remote sensing

MDPI

Article

Enhancing Remote Sensing Image Super-Resolution Guided by Bicubic-Downsampled Low-Resolution Image

Minkyung Chung [1], Minyoung Jung [2] and Yongil Kim [1,*]

[1] Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea; mkjung4876@snu.ac.kr
[2] Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA; jung411@purdue.edu
* Correspondence: yik@snu.ac.kr; Tel.: +82-880-7364

Abstract: Image super-resolution (SR) is a significant technique in image processing as it enhances the spatial resolution of images, enabling various downstream applications. Based on recent achievements in SR studies in computer vision, deep-learning-based SR methods have been widely investigated for remote sensing images. In this study, we proposed a two-stage approach called bicubic-downsampled low-resolution (LR) image-guided generative adversarial network (BLG-GAN) for remote sensing image super-resolution. The proposed BLG-GAN method divides the image super-resolution procedure into two stages: LR image transfer and super-resolution. In the LR image transfer stage, real-world LR images are restored to less blurry and noisy bicubic-like LR images using guidance from synthetic LR images obtained through bicubic downsampling. Subsequently, the generated bicubic-like LR images are used as inputs to the SR network, which learns the mapping between the bicubic-like LR image and the corresponding high-resolution (HR) image. By approaching the SR problem as finding optimal solutions for subproblems, the BLG-GAN achieves superior results compared to state-of-the-art models, even with a smaller overall capacity of the SR network. As the BLG-GAN utilizes a synthetic LR image as a bridge between real-world LR and HR images, the proposed method shows improved image quality compared to the SR models trained to learn the direct mapping from a real-world LR image to an HR image. Experimental results on HR satellite image datasets demonstrate the effectiveness of the proposed method in improving perceptual quality and preserving image fidelity.

Keywords: remote sensing images; high-resolution satellite images; super-resolution; generative adversarial network; image transfer; bicubic downsampling

check for updates

Citation: Chung, M.; Jung, M.; Kim, Y. Enhancing Remote Sensing Image Super-Resolution Guided by Bicubic-Downsampled Low-Resolution Image. *Remote Sens.* **2023**, *15*, 3309. https://doi.org/10.3390/rs15133309

Academic Editors: Igor Yanovsky and Jing Qin

Received: 4 May 2023
Revised: 23 June 2023
Accepted: 26 June 2023
Published: 28 June 2023

1. Introduction

Image super-resolution (SR) refers to the task of reconstructing high-resolution (HR) images from their low-resolution (LR) counterparts [1], and it has great significance in image processing, enabling various downstream applications [2]. However, image super-resolution is a well-known ill-posed problem because a single LR image can correspond to multiple HR images. Recent SR studies have addressed this problem by leveraging deep learning networks and achieved remarkable performance improvements compared to conventional example-based methods [3], even in the absence of prior information [4–10].

Since the advent of the deep-learning-based SR approach [4], several studies have devised deeper networks using various learning strategies such as residual [6,9,11–14], recursive [6,7], and adversarial [9,15] learning. Deep-learning-based SR models can be categorized into two groups, including convolutional neural network (CNN)-based models and generative adversarial network (GAN)-based models [16]. The first deep-learning-based SR model, SRCNN, is a CNN-based model composed of three convolutional layers, each corresponding to patch extraction, nonlinear mapping, and reconstruction [4]. Following the success of the SRCNN, CNN-based models such as VDSR [6] and EDSR [11] have been widely developed to fully leverage the learning capability of deeper networks. Currently, a residual network with a stack of residual blocks is perceived as the basic structure of CNN-based SR models. Zhang et al. [14] proposed RDN, consisting of residual dense blocks with dense local connections to enhance the residual block. In [13], a channel attention mechanism was adopted for a residual structure, demonstrating a significant improvement in SR image quality. However, because most CNN-based models are trained to optimize pixel-wise losses, such as the mean squared error (MSE) loss or L1 loss, these models are prone to producing overly smoothed SR outputs with restrictions on realistic texture recovery [17].

Compared with CNN-based models, GAN-based SR models generate more realistic textures by introducing adversarial training into existing CNN-based models [18]. The basic principle of a GAN is to train two networks (generator and discriminator) simultaneously for opposite purposes. The discriminator is trained to distinguish real HR images from SR images, whereas the generator is trained to produce realistic SR images to fool the discriminator. SRGAN [9] and ESRGAN [15] integrated two additional losses (perceptual and adversarial losses) into a loss function and improved the perceptual quality of the SR results.

A GAN-based SR approach has also been employed for remote sensing image processing to improve the perceptual quality [19]. Jiang et al. [20] complemented a GAN-based model by incorporating a subnetwork for edge enhancement, which refines edge information from satellite image datasets. Furthermore, Rabbi et al. [21] proposed EESRGAN, which trained the SR and object detection networks end-to-end and attempted to enhance the SR performance by using the detector loss from the subsequent object detection network. Liu et al. [22] proposed SG-GAN to benefit from employing a downstream task network by applying a pre-trained saliency detection model to the outputs of the SR network.

In general, deep-learning-based SR methods require LR images and their corresponding HR images as the training dataset. However, owing to the difficulties in obtaining real-world LR-HR datasets, most SR studies have only used HR images and generated LR images by applying degradations to HR images. The most commonly used method for generating LR images from HR images is downsampling by bicubic interpolation with a predefined scale factor [6,8–10,12,15,23]. However, SR models trained on simple degradation do not reflect the properties of real-world degradation, and often result in deteriorated performance when applied to real-world LR images. Therefore, some researchers have attempted to alleviate the gap between simple downsampling and real-world image degradation by applying a blur kernel and noise [4,13,14,24–27]. Conversely, several tailored datasets have been constructed, such as RealSR [28], DRealSR [29], and SR-RAW [30], which are more targeted at real-world image super-resolution. These datasets comprise real-world LR-HR image pairs obtained by adjusting the camera's focal length. Similarly, deep-learning-based SR models for remote sensing images commonly use predefined degradation to generate synthetic LR-HR datasets for training and validation [21,22]. Some recent studies adopted a degrader [31] or downsample generator [32] in the deep-learning architecture and attempted to make the model learn image degradation and super-resolution.

For HR satellite images, image datasets are usually provided as pairs of panchromatic (PAN) and multispectral (MS) images. Thus, these paired images provide a favorable opportunity for constructing real-world LR-HR image datasets. In this study, to train and validate the proposed model on real-world LR-HR image datasets, we performed pansharpening [33] using paired PAN and MS images from WorldView-3 (WV3) to generate real-world LR-HR remote sensing image datasets. The pansharpened and original MS images were then used as the HR and LR images, respectively. The scale factor was set to 4, based on the scale ratio of the PAN and MS images. The experimental results from the overall study were obtained from SR models trained on real-world LR-HR image datasets. A detailed description of the datasets used in this study is provided in Section 3.1.

Figure 1 demonstrates the difference between real-world and synthetic LR images. The ground objects are discernible in the bicubic-downsampled LR images (Figure 1a), whereas the clarity of the object boundaries is diminished in the real-world LR images (Figure 1b) because of blurring. Therefore, SR models trained on synthetic LR images from bicubic downsampling often fail to achieve satisfactory SR performance on real-world LR images. Furthermore, we observed that the SR models demonstrated better SR performance when trained on synthetic LR-HR image datasets than when trained on real-world LR-HR image datasets (see Appendix A).

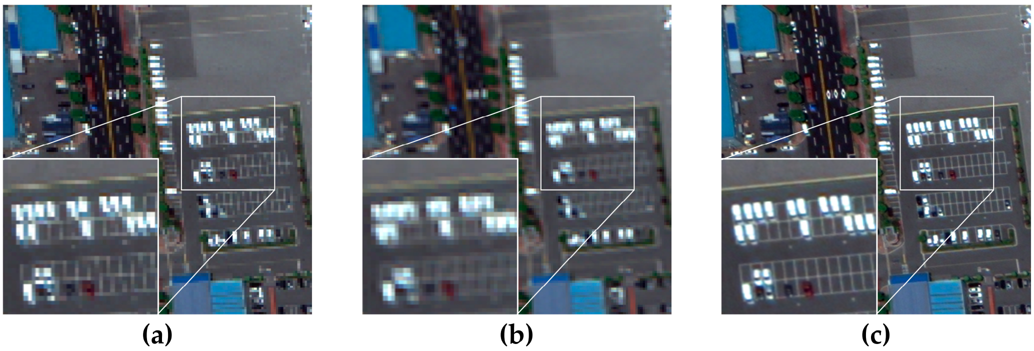


Figure 1. Comparison of the synthetic LR image with real-world LR image and HR image: (**a**) bicubic-downsampled LR image; (**b**) real-world LR image (MS image); (**c**) HR image (pansharpened MS image). For the convenience of comparison, the LR images are enlarged to the size of the HR images.

Based on these observations, we have inferred that refining the input LR image is as crucial as designing a complex SR network architecture to enhance the SR performance. Thus, this study proposed a bicubic-downsampled LR image-guided generative adversarial network (BLG-GAN) for the super-resolution of remote sensing images. The BLG-GAN performs super-resolution for real-world LR images under the guidance of clean synthetic LR images, obtained through a simple bicubic operation. By dividing the SR problem into subproblems with separate networks, the learning objective of each network becomes clearer. As a result, the training process of the BLG-GAN can be more stabilized than that of deep networks trained to learn a direct relationship between real-world LR and HR images.

To the best of our knowledge, this is the first study to introduce a training strategy that uses a synthetic LR image from bicubic downsampling to guide the supervised image super-resolution of remote sensing images. Moreover, we investigated the effectiveness of our method by comparing it with state-of-the-art methods and thoroughly analyzed the influence of its components on SR performance.

The remainder of this study is organized as follows. Section 2 presents the architecture of the proposed BLG-GAN model. Section 3 presents the experimental results of the WV3 datasets. In Section 4, the effectiveness of the proposed method was validated using ablation studies on the network architecture and type of loss. Finally, Section 5 presents the conclusions of this study.

2. Methodology

The proposed model aims to learn mapping from the real-world LR image domain X to the HR image domain Y from training using the given samples $x \in X$ and $y \in Y$ with the guidance of bicubic-downsampled LR images. While real-world LR image x is obtained from MS bands of WV3, a synthetic LR image is generated from HR images y with bicubic downsampling and denoted as $y_{\downarrow} \in Y_{\downarrow}$. Inspired by [34–36], we assumed $y_{\downarrow}$ as "clean LR image," which has less corruption in an image such as blur and noise. Thus, we used these bicubic-downsampled LR images as a bridge between the LR images and the corresponding HR images to restore clear details from the clean LR images. The prior application of image transfer to the input LR image is intended to reduce corruption within the real-world LR image and affects the quality of the output images from the following SR process.

As shown in Figure 2, the proposed BLG-GAN model consists of two stages: LR image transfer and super-resolution. In the LR image transfer stage, the LR images are processed through $G_{XY\downarrow}$ to generate LR images that have similar image characteristics or distributions with synthetic LR images, referred to as "bicubic-like LR images". The output of the LR image transfer stage is then fed into the generator with upsampling blocks ($G_{Y\downarrow Y}$) for super-resolution. Both stages include a generator and a discriminator to adopt adversarial training for the generation of bicubic-like LR and SR images. Each generator is trained to fool its corresponding discriminator and produce bicubic-like LR or SR images. Conversely, the discriminator is trained to distinguish whether the generated image is real or fake. The following subsections provide detailed explanations of each stage.

