aerospace

MDPI

Article

Remote Sensing Image Super-Resolution for the Visual System of a Flight Simulator: Dataset and Baseline

Wenyi Ge [1], Zhitao Wang [2], Guigui Wang [1], Shihan Tan [1] and Jianwei Zhang [3,*]

[1] National Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University, Chengdu 610000, China; Gwen.Scu@gmail.com (W.G.); tianshanhangui@126.com (G.W.); tanshihan_cq@163.com (S.T.)
[2] Beijing Satellite Navigation Center (BSNC), Beijing 100094, China; wangzhitao.nav@gmail.com
[3] College of Computer Science, Sichuan University, Chengdu 610000, China
* Correspondence: zhangjianwei@scu.edu.cn

Abstract: High-resolution remote sensing images are the key data source for the visual system of a flight simulator for training a qualified pilot. However, due to hardware limitations, it is an expensive task to collect spectral and spatial images at very high resolutions. In this work, we try to tackle this issue with another perspective based on image super-resolution (SR) technology. First, we present a new ultra-high-resolution remote sensing image dataset named Airport80, which is captured from the airspace near various airports. Second, a deep learning baseline is proposed by applying the generative and adversarial mechanism, which is able to reconstruct a high-resolution image during a single image super-resolution. Experimental results for our benchmark demonstrate the effectiveness of the proposed network and show it has reached satisfactory performances.

Keywords: flight simulator; remote sensing image; super-resolution; generative adversarial network

check for updates

Citation: Ge, W.; Wang, Z.; Wang, G.; Tan, S.; Zhang, J. Remote Sensing Image Super-Resolution for the Visual System of a Flight Simulator: Dataset and Baseline. *Aerospace* **2021**, *8*, 76. https://doi.org/10.3390/aerospace8030076

Academic Editor: Alexei Sharpanskykh

Received: 25 February 2021
Accepted: 11 March 2021
Published: 15 March 2021

1. Introduction

As is well known, air traffic control (ATC) is the key to ensuring the operational safety of air traffic, which highly depends on the collaboration between the air traffic controller (ATCO) and the aircrew [1]. The ATCO makes real-time decisions to direct the flight to its destination based on situational information from the ATC system, while the aircrew flies the aircraft in strict accordance with the ATCO's instruction, in an accurate and prompt manner [2]. Due to safety issues, both the ATCO and aircrew are required to be licensed by the concerned administration of their country. To obtain a valid license, they must meet specific requirements for being licensed. In addition, their skills will need to be re-examined at specified intervals. Thus, training equipment is indispensable for achieving the training of the ATCO or aircrew, and comprises an ATC simulator for the ATCO and a flight simulator for the aircrew.

Of these, the flight simulator has become a hot research topic due to its prominent significance related to flight in the air. The simulator is very important for ensuring flight safety, and is also able to greatly reduce equipment and maintenance costs [3,4]. The main purpose of the flight simulator is to provide realistic, real-time, immersive scenarios to complete the pilots' training before they fly a real aircraft. The training scenarios consist of various flight phases, including the airport ground, instrument landing, approach, and cruise. Furthermore, they also depend on the location of the target flight, for example, the scene for Chengdu airport is highly distinct from that of Beijing airport. To this end, the flight simulator puts forward higher requirements for its visual system, for which the most realistic are given a higher priority.

Currently, remote sensing images are widely applied to build the visual systems of flight simulators because of their merits of wide and accurate scenes. The development of remote sensing technology in recent years has led to a great increase in the number of satellite images. Remote sensing images have been broadly applied to various research fields, including target/object detection, temperature measurement, biophysical prediction, multi-specialist architecture, etc. However, due to hardware limitations of sensors and high costs for collecting such images, it is difficult to gain very high-resolution images. Therefore, more and more researchers are preferring to reconstruct high-resolution (HR) images from low-resolution (LR) images, rather than devoting time to physical imaging technology.

The single image super-resolution (SISR) task aims to reconstruct high-resolution images from their low-resolution counterparts. The SISR task is a significant computer vision and image processing issue that has been widely applied for all kinds of practical applications. Normally, the SISR problem can be represented by the following forward observation with a linear degradation process:

$$Y = H\tilde{X} + \eta \tag{1}$$

$Y \in \mathbb{R}^{N/s \times N/s}$ is an obtained LR image ($N/s \times N/s$ is the resolution of the LR image). $H \in \mathbb{R}^{N/s \times N/s}$ denotes a downsampling operation (typically, a bicubic interpolation) that is able to resize an HR input image $\tilde{X} \in \mathbb{R}^{N \times N}$ by a scaling factor s. In general, η is defined as an additive white Gaussian noise with a standard deviation σ. However, in real-world natural scenes, η also accounts for all possible noise during the image collection process. The noise may be the inherent sensor noise, stochastic noise, compression artifacts, etc. As is well known, the downsampling operation H is a typical ill-conditioned or singular problem, since the unknown noise (η) is usually imposed on the images. Therefore, there are many possible solutions for this task.

