J.; Zheng, Y.; Ghahremani, M.; Chen, H. Cycle Structure and Illumination Constrained GAN for Medical Image Enhancement. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; pp. 667–677.
- Ganesan, P.; Xue, Z.; Singh, S.; Long, R.; Ghoraani, B.; Antani, S. Performance Evaluation of a Generative Adversarial Network for Deblurring Mobile-phone Cervical Images. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4487–4490.
- Pacot, M.P.B.; Marcos, N. Cloud Removal from Aerial Images Using Generative Adversarial Network with Simple Image Enhancement. In Proceedings of the 2020 3rd International Conference on Image and Graphics Processing, Singapore, 8 February 2020; pp. 77–81.
- 23. LIVE Image Quality Assessment Database Release. Available online: https://live.ece.utexas.edu/research/quality (accessed on 20 January 2020).
- 24. Kalayeh, M.M.; Marin, T.; Brankov, J.G. Generalization evaluation of machine learning numerical observers for image quality assessment. *IEEE Trans. Nucl. Sci.* 2013, 60, 1609–1618. [CrossRef]
- 25. Lasserre, J.A.; Bishop, C.M.; Minka, T.P. Principled Hybrids of Generative and Discriminative Models. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006.
- Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
- 27. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein gans. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.
- 28. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* 2015, arXiv:1511.06434.
- 29. Kaji, S.; Kida, S. Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol. Phys. Technol.* **2019**, *12*, 235–248. [CrossRef]
- 30. Isola, P.; Zhu, J.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. *arXiv* 2016, arXiv:1611.07004.
- Mejjati, Y.A.; Richardt, C.; Tompkin, J.; Cosker, D.; Kim, K.I. Unsupervised attention-guided image-to-image translation. In Proceedings of the 32th Annual Conference on Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 3693–3703.
- 32. Harmel, R.D.; King, K.W.; Haggard, B.E.; Wren, D.G.; Sheridan, J.M. Practical guidance for discharge and water quality data collection on small watershed. *Trans. Am. Soc. Agric. Eng.* **2006**, *49*, 937–948.
- 33. Unsplash. Available online: https://unsplash.com (accessed on 17 March 2020).
- 34. POND5. Available online: https://www.pond5.com (accessed on 25 July 2020).
- 35. ShutterStock. Available online: https://www.shutterstock.com/ (accessed on 20 July 2020).
- 36. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, ICML, Lille, France, 6–11 July 2015; pp. 448–456.
- 37. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* 2016, arXiv:1607.08022.
- 38. Im, J.; Jensen, J.R. A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote Sens. Environ.* **2005**, *99*, 326–340. [CrossRef]
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014.
- Kodirov, E.; Xiang, T.; Gong, S. Semantic autoencoder for zero-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3174–3183.
- Scherer, D.; Müller, A.; Behnke, S. Evaluation of pooling operations in convolutional architectures for object recognition. In Proceedings of the International Conference on Artificial Neural Networks, Thessaloniki, Greece, 15–18 September 2010; pp. 92–101.
- 42. Fukushima, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Netw.* **1988**, *1*, 119–130. [CrossRef]
- 43. Agarap, A.F. Deep learning using rectified linear units (relu). arXiv 2018, arXiv:1803.08375.

- Tan, C.; Sun, F.; Kong, T.; Fang, B.; Zhang, W. Attention-based Transfer Learning for Brain-computer Interface. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1154–1158.
- Shen, T.; Zhou, T.; Long, G.; Jiang, J.; Wang, S.; Zhang, C. Reinforced Self-Attention Network: A Hybrid of Hard and Soft Attention for Sequence Modeling. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI), Stockholm, Sweden, 13–19 July 2018; pp. 4345–4352.
- 46. Xue, W.; Hu, X.; Wei, Z.; Mei, X.; Chen, X.; Xu, Y. A fast and easy method for predicting agricultural waste compost maturity by image-based deep learning. *Bioresour. Technol.* **2019**, 290, 121761. [CrossRef]
- 47. Kingma, D.P.; Adam, B.A.J. A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.
- Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machines learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannan, GA, USA, 2–4 November 2016; pp. 265–283.