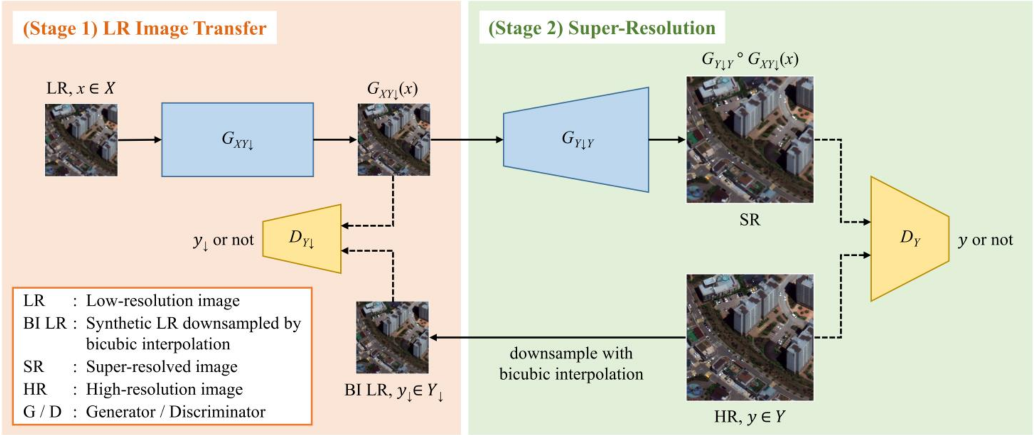

(Stage 1) LR Image Transfer
LR, $x \in X$
$G_{XY\downarrow}$
$G_{XY\downarrow}(x)$
$y_\downarrow$ or not
$D_{Y\downarrow}$
LR : Low-resolution image
BI LR : Synthetic LR downsampled by bicubic interpolation
SR : Super-resolved image
HR : High-resolution image
G / D : Generator / Discriminator
BI LR, $y_\downarrow \in Y_\downarrow$
(Stage 2) Super-Resolution
$G_{Y\downarrow Y} \circ G_{XY\downarrow}(x)$
$G_{Y\downarrow Y}$
SR
D_Y
y or not
downsample with bicubic interpolation
HR, $y \in Y$

Figure 2. Overall network architecture of the proposed BLG-GAN. The proposed network consists of two stages: LR image transfer and super-resolution. In the LR image transfer stage, the generator ($G_{XY\downarrow}$) transfers real-world LR images to bicubic-like LR images. In the super-resolution stage, the subsequent generator ($G_{Y\downarrow Y}$) learns the relationship between bicubic-like LR images and HR images.

2.1. LR Image Transfer

In LR image transfer, generator $G_{XY\downarrow}$ learns the mapping from the LR domain X to the bicubic-like LR domain $Y_\downarrow$, as illustrated in Stage 1 of Figure 2. For the given input LR image x, $G_{XY\downarrow}$ generate a bicubic-like LR image $\hat{y}_\downarrow$, which looks similar to a synthetic LR image $y_\downarrow$. This LR image transfer process can be formulated as:

$$\hat{y}_\downarrow = G_{XY\downarrow}(x). \tag{1}$$

Using adversarial training, $G_{XY\downarrow}$ was trained to fool the corresponding discriminator, $D_{Y\downarrow}$, for the generated bicubic-like LR image ($\hat{y}_\downarrow$). In the meantime, $D_{Y\downarrow}$ is trained to discern the generated LR image $\hat{y}_\downarrow$ as fake and the synthetic LR image $y_\downarrow$ as real.

The generator loss for LR Image transfer consists of two different losses: pixel-wise loss L^{LR}_{pix} and adversarial loss L^{LR}_{adv}. The pixel-wise loss calculates the l_1-distance between $\hat{y}_\downarrow$ and $y_\downarrow$. We chose LSGAN [37] for adversarial loss, which uses the form of least squares loss instead of negative log-likelihood loss. The LSGAN is known to stabilize the learning process while achieving a higher SR performance than the standard GAN [38]. The two different losses are formulated as:

$$L^{LR}_{pix} = \frac{1}{N}\sum_{i=1}^{N} \|G_{XY\downarrow}(x_i) - y_{\downarrow i}\|_1, \tag{2}$$

$$L^{LR}_{adv} = \frac{1}{N}\sum_{i=1}^{N} \|D_{Y\downarrow}(G_{XY\downarrow}(x_i)) - 1\|_2, \tag{3}$$

where N denotes the number of training samples. The discriminator loss for $D_{Y\downarrow}$ can be formulated as:

$$L_D^{LR} = \frac{1}{N}\sum_{i=1}^{N} \|D_{Y\downarrow}(y_{\downarrow i}) - 1\|_2 + \|D_{Y\downarrow}(G_{XY\downarrow}(x_i))\|_2. \quad (4)$$

Finally, the total loss for generator $G_{XY\downarrow}$ can be expressed as the weighted sum of the pixel-wise loss (L_{pix}^{LR}) and adversarial loss (L_{adv}^{LR}),

$$L_G^{LR} = L_{pix}^{LR} + \omega_1 L_{adv}^{LR}, \quad (5)$$

where ω_1 is the weight of adversarial loss for LR images.

2.2. Super-Resolution

Using the LR image generated from the prior LR image transfer as the input, the generator for super-resolution ($G_{Y\downarrow Y}$) learns the mapping relationship from the bicubic-like LR domain $Y_\downarrow$ to the HR domain Y. As shown in Stage 2 of Figure 2, the output of $G_{XY\downarrow}$, which is a bicubic-like LR image $\hat{y}_\downarrow$, is input into the SR network $G_{Y\downarrow Y}$ to produce an SR image $\hat{y}$. In the training phase, the discriminator D_Y interacts with $G_{Y\downarrow Y}$ and helps the network generate an SR image similar to the corresponding HR image, y. The super-resolution process can be formulated as follows:

$$\hat{y} = G_{Y\downarrow Y}(\hat{y}_\downarrow) = G_{Y\downarrow Y}(G_{XY\downarrow}(x)) = G_{Y\downarrow Y} \circ G_{XY\downarrow}(x). \quad (6)$$

We denote the consecutive processes of $G_{XY\downarrow}$ and $G_{Y\downarrow Y}$ as $G_{Y\downarrow Y}{}^\circ G_{XY\downarrow}$. As with the LR image transfer, $G_{Y\downarrow Y}$ is trained to fool the corresponding discriminator D_Y for the generated SR image $\hat{y}$, whereas D_Y is trained to distinguish the generated SR image $\hat{y}$ as fake and the ground truth HR image y as real.

The generator loss function for super-resolution consists of three different losses: pixel-wise loss L_{pix}^{HR}, perceptual loss L_{per}^{HR}, and adversarial loss L_{adv}^{HR}. Similar to the LR image transfer, we chose the L1 norm for pixel-wise loss and LSGAN for adversarial loss. The pixel-wise and adversarial losses for HR image are formulated as:

$$L_{pix}^{HR} = \frac{1}{N}\sum_{i=1}^{N} \|G_{Y\downarrow Y}(\hat{y}_{\downarrow i}) - y_i\|_1 \quad (7)$$

$$L_{adv}^{HR} = \frac{1}{N}\sum_{i=1}^{N} \|D_Y(G_{Y\downarrow Y}(\hat{y}_{\downarrow i})) - 1\|_2. \quad (8)$$

The discriminator loss for $D_{Y\downarrow}$ can then be formulated as:

$$L_D^{HR} = \frac{1}{N}\sum_{i=1}^{N} \|D_Y(y_i) - 1\|_2 + \|D_Y(G_{Y\downarrow Y}(\hat{y}_{\downarrow i}))\|_2. \quad (9)$$

Additionally, we added the perceptual loss L_{per}^{HR} between $\hat{y}$ and y. For perceptual loss, we adopted the learned perceptual image patch similarity (LPIPS) [39], which measures the perceptual similarity of images with multi-layer features. Recent SR studies have verified the usefulness of the LPIPS as a perceptual loss measure by achieving high ranks in challenges on SR tasks [40,41]. In Section 4.3, we also compare perceptual loss with LPIPS and the commonly used VGG-based perceptual loss.

The total loss for generator $G_{Y\downarrow Y}$ can be expressed as the weighted sum of L_{pix}^{HR}, L_{adv}^{HR}, and L_{per}^{HR}

$$L_G^{HR} = L_{pix}^{HR} + \lambda_1 L_{adv}^{HR} + \lambda_2 L_{per}^{HR}, \quad (10)$$

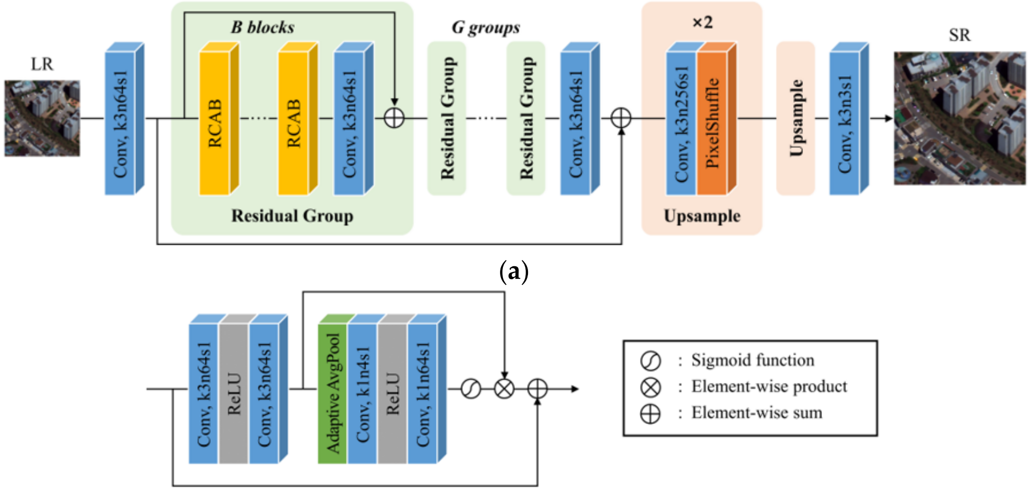
where λ_1 and λ_2 are the weights of adversarial loss and perceptual loss for HR images, respectively.

2.3. Network Architecture

The proposed SR network consists of two generators and two discriminators, each for LR image transfer and super-resolution. In this section, the architecture of each network component is described.

2.3.1. Generator

For the two generators ($G_{XY\downarrow}$ and $G_{Y\downarrow Y}$) in the proposed model, we adopted the network architecture from residual channel attention networks (RCAN) [13] (Figure 3), considering its superior SR performance even without a discriminator. RCAN is based on residual in residual (RIR) architecture with several residual groups and long skip connections. Each residual group comprises multiple residual channel attention blocks (RCABs). As shown in Figure 3b, RCAB integrates channel attention into the residual block to extract channel-wise features and achieves considerable enhancement in the image quality of the SR outputs. Further investigation of the effectiveness of the RCAN-based generator is addressed in Section 4.1 through a comparison with other generator architectures.