In this work, we attempt to utilize super-resolution technology to reconstruct the LR image into an HR one, which is further applied to build a more accurate and realistic visual system for the flight simulator. Due to the lack of a public remote sensing image dataset for the super-resolution task in this field, we first present a new dataset named Airport80, which consists of 80 ultra-high-resolution remote sensing images. This benchmark was captured from the airspace near the airports of many major cities in Asia, so it contains all kinds of natural scenes.

In succession, learning from current state-of-the-art works, we propose a simple yet powerful generative adversarial network (GAN) to achieve the remote sensing image super-resolution task. The gaming between the generative and discriminative models is expected to fit different image information caused by diverse scenes and reduces the dependencies of training samples. In general, the GAN-based SR approach is mainly to address the drawbacks of losing the high-frequency information and the fine details [5], and is able to obtain a perceptually satisfying reconstruction result.

Basically, the proposed method is based on the super-resolution generative adversarial network (SRGAN) [5] and we integrate some of the latest network design methods into the model to make it better. Since the SISR task is finally completed by the generator, our improvements mainly focus on the adjustment of the structure of the generator network. We first remove batch normalization (BN) layers from the generator. It has been confirmed that BN layers have no effect on performance in some PSNR-oriented tasks, like super-resolution. Removing BN layers helps to improve training stability and save memory usage. Second, for better ability to extract features, we replace the activation function from ReLU with PReLU [6]. Last, enlightened by [7], we also introduce deformable convolutional kernels into the generator, which can adjust the convolution sampling location by learning and focus on the extraction of local related information. Experimental results demonstrate that our approach can achieve comparable performances with state-of-the-art methods.

We summarize our primary contributions as follows.

- Due to the lack of a dataset for the super-resolution task in the research field of the visual system of a flight simulator, we present a new dataset named Airport80, which contains 80 ultra-high-resolution remote sensing images captured from the airspace near airports.

- We propose a neural network based on the GAN framework to serve as a baseline model of this dataset, in which some of the latest network designs are integrated into the model to improve the SISR performance. The proposed method is capable of generating realistic textures during a single remote sensing image super-resolution.
- Experimental results for the proposed benchmark demonstrate the effectiveness of the proposed method and show it has reached satisfactory performances. We hope that this work can bring better quality data for the visual system of a flight simulator.

2. Related Work

After decades of research, super-resolution approaches can generally be categorized into the following types: traditional methods and deep-learning-based methods. Basically, the traditional methods focus on structuring a compact dictionary or manifold space to connect patches between the low-resolution and high-resolution areas of an image. In succession, the super-resolution task can be achieved by proposing a representation scheme to conduct the super-resolution operations. A dictionary-based approach was proposed by Freeman et al. [8], in which some key dictionaries were pre-defined to present the scene pairs between the low-resolution and high-resolution patches. In this work, the nearest neighbor (NN) algorithm is applied to search the most similar patch for the input in the defined dictionary, and the corresponding high-resolution counterpart is thereby regarded as the reconstructed patch (image area). Recently, a manifold embedding technique was proposed by Change et al. [9] to replace the NN-based search strategy and showed desired performance improvements. Following this idea, the sparse coding formulation was also introduced by Yang et al. [10] to serve as an alternative solution of the NN algorithm, which further improves the performance of the super-resolution task.

Thanks to the powerful ability of the neural network to capture nonlinear transformation, deep-learning-based approaches were introduced to solve the super-resolution task and showed the performance priority over the traditional methods. A deep-learning-based model [11] was first built to achieve the image SR task in an end-to-end manner and achieved superior performance against previous works. Due to the shallow architecture, the CNN-based deep learning model [12] was designed with more convolutional layers (up to 20) to improve the final performance, in which the residual learning mechanism [13] was applied to address the gradient problems during model training. A deeper architecture (up to 52 convolutional layers), called the deep recursive residual network, was designed by Tai et al. [14] to further enhance the accuracy of the SR task. In these methods, the LR input is first upscaled to change its size to that of the HR image before feeding it into the network to complete the image reconstruction. Obviously, this design requires more computational resources (memory) and training time. To solve this issue, Shi et al. [15] proposed a sub-pixel layer, with the goal of learning a set of upsampling transformations to integrate the LR feature maps into the HR output in a more efficient way. This approach not only replaces the bicubic operation of the SR pipeline with more complex upsampling maps but also reduces the computational complexity for the overall SR operation. Recently, a deeper and wider network architecture was proposed by Lim et al. [16] to reconstruct the HR images from their LR inputs, in which the batch normalization layers [17] are removed to improve the final performance. The dense connection mechanism [18] was also adopted to complete the SR task, in which all the hierarchical features from convolutional layers are considered to generate high-resolution patches.