(**a**)

(**b**)

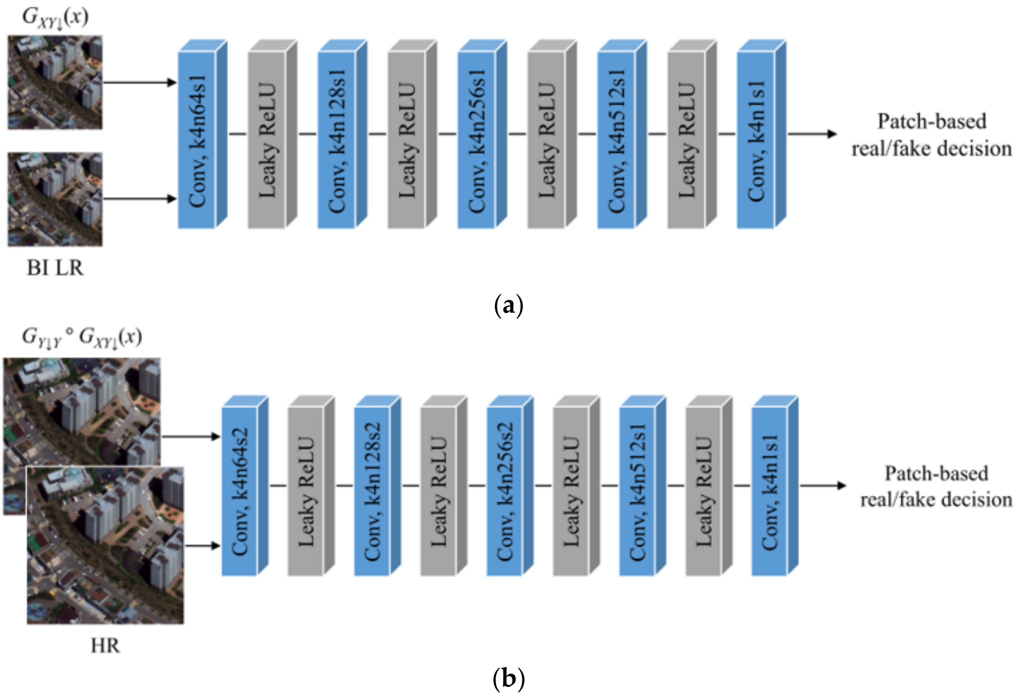
Figure 3. Architecture of the generators used in the proposed BLG-GAN: (**a**) generator with residual channel attention blocks (RCABs) [13]; (**b**) RCAB from the generator. $G_{XY\downarrow}$ and $G_{Y\downarrow Y}$ share the same framework composed of residual groups with RCABs. $G_{Y\downarrow Y}$ includes upsampling blocks to increase the size of the LR image by a factor of 4, whereas $G_{XY\downarrow}$ does not require upsampling blocks because the scale of input and output images do not change in LR image transfer.

Although the basic architecture for the two generators is almost identical, we adjusted the network capacity by setting the number of residual groups and the number of RCABs in each residual group to (5, 10) and (5, 20) for $G_{XY\downarrow}$ and $G_{Y\downarrow Y}$, respectively. Even though our total generative network ($G_{Y\downarrow Y}{}^{\circ}G_{XY\downarrow}$) is smaller in size than the original RCAN model with 10 residual groups and 20 RCABs for each residual group, BLG-GAN achieves superior SR performance by dividing the SR problem into subproblems. In addition, $G_{XY\downarrow}$ does not include upsampling blocks because the scales of the input and output images remain the same in the LR image transfer.

2.3.2. Discriminator

The discriminators $D_{Y\downarrow}$ and D_Y share the same discriminator structure based on patchGAN architecture [42] (Figure 4). The patchGAN consists of four convolutional layers with the number of features increasing from 64 to 512 by a factor of 2, followed by a final convolutional layer. The output features represent the patch-based decision of whether the image region is real or fake. To discriminate between the generated SR and HR images

(D_Y), we used a 70 × 70 patchGAN discriminator. For the discriminator for the generated bicubic-like LR images ($D_{Y\downarrow}$), we modified the stride of the first three convolutional layers of D_Y from two to one [34], because the size of the LR images is smaller than 70 × 70 pixels. As a result, the receptive field of $D_{Y\downarrow}$ is reduced to 16 × 16 for LR images.



(**a**)

(**b**)

Figure 4. Architecture of the discriminators used in the proposed BLG-GAN: (**a**) discriminator for LR image transfer ($D_{Y\downarrow}$); (**b**) discriminator for super-resolution (D_Y). $D_{Y\downarrow}$ and D_Y share the same discriminator structure based on patchGAN [42]. Considering the size of the input image, the stride of the first three convolution layers is set to one and two for $D_{Y\downarrow}$ and D_Y, respectively.

3. Experimental Results

Here, we describe the datasets used in this study and the quantitative assessment metrics used for the evaluation. Based on those metrics, the proposed method was compared with state-of-the-art techniques to verify the effectiveness of BLG-GAN.

3.1. Datasets

This study used HR satellite images from the WV3 sensor as the remote sensing images. The WV3 sensor provides PAN and MS bands with spatial resolutions of 0.31 m and 1.24 m, respectively. Two WV3 datasets, WV3-1 and WV3-2 (Figure 5), were generated using two WV3 images captured over the Pyeongdong Industrial Complex in Gwangju, Republic of Korea, with a temporal interval of approximately one year (26 May 2017 and 4 May 2018). The scene contained various land-cover types and objects, including urban areas, paddy fields, grasslands, forests, and rivers. To construct the LR-HR datasets, we adopted the Gram–Schmidt adaptive (GSA) algorithm [33] for pansharpening the paired PAN and MS images to generate HR images. The image quality assessment results for the pansharpened images are provided in Appendix B. For training and testing the SR models, the HR images were divided into sub-images of 512 × 512 pixels, which corresponded to sub-images of 128 × 128 pixels for the LR images. Consequently, the generated datasets comprised 1208 images for WV3-1 and 1136 images for WV3-2. These datasets were split into training, validation, and test datasets, with each set containing 60%, 20%, and 20% of the total dataset, respectively.

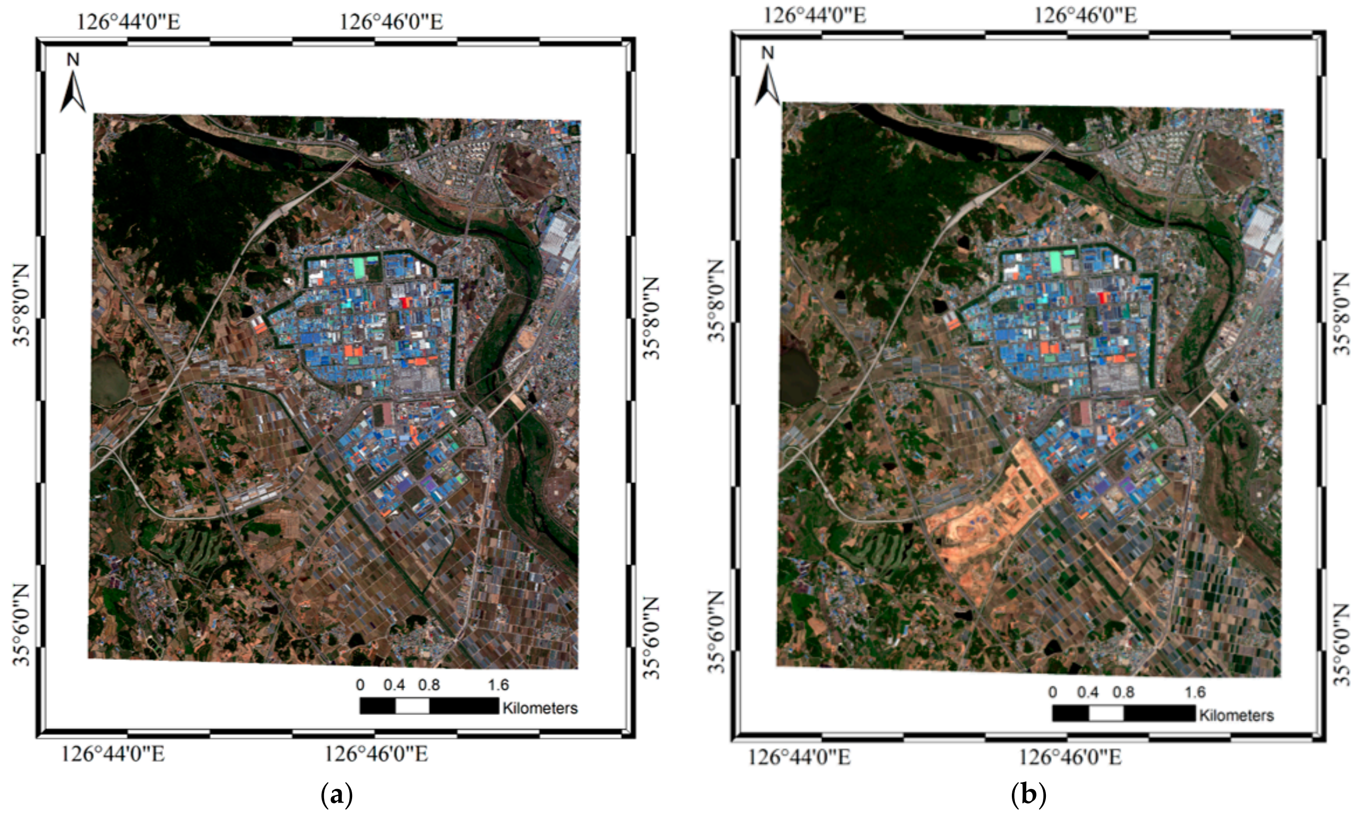


Figure 5. Remote sensing images used in this study; WorldView-3 (WV-3) images acquired over Pyeongdong Industrial Complex located in Gwangju, Republic of Korea, on (**a**) 26 May 2017 and (**b**) 4 May 2018.

3.2. Quantitative Assessment Metrics

The SR results were evaluated using several image quality metrics, including the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) [43], LPIPS [39], spectral angle mapper (SAM) [44], erreur relative globale adimensionnelle de synthèse (ERGAS) [45], universal image quality index (UIQI) [46], and natural image quality evaluator (NIQE) [47].

The PSNR is calculated as:

$$\mathrm{PSNR}(\hat{y}, y) = 10\log_{10}\frac{MAX^2}{MSE(\hat{y}, y)}, \tag{11}$$

where *MAX* is the maximum pixel value of the image and *MSE* is the mean squared error between the SR image ($\hat{y}$) and the ground truth HR image (y).

SSIM [43] measures three properties of an image: luminance, contrast, and structural characteristics. SSIM is defined as:

$$\mathrm{SSIM}(\hat{y}, y) = \frac{(2\mu_{\hat{y}}\mu_y + c_1)(2\sigma_{\hat{y}y} + c_2)}{\left(\mu_{\hat{y}}^2 + \mu_y^2 + c_1\right)\left(\sigma_{\hat{y}}^2 + \sigma_y^2 + c_2\right)}, \tag{12}$$

where $\mu_{\hat{y}}$ and μ_y are the means of $\hat{y}$ and y, respectively. $\sigma_{\hat{y}}$ and σ_y are the standard deviations of $\hat{y}$ and y, respectively, and $\sigma_{\hat{y}y}$ is the cross-covariance of images $\hat{y}$ and y. c_1 and c_2 are constants to prevent division by zero.

Although the PSNR and SSIM are the most widely used indices for evaluating the image quality of SR products, these conventional metrics focus on image fidelity rather than human perception. Therefore, we also used LPIPS, which was devised to reflect human perception and calculate the perceptual similarity of images [39]. LPIPS measures the perceptual similarity of image patches using pre-trained networks such as VGGNet and AlexNet. We used the pre-trained VGG-16 model to compute the l_2-distance of the features from multiple layers. The formulation of LPIPS is as follows:

$$\text{LPIPS}(\hat{y}, y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (f^l_{h,w} - f^l_{0,h,w})\|_2^2, \tag{13}$$

where $f^l_{h,w}$ and $f^l_{0,h,w}$ represent the features extracted from the lth layer at the locations (h, w) of images $\hat{y}$ and y, respectively. H_l and W_l are the height and width of the features from the lth layer, respectively. w_l is a learned weight vector and $\odot$ represents the element-wise product. Although higher PSNR and SSIM values indicate better image quality, a low LPIPS value is desirable because it measures the distance between the features of the input images.