Like other computer vision tasks, the perception mechanism was also introduced to the SR research. The first work is the SRGAN model [5], which is able to reconstruct perceptually more pleasant high-resolution images. In order to pay more attention to the visual quality of generated images, the perceptual loss function [19] was lately introduced into GAN-based SR approaches. In those models, an adversarial loss was also proposed to formulate a combined loss function, which can produce photo-realistic high-resolution images. To further improve the performance for the GAN-based SR models, the enhanced super-resolution generative adversarial network (ESRGAN) [20] model was proposed, where the state-of-the-art perceptual SR images can be obtained up until now. More recently, a benchmark protocol was presented by Lugmayr et al. [21] to recover real-world image corruptions, in which real-world challenge series [22] are also introduced to describe the influences of the bicubic downsampling operation and separate degradation learning for super-resolution. Later, a downsample generative adversarial network (DSGAN) [23] was proposed to capture the degradation transformation by fitting the transformation distribution in an unsupervised manner, and the ESRGAN was also modified as ESRGAN-frequency separation (FS) to further improve its accuracy in a real-world setting.

3. Methodology

3.1. Airport80 Dataset

As far as we know, there are few public remote sensing image datasets for super-resolution tasks in the visual system of a flight simulator for the air transportation industry. Therefore, we have created a new dataset named Airport80, containing 80 ultra-high-resolution remote sensing images. This benchmark was captured from the airspace near the airports of many major cities in Asia, so it contains all kinds of real-world structures. We term it Airport80 to be consistent with the naming of other super-resolution datasets, like *Set*5 [24], *Set*14 [25], and *Urban*100 [26]. Due to image content and copyright issues, this dataset is meant for research purposes.

Resolution and Diversity: Each image was captured by a remote sensing satellite with a spatial resolution of 0.6 meters. Therefore, all 80 images are ultra-high-resolution, which means each of them has 4K pixels on at least one of the axes (horizontal or vertical), and some of them even have 20,000 × 20,000 resolution. In addition, this dataset includes a wealth of real-world scenes, such as urban settings, ports, deserts, hills, lakes, rivers, and so on. We randomly selected 60 images for training and used the rest for testing. Considering the ultra-high resolution issue, we cropped the remaining 20 images to 1440 × 1440 resolution with fixed step size and obtained 250 sub-images as the final testing set. Figure 1 shows some samples from our new dataset. We hope that this dataset can supplement current super-resolution tasks for remote sensing images, which are further applied to build the visual system of a flight simulator for training a qualified pilot.

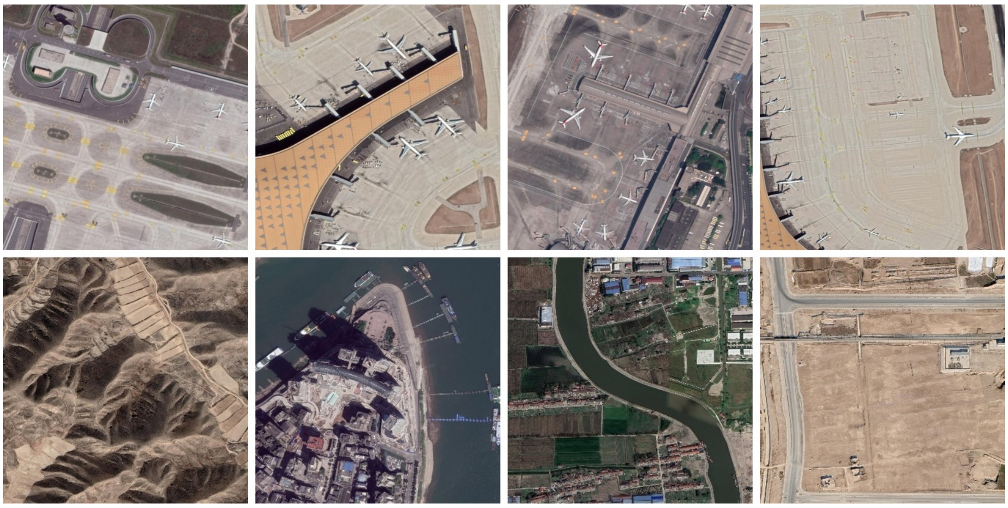


Figure 1. Selected samples from Airport80 dataset.

Evaluation Metrics: Like other super-resolution benchmarks, two commonly used metrics, peak signal-to-noise ratio (*PSNR*) and structural similarity (*SSIM*) [27] were considered to achieve a quantitative evaluation of the Airport80 dataset. *PSNR* is calculated via the mean squared error and the maximum value (denoted as L) of the images. Given the target image I and the reconstruction image $\tilde{I}$, the *PSNR* measurement can be obtained by:

$$PSNR(I,\tilde{I}) = 10 \cdot log_{10}\left(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}(I(i)-\tilde{I}(i))^2}\right) \tag{2}$$

where L equals 255 in 8-bit images. In addition, $SSIM$ is proposed for estimating the structural similarity between two images. In general, the properties of an image, including contrast, luminance, and structures, are independently evaluated to calculate a fair comparison, as shown below:

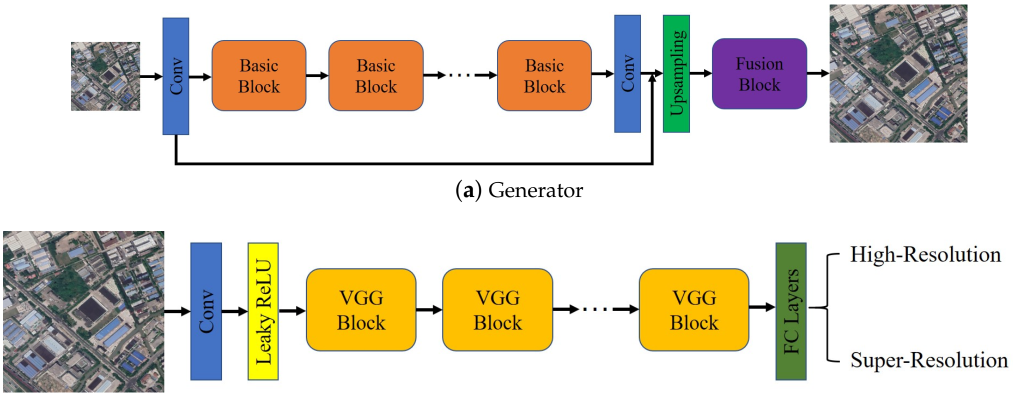
$$SSIM(I,\tilde{I}) = \frac{(2\mu_I\mu_{\tilde{I}}+c_1)(2\sigma_{I\tilde{I}}+c_2)}{(\mu_I^2+\mu_{\tilde{I}}^2+c_1)(\sigma_I^2+\sigma_{\tilde{I}}^2+c_2)} \tag{3}$$

where μ_* represents the mean of each image, σ_* represents the variance of each image, and $\sigma_{I\tilde{I}}$ represents the covariance of two images. c_1, c_2, c_3 are constants used to maintain stability.

3.2. Network Architecture

3.2.1. Baseline Model

The architecture of the proposed SR network is shown in Figure 2. Briefly, it is a typical application of the GAN [28] family in super-resolution tasks. We made some improvements to make it more suitable for the super-resolution task of remote sensing images in respect of research into the visual system of a flight simulator. It contains two individual neural networks: a generator G is designed to estimate a given LR image its HR counterpart, and a discriminator D is designed to discriminate real HR images from generated samples and ground-truth. The details of the two networks will be introduced in the following parts.



(**a**) Generator

(**b**) Discriminator

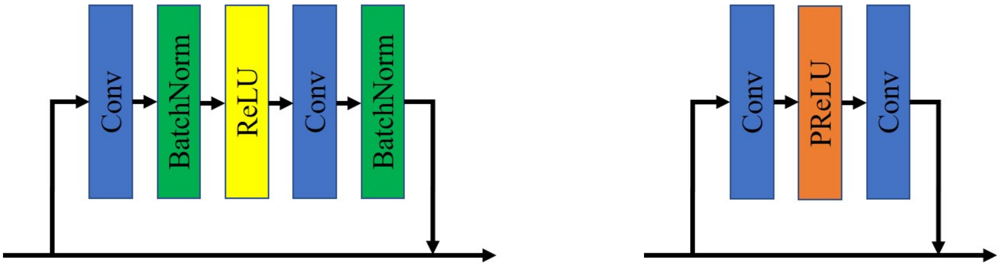
Figure 2. The architecture of the networks. Different color blocks represent different function modules.

The discriminator uses modules of the form Convolution-BatchNorm-LeakyReLU. In general, a total of 8 convolutional layers are stacked to formulate the discriminator, in which an incremental number of 3 × 3 kernels are designed, increasing by a factor of 2 from 64 to 512 like VGG [29]. The convolution operations with stride 2 are utilized to downsampling the resolution of the feature map each time, while the number of kernels will be doubled. Finally, the feature representations are converted into the probability distribution, in which two fully connected layers and a sigmoid function are applied to achieve the classification task.

3.2.2. Incremental Details

Since the SISR task is finally completed by the generator, our improvements mainly focus on the adjustment of the structure of the generator network. As depicted in Figure 2, the generator is broadly composed of three parts: (1) a series of basic blocks is responsible for extracting convolutional features for the low-resolution image, (2) a skip connection operator is designed for concatenating high-level and low-level features, and (3) a fusion block is used to fuse features and complete the final output. Compared with the original SRGAN [5], we have modified the structure of the basic blocks and fusion blocks.

In the basic blocks, we first remove batch normalization (BN) layers. BN layers have been proven to decrease performance in some PSNR-oriented tasks, like super-resolution, image deblurring, and image dehazing. Referring to ESRGAN [20], BN layers are more likely to create artifacts when the network goes deeper. These artifacts occasionally appear among iterations and different settings, violating the need for stable performance overtraining. Thus, removing BN layers will help to improve the stability of training and save memory usage, As shown in Figure 3.