In addition, SAM, ERGAS, and UIQI are conventional image quality assessment metrics that also focus on image fidelity rather than perceptual quality. SAM [44] measures the spectral angle between two images by calculating the dot product divided by the l_2-norm of each image. As the input images show high similarity, the SAM value approaches zero. The SAM is defined as

$$\text{SAM}(\hat{y}, y) = \arccos\left(\frac{\hat{y} \cdot y}{\|\hat{y}\|_2 \|y\|_2}\right). \tag{14}$$

ERGAS [45] measures the image quality in terms of the band-wise normalized mean error. A lower ERGAS value indicates higher image quality. The formulation of ERGAS is as follows:

$$\text{ERGAS}(\hat{y}, y) = 100\text{s}\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{RMSE(\hat{y}_k, y_k)}{\overline{y}_k}\right)}, \tag{15}$$

where s and N represent the scale factor and the number of spectral bands of the images being evaluated, respectively.

Wang and Bovik [46] proposed UIQI, which models image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Higher UIQI values that are close to one indicate better image quality. The UIQI is calculated as:

$$\text{UIQI}(\hat{y}, y) = \frac{4\sigma_{\hat{y}y}\mu_{\hat{y}}\mu_y}{\left(\mu_{\hat{y}}^2 + \mu_y^2\right)\left(\sigma_{\hat{y}}^2 + \sigma_y^2\right)}. \tag{16}$$

To evaluate the quality of the SR results, we also employed a no-reference quality metric. Unlike the previously mentioned metrics, a no-reference quality metric does not require a reference image for image quality assessment. NIQE [47] operates by extracting the natural scene statistics (NSS) features from the image patches and fitting them with a multivariate Gaussian (MVG) model. The derived MVG model is then compared with the MVG model obtained from a natural image database. A lower NIQE value indicates better image quality.

3.3. Implementation Details

To stabilize the training process, two generators in the proposed model ($G_{XY\downarrow}$ and $G_{Y\downarrow Y}$) were pre-trained using only the pixel-wise loss (L1 loss). Subsequently, based on the pre-trained generators, discriminators were added for adversarial training and jointly trained to generate SR images from real-world LR images (MS images).

In the training phase, we randomly cropped eight LR image patches with a size of 32 × 32 pixels for every iteration and augmented the data with random flip (horizontal and vertical) and rotation (90°, 180°, or 270°) to complement the limited number of training samples. The generators and discriminators were trained using the Adam optimizer with β_1 = 0.5, β_2 = 0.999, and ε = 10^{-8}, except for $G_{Y\downarrow Y}$, which uses β_1 = 0.9 instead. The learning rate was initialized to 10^{-4} and halved every 100 epochs. To ensure a fair comparison among different SR models, we trained all the models from scratch on our training datasets rather than using the pre-trained models. We set the weights for loss as ω_1 = 0.001, λ_1 = 0.001, and λ_2 = 0.01.

3.4. Comparison with State-of-the-Art Methods

To validate the effectiveness of the proposed BLG-GAN, we implemented several state-of-the-art methods, including seven CNN-based methods and five GAN-based methods. The CNN-based methods implemented were EDSR [11], D-DBPN [12], RRDB [15], RDN [14], RCAN [13], HAN [27], and DRN-L [48]. Additionally, we implemented the following GAN-based methods: SRGAN [9], ESRGAN [15], ESRGAN-FS [40], EESRGAN [21], and SG-GAN [22]. For all these models, real-world LR images (MS images) were used as inputs, and the corresponding SR images were obtained directly from a single SR network or generator.

SRGAN [9] and ESRGAN [15] are widely recognized studies that introduced GANs to solve the SR problem, and their basic structures have been adopted by many researchers. ESRGAN-FS [40] builds upon the structure of ESRGAN and incorporates a frequency separation training strategy. As these GAN-based models were originally developed for images used in computer vision, we also compared our method with two GAN-based SR methods designed for remote sensing images: EESRGAN [21] and SG-GAN [22]. EESRGAN is an improved model based on EEGAN [20], which aims to enhance the edges extracted from an image by adding an edge-enhancement network to the back of the generator. On the other hand, the SG-GAN adopts a salient object detection network [49] at the rear of the SRGAN-based generator, leveraging saliency information to generate more detailed SR outputs.

The quantitative evaluation results for the WV3-1 and WV3-2 datasets are presented in Tables 1 and 2. Consistent with previous research [9,40], CNN-based methods tend to achieve high PSNR and SSIM values but they also exhibit high LPIPS values. This is because using only pixel-wise loss (e.g., MSE loss or L1 loss) for the SR model training often results in blurry and overly smoothed SR outputs with low perceptual quality. In contrast, GAN-based methods generated visually pleasing SR results with low LPIPS and NIQE values. However, this comes at the expense of decreased PSNR, SSIM, and UIQI values, as well as increased SAM and ERGAS values, which can be attributed to the introduction of pseudo-texture through adversarial training. Therefore, it is crucial to effectively suppress the pseudo-texture while preserving high image fidelity to construct a successful GAN-based SR model. Furthermore, it is worth noting that, although GAN-based methods yield better NIQE values than CNN-based methods, it can be difficult to distinguish subtle performance differences among the CNN-based or GAN-based methods using NIQE alone. This limitation arises from the inherent nature of NIQE as a no-reference image quality index derived from a natural image database [47]. As remote sensing images have distinct image characteristics compared to natural images, the evaluation results using NIQE often deviate from human perceptions [31,50]. Hence, it is preferable to consider the limitations associated with using a no-reference index when evaluating the performance of SR methods for remote sensing images.

Table 1. Quantitative comparison with state-of-the-art methods on the WV3-1 dataset. The best and the second-best performances for each method are indicated in bold and underlined, respectively.

Method		PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
Bicubic		30.2986	0.8173	0.0242	63.0769	0.4173	0.3545	7.2993
CNN-based	EDSR [11]	31.9558	0.8586	0.0217	52.4538	0.4952	0.3247	7.3323
	D-DBPN [12]	31.1050	0.8397	0.0248	57.6184	0.4528	0.3390	**7.2362**
	RRDBNet [15]	31.7101	0.8540	0.0238	53.9245	0.4855	0.3288	7.7258
	RDN [14]	32.5940	0.8704	0.0223	49.0862	0.5219	0.3092	7.3097
	RCAN [13]	32.1932	0.8626	0.0234	51.1209	0.5066	0.3107	7.2772
	HAN [27]	**32.8207**	**0.8752**	**0.0215**	**47.8851**	**0.5359**	**0.2980**	7.5154
	DRN-L [48]	32.0414	0.8615	0.0221	51.9685	0.5014	0.3222	7.7383
GAN-based	SRGAN [9]	29.1961	0.7702	0.0560	72.4688	0.3420	0.3231	**4.8997**
	ESRGAN [15]	29.2197	0.7892	0.0449	72.1651	0.3904	0.2870	5.0202
	ESRGAN-FS [40]	28.9710	0.7827	0.0504	74.3983	0.3881	0.2852	4.9360
	EESRGAN [21]	30.4883	0.8138	0.0350	62.1157	0.4329	0.2669	5.5291
	SG-GAN [22]	30.9505	0.8310	0.0293	58.6363	0.4378	0.3073	5.5822
	BLG-GAN (1-stage)	31.4131	0.8373	0.0272	55.8005	0.4557	0.2740	5.7224
	BLG-GAN	**32.1416**	**0.8518**	**0.0247**	**51.8453**	**0.4883**	**0.2349**	5.7999

Table 2. Quantitative comparison with state-of-the-art methods on the WV3-2 dataset. The best and the second-best performances for each method are indicated in bold and underlined, respectively.

Method		PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
Bicubic		30.1314	0.8158	0.0246	54.9347	0.4392	0.3601	7.2807
CNN-based	EDSR [11]	31.5149	0.8509	0.0244	47.0863	0.5032	0.3342	7.9368
	D-DBPN [12]	31.1819	0.8423	0.0249	48.8536	0.4822	0.3444	7.9284
	RRDBNet [15]	31.2424	0.8463	0.0245	48.4882	0.4920	0.3424	8.2237
	RDN [14]	31.7136	0.8566	0.0248	46.0044	0.5136	0.3246	7.8459
	RCAN [13]	31.5306	0.8525	0.0235	46.8795	0.5048	0.3270	**7.6668**
	HAN [27]	**31.8671**	**0.8612**	0.0239	**45.2343**	**0.5245**	**0.3151**	7.9126
	DRN-L [48]	31.5219	0.8537	**0.0231**	46.9640	0.5080	0.3310	8.0392
GAN-based	SRGAN [9]	28.9943	0.7653	0.0544	62.8320	0.3508	0.3301	5.1271
	ESRGAN [15]	29.0857	0.7798	0.0582	62.1098	0.4006	0.3036	**4.9066**
	ESRGAN-FS [40]	29.2271	0.7899	0.0435	61.7939	0.4184	0.2894	5.1028
	EESRGAN [21]	30.3289	0.8076	0.0339	53.6708	0.4381	0.2781	5.3389
	SG-GAN [22]	30.4923	0.8200	0.0312	52.7853	0.4532	0.3181	5.4784
	BLG-GAN (1-stage)	30.8558	0.8267	0.0306	50.4641	0.4580	0.2858	5.5300
	BLG-GAN	**31.1871**	**0.8331**	**0.0272**	**48.8193**	**0.4769**	**0.2493**	5.6032

As shown in Tables 1 and 2, HAN [27] and RDN [14] exhibited superior SR performance among the CNN-based methods. Among the GAN-based methods, the proposed BLG-GAN model achieved superior SR performance for both the WV3-1 and WV3-2 datasets. Although HAN shows better image quality in terms of image fidelity than BLG-GAN, the LPIPS and NIQE values of HAN are significantly higher than those of BLG-GAN. This indicates a limitation of CNN-based methods in respect of perceptual quality. To further investigate the performance, we also implemented a one-stage version of the BLG-GAN model, denoted as "BLG-GAN (1-stage)" in Tables 1 and 2. This one-stage model consists of $G_{Y\downarrow Y}$ and D_Y, which employ the same network structures as the proposed two-stage BLG-GAN model. The one-stage BLG-GAN is intended to learn a direct relationship between real-world LR and HR images. While the one-stage BLG-GAN model outperforms other GAN-based models, the two-stage BLG-GAN model achieves superior SR performance in terms of both image fidelity and perceptual quality. These experimental results verify that the proposed BLG-GAN can generate clearer SR images than other methods by utilizing bicubic-like LR images obtained through LR image transfer as input to the SR model. In particular, our method significantly reduces the LPIPS values while maintaining high values for image fidelity metrics, outperforming all other GAN-based methods. The enhancement of the perceptual quality of the SR outputs can also be observed in Figure 6.

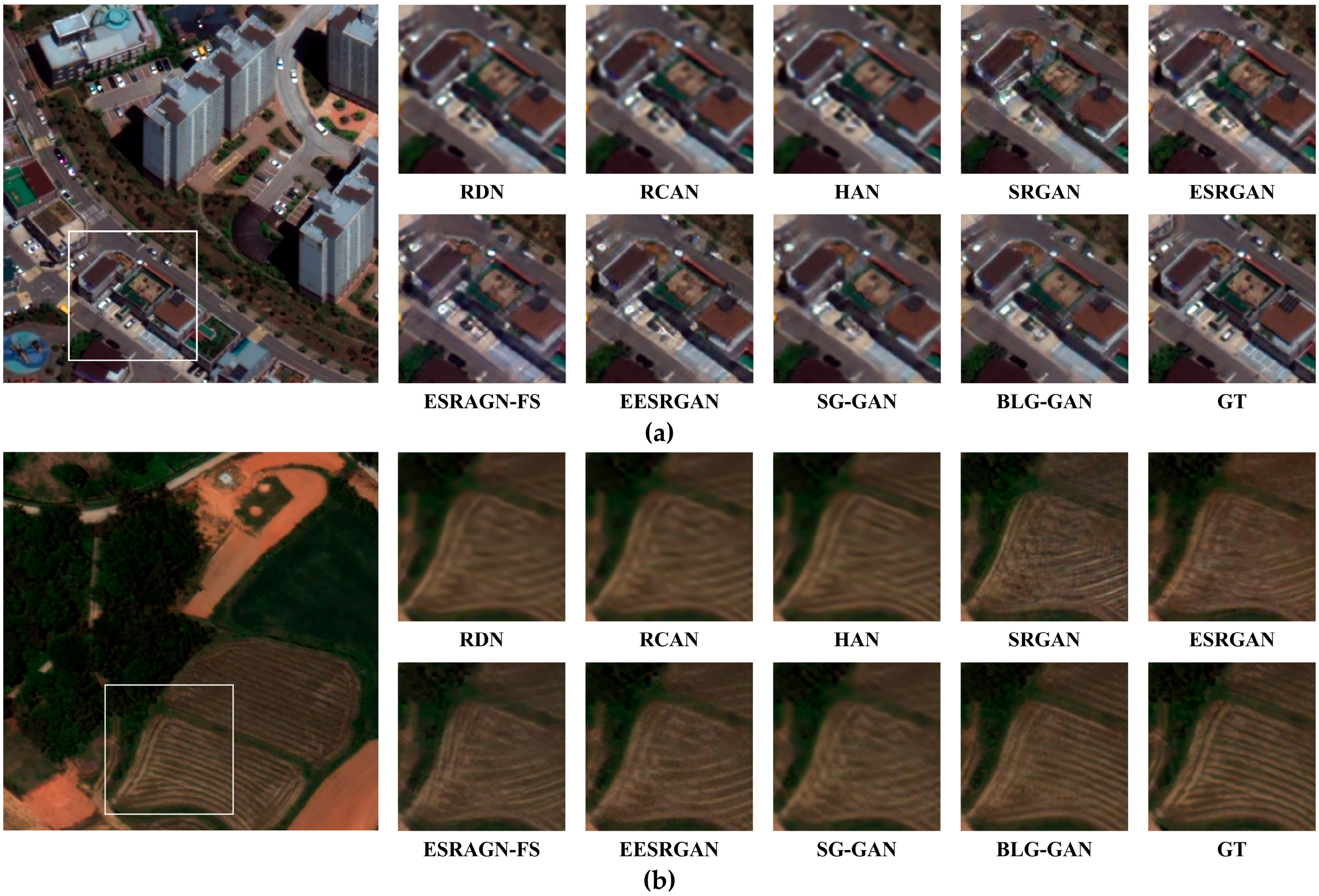


Figure 6. Visual comparison on two remote sensing datasets with a scale factor of four. Examples of SR results from the (**a**) WV3-1 and (**b**) WV3-2 datasets.

Figure 7 illustrates the SR process of the BLG-GAN model for real-world remote sensing images using LR image transfer to bicubic-like LR images. Once the real-world LR image (Figure 7a) is fed into the image transfer network, the input image is restored to a less blurry bicubic-like LR image (Figure 7b) with sharper edges. The subsequent SR network can use these edges to generate SR images with clear details. As shown in Figure 7c, the BLG-GAN successfully recovers various ground features, including road lanes, parking lines in the parking lot, and the rectangular shape of building roofs.

Furthermore, we evaluated the computational efficiency of BLG-GAN by considering the number of network parameters (M) and SR performance. As illustrated in Figure 8, the CNN-based models showed higher PSNR and LPIPS values with fewer network parameters than the GAN-based models. This is because GAN-based models incorporate additional parameters from the discriminator. Most GAN-based models showed slightly lower PSNR values than the CNN-based models while improving the perceptual quality of the SR outputs, as indicated by low LPIPS values. Remarkably, BLG-GAN achieved superior results in terms of both PSNR and LPIPS compared to the state-of-the-art models, even with a smaller overall capacity of the SR network. This indicates that our approach of dividing the SR problem into subproblems is valid for real-world remote sensing images.

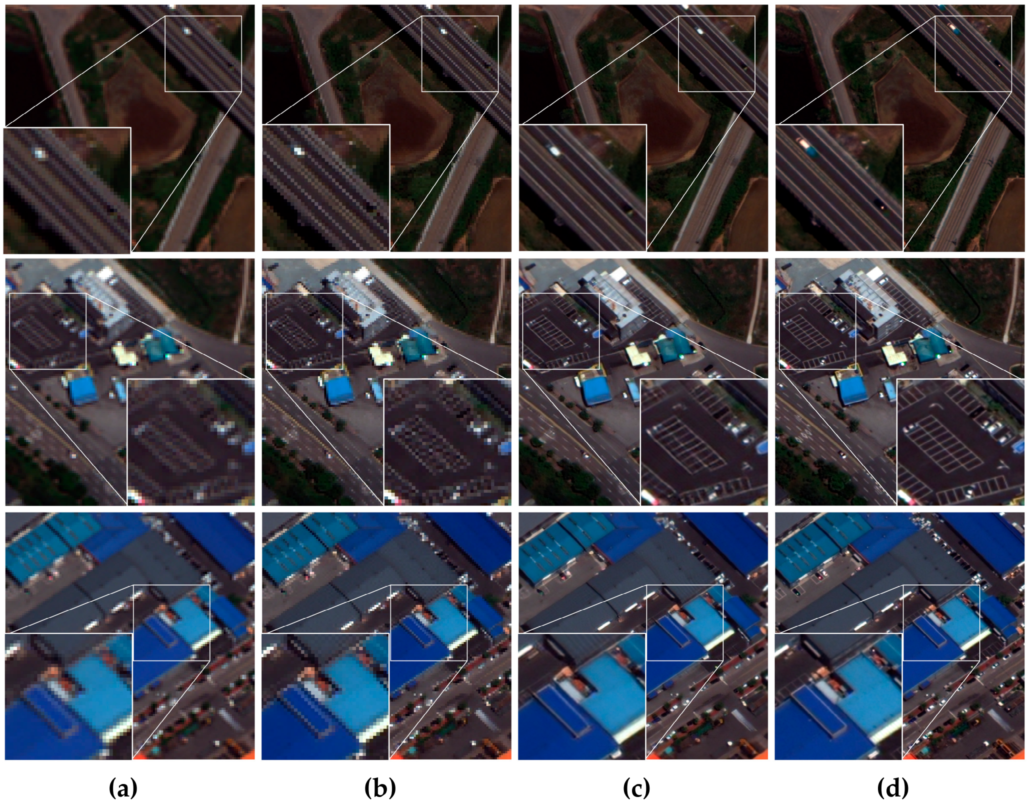


Figure 7. Examples of SR results from BLG-GAN: (**a**) input real-world LR images; (**b**) the generated bicubic-like LR images; (**c**) output SR images; (**d**) reference HR images.

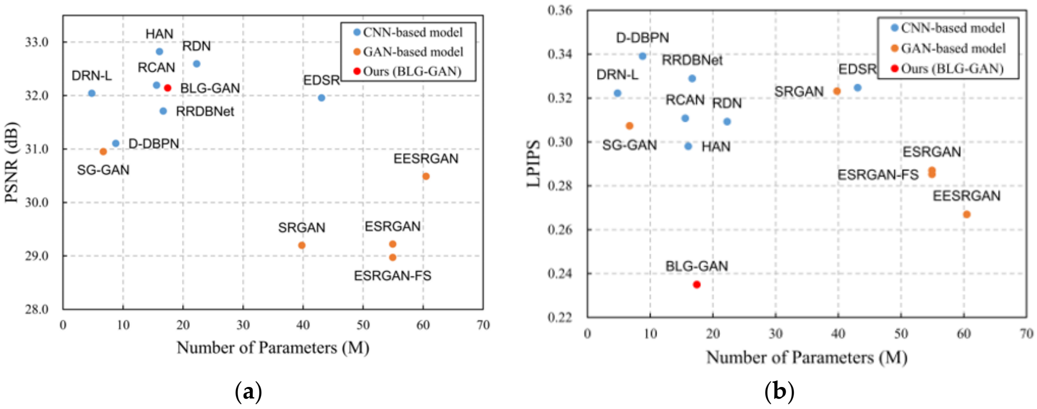


Figure 8. Comparison of the number of network parameters (M) and SR performance using (**a**) PSNR and (**b**) LPIPS metrics. The proposed BLG-GAN model is indicated as a red point and the CNN- and GAN-based models are shown as blue and orange points, respectively.

4. Discussion

To verify the effectiveness of each component of the proposed BLG-GAN, we analyzed the influence of the generator and discriminator architectures, the type of GAN loss, and the type of perceptual loss on the SR performance. The final architecture and loss function of the BLG-GAN were determined based on the results of the following analyses.

4.1. Generator Architecture

We compared the SR performance of the generators $G_{Y\downarrow Y}$ using four different basic blocks: the residual block from SRResNet [9], the residual block based on ResNet-18/34, the residual in residual dense block (RRDB) from ESRGAN [15] (Figure 9), and RCAB from

RCAN [13] (Figure 3b). The residual block from SRRestNet consists of two convolutional layers followed by batch normalization (BN) and uses ParametricReLU (PReLU) as the activation function [9] (Figure 9a). The main difference between the residual blocks from SRResNet and ResNet-18/34 is whether the block contains BN layers or not. Previous studies [11,15] have shown that BN layers are preferentially removed from the SR network because normalization of the features can restrain the generalization ability and deteriorate the SR performance. The RRDB also eliminates BN layers from the block architecture and integrates dense connections into a multilevel residual network to increase the network capacity. As mentioned in Section 2.3, the RCAB was proposed as a basic block for RCAN [13]. The RCAN model is based on the RIR structure with multiple residual groups. Each residual group comprises RCABs, which utilize a channel attention mechanism to extract more informative features from the input.

The number of blocks was set to 16 for the generators using residual blocks from SRResNet and ResNet-18/34, following the configuration in [9]. For the RRDB, we used the same number of blocks as in the original study [15], which is 23. While the original RCAN model had 10 residual groups with 20 residual blocks [13], we reduced it to five residual groups with 20 residual blocks in our implementation. Despite the reduced network capacity, we could achieve satisfactory SR results from the generator. To ensure a fair comparison among the generators with different basic blocks, all the input LR images were obtained from the LR image transfer generator $G_{XY\downarrow}$, which is described in Section 2.1. In the training phase, we employed patchGAN [42] as the discriminator for the SR outputs and trained the generators using pixel-wise and adversarial losses in all cases.

From the evaluation results presented in Tables 3 and 4, it was confirmed that utilizing residual blocks without BN instead of residual blocks with BN can improve the SR performance, which is consistent with previous observations [11,15]. The generator that employed the RCAB exhibited a superior performance, achieving the highest values for PSNR, SSIM, and UIQI, as well as the lowest values for SAM, ERGAS, and LPIPS values for both datasets. As a result, RCAB was chosen as the basic block for $G_{Y\downarrow Y}$, and further analysis was conducted on discriminators and losses.

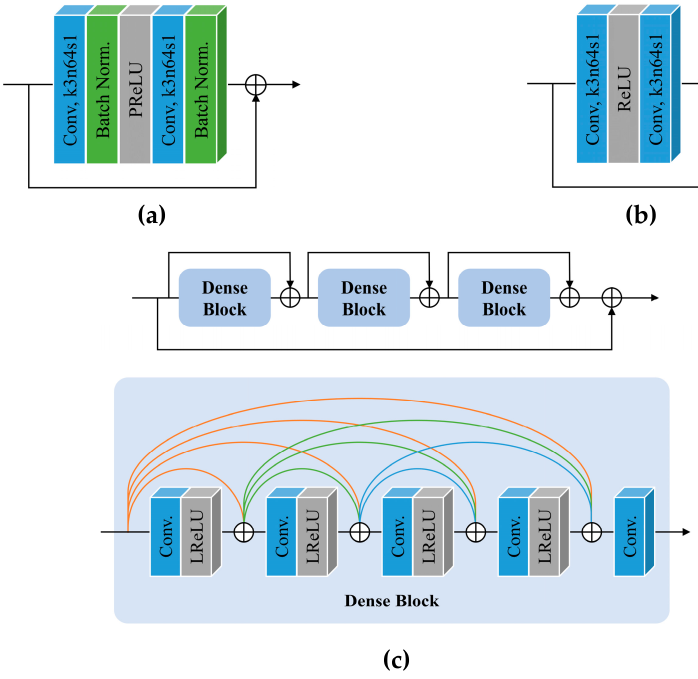


Figure 9. Basic blocks for generator $G_{Y\downarrow Y}$ used in this study to analyze the influence of generator architecture on SR performance: (**a**) residual block from SRResNet [9] (residual block with BN); (**b**) residual block based on ResNet-18/34 (residual block without BN); (**c**) residual in residual dense block (RRDB) [15].

Table 3. Analysis of the effect of the basic block type for generator $G_{Y\downarrow Y}$ on SR performance on the WV3-1 dataset. The best and second-best performances are indicated in bold and underlined, respectively.

Type of Basic Block for Generator $G_{Y\downarrow Y}$	**PSNR**	**SSIM**	**SAM**	**ERGAS**	**UIQI**	**LPIPS**	**NIQE**
Residual block with BN [9]	31.5495	0.8392	0.0361	60.1968	0.4645	0.2966	5.5730
Residual block without BN	32.0368	0.8486	0.0257	52.3808	0.4740	0.2845	5.6019
RRDB [15]	32.1078	0.8516	0.0255	51.9306	0.4831	0.2775	**5.2930**
RCAB [13]	**32.2062**	**0.8552**	**0.0242**	**51.4806**	**0.4927**	**0.2636**	5.8955

Table 4. Analysis of the effect of the basic block type for generator $G_{Y\downarrow Y}$ on SR performance on the WV3-2 dataset. The best and second-best performances are indicated in bold and underlined, respectively.

Type of Basic Block for Generator $G_{Y\downarrow Y}$	**PSNR**	**SSIM**	**SAM**	**ERGAS**	**UIQI**	**LPIPS**	**NIQE**
Residual block with BN [9]	30.8659	0.8255	0.0289	51.2175	0.4572	0.3057	5.4436
Residual block without BN	30.9909	0.8275	0.0269	49.7892	0.4620	0.3025	5.4305
RRDB [15]	31.1359	0.8308	0.0264	49.0844	0.4664	0.2953	**5.1734**
RCAB [13]	**31.2822**	**0.8362**	**0.0255**	**48.2146**	**0.4806**	**0.2815**	5.6038

4.2. Discriminator Architecture and GAN Loss

To compare the discriminator architectures, we selected two widely used networks, SRGAN [9] and patchGAN [42], for adversarial training of the SR network. The SRGAN discriminator (SRGAN-D) originates from [9] and is used in two state-of-the-art GAN-based SR methods, SRGAN and ESRGAN. SRGAN-D consists of eight convolutional layers with increasing features from 64 to 512 by a factor of two. The output features are then passed through two fully connected dense layers, followed by a sigmoid activation function. On the other hand, the patchGAN discriminator [42] provides a patch-based decision on whether the input image patch is real or fake. The detailed architecture of the patchGAN discriminator is presented in Section 2.3.2 (Figure 4b). We tested two types of GAN losses for both discriminators: the standard GAN loss [51] and the LSGAN loss [37]. Additionally, we applied the relativistic average GAN (RaGAN) [52] to SRGAN-D, as proposed in [15]. Besides the LSGAN, which is already explained in Section 2.1, the formulations for the standard GAN and RaGAN losses are provided below for comparison. The standard GAN loss for the generator (L_G) and discriminator (L_D) can be formulated as:

$$L_G = \frac{1}{N}\sum_{i=1}^{N} -\log D_Y\left(x_{f,i}\right), \tag{17}$$

$$L_D = \frac{1}{N}\sum_{i=1}^{N} -\log(D_Y(x_{r,i})) - \log\left(1 - D_Y\left(x_{f,i}\right)\right), \tag{18}$$

where N denotes the number of training samples. x_r and x_f represent the real data (HR image) and fake data (SR image), respectively. While standard GAN loss for the generator uses 1-side loss, RaGAN utilizes both real and fake data in adversarial training. The RaGAN losses for the generator (L_G^{Ra}) and discriminator (L_D^{Ra}) are formulated as:

$$L_G^{Ra} = -\mathbb{E}_{x_r}\left[\log\left(1 - D_Y^{Ra}\left(x_r, x_f\right)\right)\right] - \mathbb{E}_{x_f}\left[\log D_Y^{Ra}\left(x_f, x_r\right)\right], \tag{19}$$

$$L_D^{Ra} = -\mathbb{E}_{x_r}\left[\log D_Y^{Ra}\left(x_r, x_f\right)\right] - \mathbb{E}_{x_f}\left[\log\left(1 - D_Y^{Ra}\left(x_f, x_r\right)\right)\right], \tag{20}$$

where $D^{Ra}\left(x_r, x_f\right) = \sigma\left(C(x_r) - \mathbb{E}_{x_f}\left[C\left(x_f\right)\right]\right)$. σ is the sigmoid function and $C(x)$ is the output of the discriminator before applying the final sigmoid function. $\mathbb{E}_{x_f}[\cdot]$ means averaging the inputs for fake data (x_f) in the batch.

Based on the evaluation results presented in Tables 5 and 6, we verified the effectiveness of the PatchGAN discriminator for SR model training. The SR results obtained using the PatchGAN discriminator showed better LPIPS values than SRGAN-D for all types of GAN loss, while maintaining high values for PSNR, SSIM, and UIQI, and low values for SAM and ERGAS. Among the different types of GAN loss, LSGAN was found to improve image fidelity metrics more than the standard GAN. On the other hand, RaGAN showed inconsistent performance across the two test datasets, which can be attributed to the distortions introduced in the SR outputs by excessive pseudo-textures. Therefore, for our proposed BLG-GAN model, we selected PatchGAN as the discriminator and LSGAN loss as the GAN loss, considering the superior perceptual quality of the SR outputs along with reasonably good image fidelity.

Table 5. Analysis of the effect of type of discriminator architecture and GAN loss on SR performance on the WV3-1 dataset. The best and second-best performances are indicated in bold and underlined, respectively.

Type of Discriminator	Type of GAN Loss	PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
SRGAN-D [9]	Standard [51]	31.1322	0.8182	0.0345	58.0834	0.4363	0.2905	5.1780
	LSGAN [37]	**32.6198**	**0.8726**	**0.0219**	**48.9879**	**0.5287**	0.2977	7.9888
	RaGAN [52]	30.7263	0.8099	0.0380	61.2168	0.4356	0.2923	**5.1638**
PatchGAN [42]	Standard [51]	31.9389	0.8455	0.0255	53.1523	0.4805	0.2651	5.6226
	LSGAN [37]	32.2062	0.8552	0.0242	51.4806	0.4927	**0.2636**	5.8955

Table 6. Analysis of the effect of type of discriminator architecture and GAN loss on SR performance on the WV3-2 dataset. The best and second-best performances are indicated in bold and underlined, respectively.

Type of Discriminator	Type of GAN Loss	PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
SRGAN-D [9]	Standard [51]	30.4062	0.8084	0.0341	53.2263	0.4388	0.3042	**5.3235**
	LSGAN [37]	**31.6627**	**0.8575**	**0.0233**	**46.2162**	**0.5163**	0.3155	8.0493
	RaGAN [52]	30.5378	0.8237	0.0344	52.4700	0.4546	0.3233	5.6434
PatchGAN [42]	Standard [51]	30.9809	0.8253	0.0271	49.9844	0.4666	0.2852	5.3255
	LSGAN [37]	31.2822	0.8362	0.0255	48.2146	0.4806	**0.2815**	5.6038

4.3. Perceptual Loss

In most CNN-based methods, the perceptual quality of SR images is often limited due to the optimization using only pixel-wise MSE loss or L1 loss. To address this limitation and improve the perceptual quality of SR images, Ledig et al. [9] introduced perceptual loss in their GAN-based SR model. However, some studies [53] reported that the use of perceptual loss can introduce color variations and alter the original spectral information of the images. Therefore, we aimed to investigate the effect of different perceptual losses on the performance of SR models, using the model architecture determined in Section 4.1 and Section 4.2. We employed the same generator and discriminator architecture and tested three different perceptual losses: (1) VGG loss based on the pre-trained VGG-19 model, computed in the L2 norm ($L_{vgg19-L2}$), as proposed in [9]; (2) VGG loss computed in the L1 norm ($L_{vgg19-L1}$); and (3) perceptual loss based on LPIPS, which utilizes features extracted from the pre-trained VGG-16 model. The two VGG-19 model-based perceptual losses are defined as:

$$L_{vgg19-L1}(\hat{y}, y) = \|\phi(\hat{y}) - \phi(y)\|_1, \tag{21}$$

$$L_{vgg19-L2}(\hat{y}, y) = \|\phi(\hat{y}) - \phi(y)\|_2, \tag{22}$$

where $\phi(\cdot)$ represents the output features from the pre-trained VGG-19 model.

As shown in Tables 7 and 8, the utilization of any perceptual loss resulted in a slight decrease in the values of PSNR, SSIM, and UIQI, and an increase in the values of SAM and ERGAS. However, it significantly enhanced the perceptual quality of the SR images, as evidenced by the decreased values of LPIPS. Among the three perceptual losses tested, the LPIPS loss exhibited the highest values for image fidelity metrics and achieved a remarkable enhancement in perceptual image quality. These trends remained consistent across both test datasets, validating the effectiveness of employing LPIPS as a perceptual loss for improving the perceptual quality of SR images. The pre-trained models used for calculating perceptual loss were originally trained for high-level tasks, such as VGG models trained for classification. Leveraging these pre-trained models in training SR models proves highly beneficial because it enables the integration of high-level features into low-level tasks, such as image super-resolution.

Table 7. Analysis of the effect of type of perceptual loss on SR performance on the WV3-1 dataset. The best and second-best performances are indicated in bold and underlined, respectively.

Type of Perceptual Loss	PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
No perceptual loss	**32.2062**	**0.8552**	**0.0242**	**51.4806**	**0.4927**	0.2636	5.8955
LPIPS [39]	32.1416	0.8518	0.0247	51.8453	0.4883	**0.2349**	5.7999
VGG19-L1	32.0325	0.8474	0.0264	52.4761	0.4802	0.2451	6.0107
VGG19-L2	31.9628	0.8440	0.0278	52.8709	0.4734	0.2459	**5.7429**

Table 8. Analysis of the effect of type of perceptual loss on SR performance on the WV3-2 dataset. The best and second-best performances are indicated in bold and underlined, respectively.

Type of Perceptual Loss	PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
No perceptual loss	**31.2822**	**0.8362**	**0.0255**	**48.2146**	**0.4806**	0.2815	5.6038
LPIPS [39]	31.1871	0.8331	0.0272	48.8193	0.4769	**0.2493**	5.6032
VGG19-L1	31.0858	0.8319	0.0272	49.2414	0.4714	0.2629	5.5312
VGG19-L2	30.9904	0.8263	0.0293	49.8302	0.4617	0.2653	**5.5133**

5. Conclusions

In this study, we proposed a novel two-stage SR model for real-world remote sensing images. The proposed BLG-GAN method divides the image super-resolution procedure into two stages: LR image transfer and super-resolution. In the LR transfer stage, our proposed method refines the input LR images by transforming them into less blurry and noisy bicubic-like LR images using the guidance from synthetic LR images obtained through bicubic downsampling. The refined LR images are then fed into the SR network, which learns the relationship between the bicubic-like LR images and their corresponding HR images. By utilizing bicubic-downsampled LR images as a bridge between the real-world LR and HR images, our BLG-GAN method achieves a superior SR performance in terms of both image fidelity and perceptual quality. Moreover, since synthetic LR images can be easily obtained through bicubic downsampling, BLG-GAN can be easily implemented with a lower computational burden. In future studies, our method can be further validated using remote sensing images from other sources. Incorporating multi-source remote sensing images would enable the construction of large-scale datasets and facilitate comparisons with data-intensive models, which was not feasible in this study due to limited dataset size. Furthermore, the proposed method can be enhanced by integrating transfer learning techniques within the framework to address real-world remote sensing images without reference.

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

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant NRF-2023R1A2C2005548) and by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00155763).

Data Availability Statement: The remote sensing imagery used in this study (WorldView-3) is subject to a restrictive commercial license and cannot be shared publicly.

Acknowledgments: The Institute of Engineering Research at Seoul National University provided research facilities for this work.

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

Appendix A

Tables A1 and A2 provide quantitative assessment results of SR models trained on synthetic datasets. The results demonstrate that the SR models trained on synthetic LR-HR image datasets achieve better SR performance than those trained on real-world LR-HR image datasets. To ensure a fair comparison, all models were trained from scratch using the same hyperparameters for synthetic and real-world datasets. The comparison between Tables 1 and 2 in Section 3.4 and Tables A1 and A2 reveals that generating HR images from real-world LR images is more challenging than synthetic LR images.

Table A1. Quantitative assessment results of state-of-the-art SR models trained on synthetic LR-HR image dataset (WV3-1 dataset).

Method		PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
Bicubic		32.8364	0.8846	0.0206	47.5450	0.6291	0.2757	6.6302
CNN-based	EDSR [11]	34.3155	0.9103	0.0180	40.7123	0.6698	0.2300	7.8177
	D-DBPN [12]	33.8119	0.9006	0.0204	42.8439	0.6423	0.2534	7.4990
	RRDBNet [15]	33.9596	0.9045	0.0200	42.2957	0.6514	0.2488	7.5386
	RDN [14]	34.5049	0.9134	0.0185	39.8893	0.6755	0.2197	7.6263
	RCAN [13]	34.2489	0.9094	0.0191	41.0236	0.6650	0.2294	7.4470
	HAN [27]	34.6353	0.9155	0.0182	39.3325	0.6814	0.2142	7.8436
	DRN-L [48]	34.4271	0.9123	0.0186	40.2755	0.6740	0.2236	7.6602
GAN-based	SRGAN [9]	31.5388	0.8355	0.0525	56.1957	0.4871	0.2546	5.3869
	ESRGAN [15]	31.4596	0.8467	0.0452	56.5095	0.5274	0.2290	5.2891
	ESRGAN-FS [40]	31.4043	0.8491	0.0426	57.9827	0.5370	0.2188	5.3988
	EESRGAN [21]	32.6946	0.8755	0.0305	48.9042	0.5939	0.2046	5.8502
	SG-GAN [22]	32.3234	0.8629	0.0356	51.0380	0.5435	0.2655	5.5719

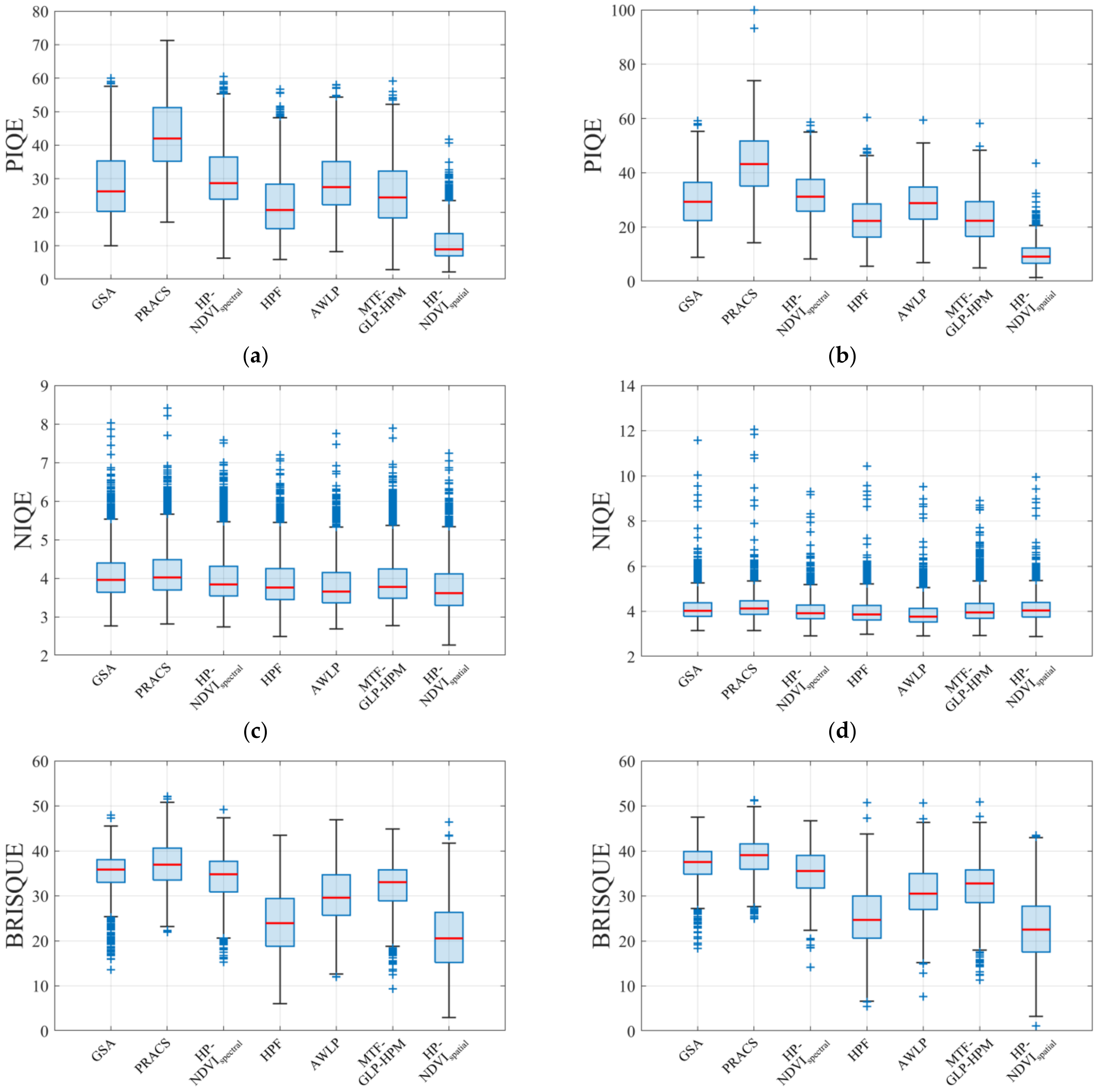
Table A2. Quantitative assessment results of state-of-the-art SR models trained on synthetic LR-HR image dataset (WV3-2 dataset).

Method		PSNR	SSIM	SAM	ERGAS	UIQI	LPIPS	NIQE
Bicubic		33.1650	0.8921	0.0185	38.7233	0.6612	0.2636	6.6402
CNN-based	EDSR [11]	34.4422	0.9127	0.0168	33.6200	0.6912	0.2325	7.6672
	D-DBPN [12]	34.0034	0.9046	0.0190	35.2938	0.6694	0.2509	7.2414
	RRDBNet [15]	34.1724	0.9090	0.0195	34.6322	0.6786	0.2435	7.1857
	RDN [14]	34.6515	0.9164	0.0172	32.8534	0.6982	0.2235	7.4615
	RCAN [13]	34.4469	0.9130	0.0176	33.6167	0.6911	0.2275	7.4895
	HAN [27]	34.8012	0.9188	0.0169	32.3318	0.7048	0.2174	7.4457
	DRN-L [48]	34.5861	0.9151	0.0173	33.0998	0.6972	0.2262	7.4913
GAN-based	SRGAN [9]	31.3895	0.8367	0.0484	47.7144	0.5120	0.2661	5.4253
	ESRGAN [15]	31.3861	0.8535	0.0418	47.9343	0.5639	0.2260	5.0752
	ESRGAN-FS [40]	31.6624	0.8567	0.0383	46.2197	0.5689	0.2183	5.4178
	EESRGAN [21]	33.0359	0.8793	0.0270	39.5130	0.6225	0.1987	5.6776
	SG-GAN [22]	33.8097	0.9038	0.0247	36.0339	0.6650	0.2388	7.1215

Appendix B

To construct the real-world LR-HR datasets, we evaluated several pansharpening methods, including component substitution (CS)-based, multiresolution analysis (MRA)-based, and hybrid methods. The CS-based methods employed were GSA [33], partial replacement adaptive component substitution (PRACS) [54], and hybrid pansharpening algorithm using NDVI in spectral mode (HP-NDVI$_{\text{spectral}}$) [55]. In addition, we implemented the following MRA-based methods: high pass filtering (HPF) algorithm [56], additive wavelet luminance proportional (AWLP) [57], and generalized Laplacian pyramid with modulation transfer function and high-pass modulation (MTF-GLP-HPM) algorithm [58]. Furthermore, the hybrid pansharpening algorithm using NDVI in spatial mode (HP-NDVI$_{\text{spatial}}$) [55] was also implemented as a hybrid method. The detailed explanation of each pansharpening method is beyond the scope of this study. For detailed methodological information, please refer to the original papers.

Due to the unavailability of reference HR images, the image quality of pansharpened images from the WV3-1 and WV3-2 datasets was evaluated using no-reference metrics, including perception-based image quality evaluator (PIQE) [59], NIQE [47], and blind/referenceless image spatial quality evaluator (BRISQUE) [60]. Lower values of PIQE, NIQE, and BRISQUE indicate better image quality. Figure A1 presents the box plots of the image quality assessment results obtained from different pansharpening methods. The MRA-based methods exhibited slightly lower PIQE and BRISQUE values than the CS-based methods. However, the MRA-based methods tended to generate blurry images in comparison to the CS-based methods, as depicted in Figure A2. This suggests that the blurriness in remote sensing images may be perceived as smoothness in natural images. Among the CS-based methods, GSA showed stable performance with low PIQE values and concentrated distributions for BRISQUE values. The distributions of NIQE values were similar across the different pansharpening methods, indicating no significant difference. In addition, the HP-NDVI$_{\text{spatial}}$ method exhibited superior performance in terms of PIQE and BRISQUE values, but the introduction of excessive spatial information often led to undesired artifacts such as pseudo-textures. Therefore, the GSA algorithm was chosen to generate HR images for the LR-HR datasets due to its stable performance and visually clear pansharpened images. Nevertheless, further investigation on large-scale datasets is necessary to generalize these observations.



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

(**c**) (**d**)

(**e**) (**f**)

Figure A1. Comparison of image quality assessment results of pansharpening images obtained from the WV3-1 dataset using (**a**) PIQE, (**c**) NIQE, and (**e**) BRISQUE, and from the WV3-2 dataset using (**b**) PIQE, (**d**) NIQE, and (**f**) BRISQUE. On each box, the red line indicates the median and the outliers are plotted as blue '+' markers.

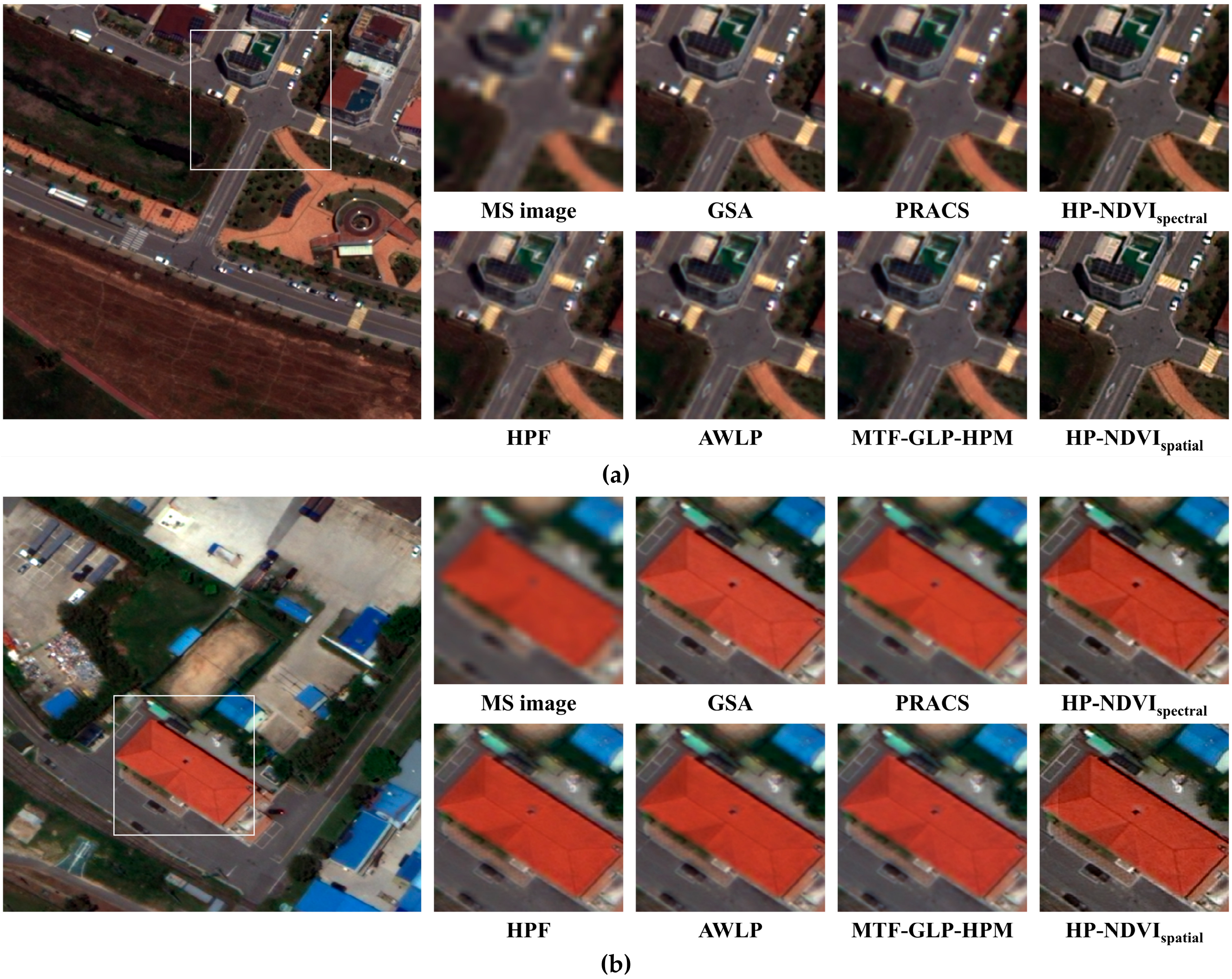


Figure A2. Visual comparison on two remote sensing datasets: Examples of pansharpened images obtained from the (**a**) WV3-1 and (**b**) WV3-2 datasets.

References

1. Freeman, W.T.; Pasztor, E.C.; Carmichael, O.T. Learning Low-Level Vision. *Int. J. Comput. Vis.* **2000**, *40*, 25–47. [CrossRef]
2. Yue, L.; Shen, H.; Li, J.; Yuan, Q.; Zhang, H.; Zhang, L. Image Super-Resolution: The Techniques, Applications, and Future. *Signal Process.* **2016**, *128*, 389–408. [CrossRef]
3. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-Based Super-Resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65. [CrossRef]
4. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]
5. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 391–407.
6. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
7. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
8. Tai, Y.; Yang, J.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3148–3155.
9. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.

10. Sajjadi, M.S.M.; Schölkopf, B.; Hirsch, M. EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4491–4500.
11. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
12. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep Back-Projection Networks for Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1664–1673.
13. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 294–310.
14. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
15. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision Workshops (ECCVW), Munich, Germany, 8–14 September 2018; pp. 63–79.
16. Ha, V.K.; Ren, J.-C.; Xu, X.-Y.; Zhao, S.; Xie, G.; Masero, V.; Hussain, A. Deep Learning Based Single Image Super-resolution: A Survey. *Int. J. Autom. Comput.* **2019**, *16*, 413–426. [CrossRef]
17. Chen, H.; He, X.; Qing, L.; Wu, Y.; Ren, C.; Sheriff, R.E.; Zhu, C. Real-World Single Image Super-Resolution: A Brief Review. *Inf. Fusion* **2022**, *79*, 124–145. [CrossRef]
18. Singla, K.; Pandey, R.; Ghanekar, U. A Review on Single Image Super Resolution Techniques Using Generative Adversarial Network. *Optik* **2022**, *266*, 169607. [CrossRef]
19. Jozdani, S.; Chen, D.; Pouliot, D.; Johnson, B.A. A Review and Meta-Analysis of Generative Adversarial Networks and Their Applications in Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102734. [CrossRef]
20. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-Enhanced GAN for Remote Sensing Image Superresolution. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5799–5812. [CrossRef]
21. Rabbi, J.; Ray, N.; Schubert, M.; Chowdhury, S.; Chao, D. Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network. *Remote Sens.* **2020**, *12*, 1432. [CrossRef]
22. Liu, B.; Zhao, L.; Li, J.; Zhao, H.; Liu, W.; Li, Y.; Wang, Y.; Chen, H.; Cao, W. Saliency-Guided Remote Sensing Image Super-Resolution. *Remote Sens.* **2021**, *13*, 5144. [CrossRef]
23. Lai, W.-S.; Huang, J.-B.; Ahuja, N.; Yang, M.-H. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5835–5843.
24. Zhang, K.; Zuo, W.; Gu, S.; Zhang, L. Learning Deep CNN Denoiser Prior for Image Restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2808–2817.
25. Zhang, K.; Zuo, W.; Zhang, L. Learning a Single Convolutional Super-Resolution Network for Multiple Degradations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3262–3271.
26. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.-T.; Zhang, L. Second-Order Attention Network for Single Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11057–11066.
27. Niu, B.; Wen, W.; Ren, W.; Zhang, X.; Yang, L.; Wang, S.; Zhang, K.; Cao, X.; Shen, H. Single Image Super-Resolution via a Holistic Attention Network. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 191–207.
28. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward Real-World Single Image Super-Resolution: A New Benchmark and a New Model. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3086–3095.
29. Wei, P.; Xie, Z.; Lu, H.; Zhan, Z.; Ye, Q.; Zuo, W.; Lin, L. Component Divide-and-Conquer for Real-World Image Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 101–117.
30. Zhang, X.; Chen, Q.; Ng, R.; Koltun, V. Zoom to Learn, Learn to Zoom. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3757–3765.
31. Zhang, N.; Wang, Y.; Zhang, X.; Xu, D.; Wang, X.; Ben, G.; Zhao, Z.; Li, Z. A Multi-Degradation Aided Method for Unsupervised Remote Sensing Image Super Resolution With Convolution Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [CrossRef]
32. Zhang, J.; Xu, T.; Li, J.; Jiang, S.; Zhang, Y. Single-Image Super Resolution of Remote Sensing Images with Real-World Degradation Modeling. *Remote Sens.* **2022**, *14*, 2895. [CrossRef]
33. Aiazzi, B.; Baronti, S.; Selva, M. Improving Component Substitution Pansharpening Through Multivariate Regression of MS+Pan Data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [CrossRef]

34. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised Image Super-Resolution Using Cycle-in-Cycle Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 814–823.
35. Lugmayr, A.; Danelljan, M.; Timofte, R. Unsupervised Learning for Real-World Super-Resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 3408–3416.
36. Maeda, S. Unpaired Image Super-Resolution Using Pseudo-Supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 291–300.
37. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.K.; Wang, Z.; Smolley, S.P. Least Squares Generative Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2813–2821.
38. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.K.; Wang, Z.; Smolley, S.P. On the Effectiveness of Least Squares Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2947–2960. [CrossRef]
39. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
40. Fritsche, M.; Gu, S.; Timofte, R. Frequency Separation for Real-World Super-Resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 3599–3608.
41. Jo, Y.; Yang, S.; Kim, S.J. Investigating Loss Functions for Extreme Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1705–1712.
42. 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. 1125–1134.
43. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
44. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Discrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. In Proceedings of the Summaries of the Third Annu. JPL Airborne Geoscience Workshop, Pasadena, CA, USA, 1–5 June 1992; pp. 147–149.
45. Liu, J.G. Smoothing Filter-based Intensity Modulation: A Spectral Preserve Image Fusion Technique for Improving Spatial Details. *Int. J. Remote Sens.* **2000**, *21*, 3461–3472. [CrossRef]
46. Wang, Z.; Bovik, A.C. A Universal Image Quality Index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
47. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [CrossRef]
48. Guo, Y.; Chen, J.; Wang, J.; Chen, Q.; Cao, J.; Deng, Z.; Xu, Y.; Tan, M. Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5406–5415.
49. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. BASNet: Boundary-Aware Salient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7471–7481.
50. Lei, S.; Shi, Z.; Zou, Z. Coupled Adversarial Training for Remote Sensing Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3633–3643. [CrossRef]
51. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
52. Jolicoeur-Martineau, A. The Relativistic Discriminator: A Key Element Missing from Standard GAN. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019; pp. 1–26.
53. Zhou, Y.; Deng, W.; Tong, T.; Gao, Q. Guided Frequency Separation Network for Real-World Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1722–1731.
54. Choi, J.; Yu, K.; Kim, Y. A New Adaptive Component-Substitution-Based Satellite Image Fusion by Using Partial Replacement. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 295–309. [CrossRef]
55. Choi, J.; Kim, G.; Park, N.; Park, H.; Choi, S. A Hybrid Pansharpening Algorithm of VHR Satellite Images that Employs Injection Gains Based on NDVI to Reduce Computational Costs. *Remote Sens.* **2017**, *9*, 976. [CrossRef]
56. Chavez, P.; Sides, S.C.; Anderson, J.A. Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic. *Photogramm. Eng. Remote Sens.* **1991**, *57*, 295–303.
57. Otazu, X.; Gonzalez-Audicana, M.; Nunez, O.F.J. Introduction of Sensor Spectral Response into Image Fusion Methods. Application to Wavelet-Based Methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2376–2385. [CrossRef]
58. Vivone, G.; Restaino, R.; Mura, M.D.; Licciardi, G.; Chanussot, J. Contrast and Error-Based Fusion Schemes for Multispectral Image Pansharpening. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 930–934. [CrossRef]

59. Venkatanath, N.; Praneeth, D.; Chandrasekhar, B.M.; Channappayya, S.S.; Medasani, S.S. Blind Image Quality Evaluation Using Perception Based Features. In Proceedings of the 21st National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015; pp. 1–6.
60. Mittal, A.; Moorthy, A.K.; Bovik, A.C. Blind/Referenceless Image Spatial Quality Evaluator. In Proceedings of the 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 6–9 November 2011; pp. 723–727.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.