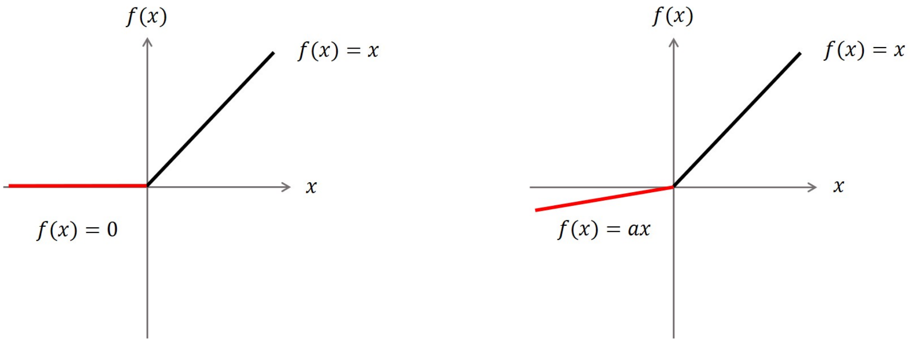
(**a**) SRGAN (**b**) Ours

Figure 3. The structures of different basic blocks. Different color blocks represent different layers.

In addition, for better ability to extract features, we changed the activation function from ReLU to parameteric rectified linear unit (PReLU) [6]. It is expressed as:

$$f(x) = \begin{cases} ax, & x \le 0 \\ x, & x > 0 \end{cases} \tag{4}$$

The parameter a is initially set to 0.25, and it will be updated automatically while training. Because there are only a few parameters added to the network, the computation and risks of over-fitting will not increase too much. The curves of two activation functions are shown in Figure 4.



(**a**) ReLU (**b**) PReLU

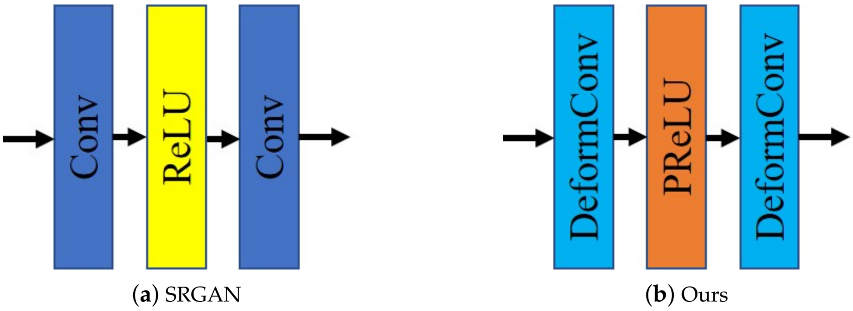
Figure 4. Curves of different activation functions. The red line represents the different parts of the two activation functions.

The inherent limitation with standard convolutional networks is that they are unable to handle geometric transformations due to their fixed shape kernel. Although some extension types like dilated convolution [30] are presented to alleviate this issue, it is still challenging for the standard kernel to align the related locations or salient features in the input image. To solve this issue, recent work [7] introduced the deformable convolutional kernel [31] into the super-resolution task to improve the capability of modeling geometric transformations by adding flexible and learnable offsets. Following this strategy, we simply replaced the standard convolutional kernel with the deformable one, as depicted in Figure 5. The standard convolution of each position p_0 in the image is expressed as

$$y(p_0) = \sum_{p_n \in R} w(p_n) x(p_0 + p_n) \tag{5}$$

where x means the feature maps or inputs, w means the sampled weights and R represents the size of the receptive field. In the deformable kernel, R is augmented with offsets $\{\Delta p_n | n = 1, ..., N\}$

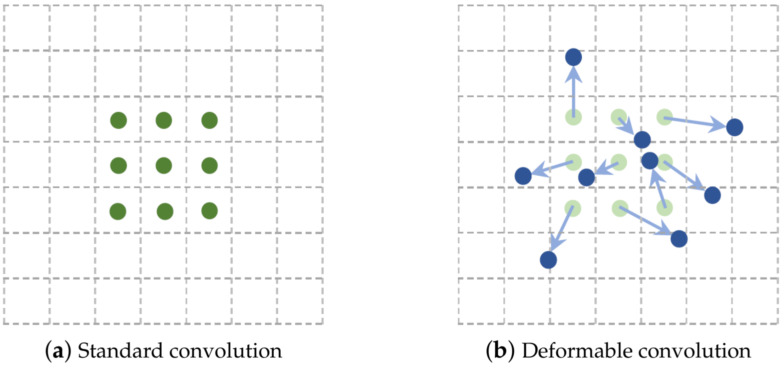
$$y(p_0) = \sum_{p_n \in R} w(p_n) x(p_0 + p_n + \Delta p_n) \tag{6}$$



(**a**) SRGAN (**b**) Ours

Figure 5. The structures of different fusion blocks. Different color blocks represent different layers.

The offsets can be learned automatically during the training phase. The standard convolution with a fixed receptive field will introduce irrelevant background noise. By introducing the deformable convolutional kernel, we hope that the network can learn convolution sampling locations autonomously and focus more on the extraction of local-related information. Figure 6 shows the sampling locations of two convolutions.



(**a**) Standard convolution (**b**) Deformable convolution

Figure 6. Illustration of the sampling locations in standard and deformable convolutions. Blue points in (**b**) represent the final sampling locations in deformable convolutions. Images come from [31].

3.3. Loss Function

The model is trained to simultaneously minimize perceptual loss L_{percep}, adversarial loss L_G^{Ra}, and context loss L_1.

Different from the pixel-wise losses, the perceptual loss [19] leverages multi-scale features extracted by a pretrained classification network to estimate high-level perceptual and semantic information differences between images. In our implementation, the loss makes use of VGG-19 [29] pretrained on ImageNet [32] as the loss network ϕ and extracts the features from the last layer of each of the first three stages. The perceptual loss is defined as

$$L_{percep} = \sum_{j=1}^{3} \frac{1}{C_j H_j W_j} ||\phi_j(J^{'}) - \phi_j(J)||_2^2 \tag{7}$$

where $\phi_j(J^{'})\phi_j(J), j = 1, 2, 3$ denote the aforementioned three VGG-19 feature maps associated with the dehazed image $J^{'}$ and the clear image J, and C_j, H_j, and W_j specify the dimension of $\phi_j(J^{'})\phi_j(J)$.

In addition, we modified the standard discriminator to the relativistic average discriminator (RaD) [33], denoted as D_{Ra}. The standard discriminator is defined as $D(x) = \sigma(C(x))$, σ means *sigmoid* function and $C(x)$ represents the non-transformed discriminator output. Thus, the RaD can be formulated as $D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}[C(x_f)])$, and $\mathbb{E}[]$ means the average of all generated samples in the mini-batch. The loss of the discriminator is then defined as:

$$L_D^{Ra} = -\mathbb{E}_{x_r}[log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[log(1 - D_{Ra}(x_f, x_r))] \tag{8}$$

The adversarial loss for the generator is in a symmetrical form:

$$L_G^{Ra} = -\mathbb{E}_{x_r}[1 - log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[log(D_{Ra}(x_f, x_r))] \tag{9}$$

where $x_f = G(x_i)$ and x_i stands for the input LR image. At last, L_1 loss is regarded as the context loss formulated by $L_1 = \mathbb{E}_{x_t}||G(x_i) - y||_1$ that evaluates the 1-norm distance between reconstructed image $G(x_i)$ and the ground-truth y. Overall, the multi-task loss L is a weighted sum of those losses:

$$L = \lambda_1 L_{percep} + \lambda_2 L_G^{Ra} + \lambda_3 L_1 \tag{10}$$

where $\lambda_1, \lambda_2, \lambda_3$ are predefined constants indicating the relative strength of each component. To keep the balance of different losses, we set them to 1.0, 5×10^{-3}, and 1×10^{-2}, respectively.

4. Experiments

4.1. Training Details

Like SRGAN [5] and ESRGAN [20], all of our experiments were performed with a scaling factor of ×4 between HR and LR images. It is worth noting that only the Airport80 dataset was used as the training data, and no images from the extra dataset were involved in the training phase. We kept all the training parameters of the unofficial SRGAN implementation provided by *MMEditing* (https://github.com/open-mmlab/mmediting/ (accessed on 25 February 2021)). We crop 128 × 128 HR sub-images and set the batch size to 16. Unlike the original SRGAN [5], we did not utilize a PSNR-oriented pretrained model to initialize the generator. The model was optimized by Adam [34] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was initially set to 1×10^{-4} and halved at [50k, 100k, 200k, 300k] iterations. All experiments were carried out on a standard PC with Intel (Santa Clara, USA) i7-6800k and two NVIDIA (Santa Clara, USA) TITAN RTX GPUs.

4.2. Ablation Study

In order to investigate the effectiveness of our improvements, we first trained some PSNR-oriented models and conducted several ablation studies. As we mentioned above, our network was built on SRGAN [5], thus the generator named SRResNet in SRGAN was selected as our baseline model. The PSNR-oriented model was only trained with the L_1 loss, and the learning rate was initially set to 2×10^{-4} and halved every 2×10^5 of iterations. The comparison results are listed in Table 1. Apparently, we can see that the adaptations of our model achieve progress on the two metrics compared to the baseline model. Compared with others, the performance improvement obtained by replacing the activation function is not very obvious. However, this adjustment is easy to implement and makes little change to the network, so we still added it to get a better performance. Finally, we integrated all of improvements and obtained a further promotion of each evaluation value, which demonstrates the proposed components are effective for the super-resolution task.

Table 1. The impact of different network designs. Each model was evaluated on the Airport80 dataset.

BN Removal	PReLU	DeformConv	PSNR	SSIM
			26.08	0.7054
√			26.75 (↑ 0.67)	0.7251 (↑ 0.0197)
	√		26.34 (↑ 0.26)	0.7156 (↑ 0.0102)
		√	26.68 (↑ 0.60)	0.7215 (↑ 0.0161)
√	√	√	27.01 (↑ 0.93)	0.7292 (↑ 0.0238)

4.3. Experimental Results

For fair comparison, we evaluated the proposed network on the Airport80 dataset for quantitative comparisons with other methods, including nearest-neighbor interpolation, bicubic interpolation, SRCNN [11], SRGAN [5], and SRResNet [5]. In addition, all the implementations came from the *MMEditing* image and video editing toolbox. It is worth noting that SRCNN and SRResNet belong to PSNR-oriented methods, while the SRGAN and our method belong to the perceptual-driven approaches. Referring to [35], the PSNR only deals with the differences between corresponding pixels instead of visual perception, which usually leads to unsatisfactory performance in representing the reconstruction quality in natural scenes, where we are usually more concerned with human perceptions. Therefore, the PSNR and SSIM in Table 2 are provided for reference.

Table 2. Quantitative comparisons for Airport80 using different methods. "Ours*" represents the generator of our method, trained with the PSNR-oriented task.

Metric	Nearest	Bicubic	SRCNN	SRGAN	SRResNet	Ours*	Ours
PSNR	23.47	25.12	25.74	23.22	26.08	**27.01**	24.59
SSIM	0.6109	0.6744	0.6896	0.6184	0.7054	**0.7292**	0.6375

It can be observed from Figure 7 that the proposed method outperforms the above mentioned approaches in both detail and sharpness. Although Bicubic and SRCNN obtain higher PSNR and SSIM, their reconstructions are generally fuzzy, and the human perception is not very good. On the contrary, SRGAN and our method, which are based on a perceptual-driven approach, achieve better edge and texture details. That also proves that PSNR and SSIM are not effective metrics for perceptual quality. Compared with SRGAN [5], our method controls the color consistency better, as shown in Figure 7, and some unpleasant color patches appear in the resulting image of SRGAN. It is worth noting that none of the above methods can handle fine textures, such as the farmland in the lower right corner of the fourth sample.

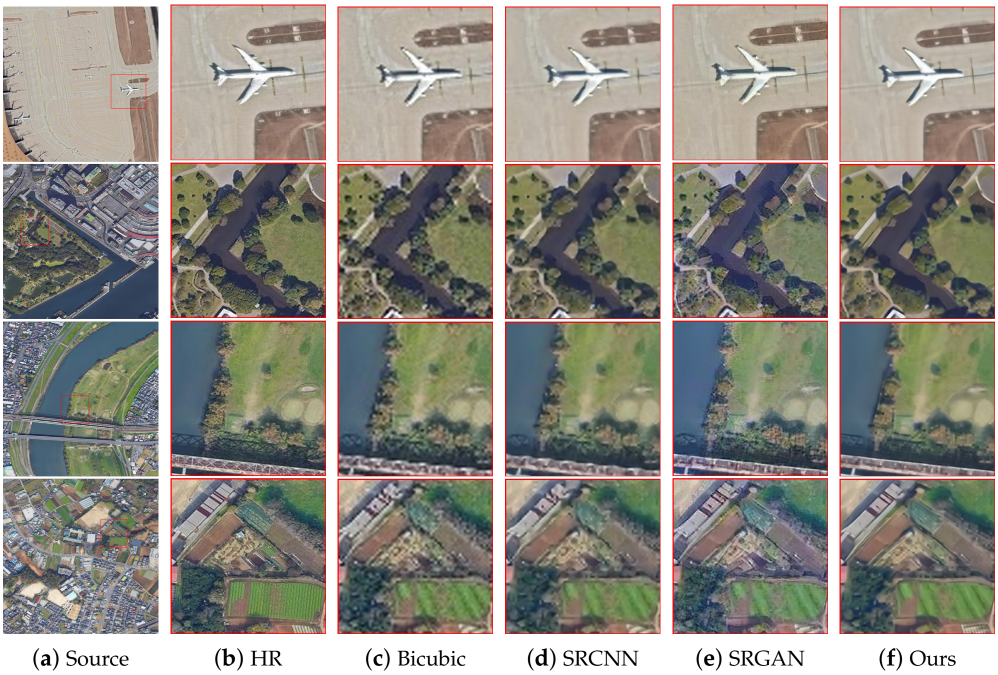

(**a**) Source (**b**) HR (**c**) Bicubic (**d**) SRCNN (**e**) SRGAN (**f**) Ours

Figure 7. Qualitative comparisons of Airport80 for different methods. In order to show the details better, the original image of all the resulting images was cropped from the test dataset.

5. Conclusions

In this paper, we started from the perspective of computer vision and utilized super-resolution technology to tackle the problem of high-resolution remote sensing image acquisition for the visual system of a flight simulator. First, due to the lack of relevant datasets in this field, we created a new dataset named Airport80, which contains 80 ultra-high-resolution remote sensing images and can be used for training and testing super-resolution algorithms. Second, a baseline model based on GAN and integrating some of the latest network designs was presented to generate realistic high-resolution images from low-resolution ones. Finally, the experimental results for our dataset demonstrate the effectiveness of the proposed method and show it has reached satisfactory performances. We hope that the above work can make a supplement to the current remote sensing image super-resolution field.

In the next step, we plan to combine some object detectors with our super-resolution network and test its application in real scenes. For example, detecting vehicles, ships and buildings in low-resolution remote sensing images

Author Contributions: Conceptualization, W.G., J.Z.; investigation, J.Z., S.T.; resources, Z.W., G.W.; writing—original draft preparation, W.G.; writing—review and editing, J.Z.; visualization, S.T., G.W.; project administration, Z.W.; funding acquisition, J.Z., Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was co-supported by Sichuan Science and Technology Program (No. 2020YFG0306, 2020YFG0327).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions eg privacy or ethical.

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

References

1. Lin, Y.; Deng, L.; Chen, Z.; Wu, X.; Zhang, J.; Yang, B. A Real-Time ATC Safety Monitoring Framework Using a Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2019**, 4572–4581. [CrossRef]
2. Lin, Y.; Guo, D.; Zhang, J.; Chen, Z.; Yang, B. A Unified Framework for Multilingual Speech Recognition in Air Traffic Control Systems. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, 1–13. [CrossRef] [PubMed]
3. Lin, Y.; Li, L.; Jing, H.; Ran, B.; Sun, D. Automated traffic incident detection with a smaller dataset based on generative adversarial networks. *Accid. Anal. Prev.* **2020**, *144*, 105628. [CrossRef]
4. Li, L.; Lin, Y.; Du, B.; Yang, F.; Ran, B. Real-time traffic incident detection based on a hybrid deep learning model. *Transportmetrica* **2020**, 1–21. [CrossRef]
5. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Aitken, P.A.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the CVPR, Honolulu, HI, USA, 21–26 July 2017.
6. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the ICCV, Santiago, Chile, 7–13 December 2015.
7. Xu, X.; Xiong, X.; Wang, J.; Li, X. Deformable Kernel Convolutional Network for Video Extreme Super-Resolution. In Proceedings of the ECCV Workshops, Glasgow, UK, 23–28 August 2020; pp. 82–98.
8. Freeman, T.W.; Jones, R.T.; Pasztor, C.E. Example-Based Super-Resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65. [CrossRef]
9. Chang, H.; Yeung, D.Y.; Xiong, Y. Super-resolution through neighbor embedding. In Proceedings of the CVPR, Washington, DC, USA, 27 June–2 July 2004.
10. Yang, J.; Wright, J.; Huang, S.T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the CVPR, Anchorage, AK, USA, 24–26 June 2008; pp. 1–8.
11. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef] [PubMed]
12. Kim, J.; Lee, K.J.; Lee, M.K. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016.
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016.
14. Tai, Y.; Yang, J.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 2790–2798.
15. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, P.A.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016.
16. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, M.K. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the CVPR Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
17. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015.
18. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the CVPR, Anchorage, AK, USA, 24–26 June 2018; pp. 2472–2481.
19. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 694–711.
20. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C.C.; Qiao, Y.; Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
21. Lugmayr, A.; Danelljan, M.; Timofte, R. Unsupervised Learning for Real-World Super-Resolution. In Proceedings of the ICCV Workshops, Seoul, Korea, 27 October–2 November 2019; pp. 3408–3416.
22. Lugmayr, A.; Joon, H.N.; Won, S.Y.; Kim, G.; Kwon, D.; Hsu, C.C.; Lin, C.H.; Huang, Y.; Sun, X.; Lu, W.; et al. AIM 2019 Challenge on Real-World Image Super-Resolution—Methods and Results. In Proceedings of the ICCV Workshops, Seoul, Korea, 27 October–2 November 2019; pp. 3575–3583.
23. Fritsche, M.; Gu, S.; Timofte, R. Frequency Separation for Real-World Super-Resolution. In Proceedings of the ICCV Workshops, Seoul, Korea, 27 October–2 November 2019; pp. 3599–3608.
24. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. In Proceedings of the BMVC, Surrey, UK, 3–7 September 2012; pp. 1–10.
25. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In *Curves and Surfaces*; Springer: Berlin/Heisenberg, Germany, 2010; pp. 711–730.
26. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
27. Wang, Z.; Bovik, C.A.; Sheikh, R.H.; Simoncelli, P.E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
28. Goodfellow, J.I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, C.A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014), Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
30. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
31. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the ICCV, Venice, Italy, 22–29 October 2017.
32. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; others. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
33. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Wang, Z.; Chen, J.; Hoi, C.H.S. Deep Learning for Image Super-resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef]