 doi:10.20944/preprints202104.0556.v1

Article

Super-Resolution of sentinel-2 images at 10m resolution without reference images

Yunhe Li [1,*], Bo Li [2]

[1] School of Electronic and Electrical Engineering, Zhaoqing University, Zhaoqing 526061, China;
* Correspondence: liyunhe@zqu.edu.cn

Abstract: Sentinel-2 can provide multi-spectral optical remote sensing images in RGBN bands with a spatial resolution of 10m, but the spatial details provided are not enough for many applications. WorldView can provide HR multi-spectral images less than 2m, but it is a commercial paid resource with relatively high usage costs. In this paper, without any available reference images, Sentinel-2 images at 10m resolution are improved to a resolution of 2.5m through super-resolution (SR) based on deep learning technology. Our model, named DKN-SR-GAN, uses degradation kernel estimation and noise injection to construct a dataset of near-natural low-high-resolution (LHR) image pairs, with only low-resolution (LR) images and no high-resolution (HR) prior information. DKN-SR-GAN uses the Generative Adversarial Networks (GAN) combined of ESRGAN-type generator, PatchGAN-type discriminator and the VGG-19-type feature extractor, using perceptual loss to optimize the network, so as to obtain SR images with clearer details and better perceptual effects. Experiments demonstrate that in the quantitative comparison of the non-reference image quality assessment (NR-IQA) metrics like NIQE, BRISQUE and PIQE, as well as the intuitive visual effects of the generated images, compared with state-of-the-art models such as EDSR8-RGB, RCAN and RS-ESRGAN, our proposed model has obvious advantages.

Keywords: super-resolution; generative adversarial network; Sentinel-2

1. Introduction

Satellite remote sensing images have important applications in many fields, such as agriculture, environmental protection, land use, urban planning, natural disasters, hydrology, climate and so on [1]. With the continuous updating of optical instruments and other equipment, the spatial resolution of satellite images is constantly improving. For example, Worldview-3/4 satellite can collect 8 bands of multi-spectral data with a ground resolution of 1.2m [2]. However, Worldview-3/4 satellite data need to be paid for its use, and when covering a large area or performing a multi-temporal analysis, it will be restricted by the data cost. Therefore, open access data with acceptable spatial quality can be considered, such as Landsat [3] or Sentinel [4]. Sentinel-2 updates remote sensing images of every location in the world for free approximately every 5 days, and these remote sensing images are becoming more and more important resources for applications. Sentinel-2 uses two satellites to achieve remote sensing coverage at the equator on a global scale and provides a multi-resolution layer composed of 13 spectral bands, among which 10m resolution images are provided in 4 bands of the visible lights in red (B4), green (B3) and blue (B2) and the near-infrared (B8), 20m resolution images provided in 6 bands, and 60m resolution images provided in the other 3 bands respectively [4]. The bands of 10m and 20m resolution are usually used for land cover or water mapping, agriculture or forestry, while the band of 60m with lower resolution is mainly used for water vapor monitoring [5]. Due to the open data distribution strategy, the 10m resolution remote sensing images provided by Sentinel-2 are becoming important resources for some applications. However, such spatial resolution is still slightly insufficient in many applications. In order to make full use of the free availability of Sentinel-2 images, and to achieve the spatial

resolution of about 2m, it is worth considering some post-processing methods to obtain the spatial enhancement of LR images, and recover the high-frequency details to generate HR images. In order to improve the spatial resolution of Sentinel-2 images, some researchers [6-12] fused the data of the several bands of Sentinel-2 with 60m, 20m and 10m spatial resolution to obtain higher spatial resolution images, however this paper focuses on SR directly using 10m resolution images.

At the earlier period, Yang et al. [13] and Gou et al. [14] studied the supervised SR model based on dictionary learning, and provided an effective solution by using sparse coding technology. Pan et al. [15] applied structural self-similarity and compressed sensing to SR tasks. Zhang et al. [16] and Li et al. [17] adopted several different image representation spaces in SR to achieve higher performance.

Deep learning has attracted more and more attention in the field of SR [18]. Deep learning does not need to directly map the relationship between HR and LR domains. As long as there is enough training data, a deep learning network in principle can learn very complex non-linear relationships [19]. Among them, the model based on convolutional neural network (CNN) can make better use of the high-order features of the images to conduct HR images, and can significantly improve the performance of SR [18]. Dong et al. [19] proposed SRCNN network with high learning ability based on CNN, and adopted pixel loss to optimize the network, but the result was too smooth without consideration of the perceptual quality. And on this basis, Kim et al. [20] and Zhang et al. [21] introduced residual learning models, Tai et al. [22] introduced recursive learning models, and Hu et al. [23] introduced attention mechanism to optimize the deep learning architecture to improve performance, but these models also had the problem of over-smoothing because they all solely relied on pixel loss to optimize the network.

Goodfellow et al. [24] proposed GAN training two models at the same time, one of which was called generator (G), and the other was called discriminator (D). The generator took the random samples in the potential space as the input, and its output needed to imitate the real samples in the training set as much as possible. The input of the discriminator was the real sample or the output of the generator, and its goal was to distinguish the output of the generator from the real sample as far as possible, while the goal of the generator was to cheat the discriminator as far as possible. SRGAN [25] proposed by Leding et al. was a pioneering work to implement SR based on GAN theory, and because of its ability to generate images with rich texture and high quality, GAN has been widely used in SR. Wang et al. [26] further improved SRGAN model, proposed ESRGAN, used a more complex and denser residual layer combination in the generator and deleted batch normalization layer. As SR model based on GAN was gradually applied in the field of satellite remote sensing, Ma et al. [27] proposed transfer GAN (TGAN) to solve the shortcomings of poor quantity and quality of remote sensing data. Haut et al. [28,29] and Lei et al. [30] designed the network to form LHR image pairs by downsampling the public remote sensing images, and tested different network architectures. Aiming at the remote sensing image provided by Sentinel-2, Gong et al. [31] proposed Enlighten-GAN SR model, which adopted internal inconsistency loss and cropping strategy and achieved good results in gradient similarity measurement (GSM) for the medium resolution remote sensing images of Sentinel-2. Sentinel-2 can provide the images with spatial resolution up to 10m. In the task of upgrading the resolution from 10m to 2m, SR model based on GAN has encountered a great challenge that mainly comes from the lack of HR real images at 2m resolution. In recent years, some researchers used the 10m resolution images of Sentinel-2 and 2m HR images of worldview satellite to form LHR image pairs to construct training dataset. For example, Galar et al. [32] proposed a SR model based on enhanced depth residual network (EDSR), and Salgueiro et al. [33] proposed an RS-ESRGAN model based on ESRGAN. All the proposed models could enhance the 10m channel of Sentinel-2 to 2m. However, by using the unnatural low-high image pairs consisting of Sentinel-2 and Worldview images, and other models using the BiCubic downsampling to construct LHR image pairs [21,25,26,34-37], the track details related to frequency will be lost [38]. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

order to solve this problem, inspired by blind SR model KernelGAN [39] and the blind image denoising model [40], we explicitly estimate the degradation kernel of LHR image pairs of natural images through GAN, estimate the distribution of the degraded noises at the same time and degrade the 10m resolution images of Sentinel-2 to construct near-natural LHR image datasets. On the basis of these datasets, with the references to SRGAN, PatchGAN and VGG-128 network structure, DKN-SR-GAN is designed to implement SR of Sentinel-2 images from 10m to 2.5m.

2. Dataset

For the convenience of the following analysis, we initially present the datasets used in training and testing. The model proposed in this paper is aimed at Sentinel-2 images, so we use SEN12MS [41] dataset to train and test the models. SEN12MS contains complete multi-spectral information in geocoded images, it also includes SAR and multi-spectral images provided by Sentinel-1 and Sentinel-2, and adds land cover information obtained by MODIS system. This paper mainly focuses on 10m resolution images of red (B4), green (B3) and blue (B2) bands in multi-spectral images, namely, RGB color images with 10m resolution. SEN12MS gives Sentinel-2 cloudless images of the region of interest (ROI) at specified time intervals. SEN12MS divides the images into patches with 256x256 pixels, which span 128 pixels so that the overlap rate between the adjacent patches is 50%. SEN12MS takes 50% overlap as an ideal compromise between the independence of patches and the maximum number of samples. SEN12MS dataset obtains randomly sampled ROI based on four seeds(1158, 1868, 1970 and 2017), and the distribution of ROI is shown in Figure 1.

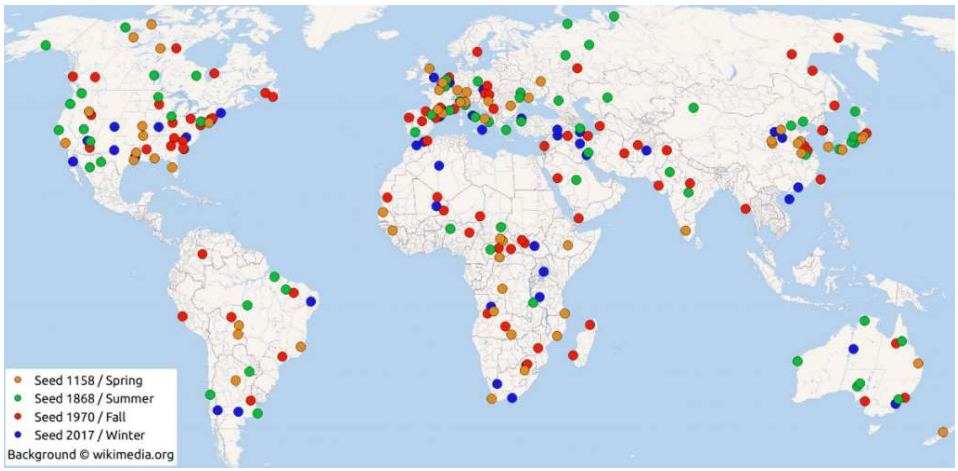

Figure 1. Distribution of regions of interest corresponding to four random seeds

In this paper, DKN-SR-GAN uses a dataset of SEN12MS, named ROIs1158, which is composed of 56 regions of interest across globe generated from 1158 seeds from June 1, 2017 to August 31, 2017. ROIs1158 is divided into 56 subsets by region, totally 40883 pieces of 256x256 pixel images. This paper randomly selects the subset "ROIs1158_spring_106" as the test dataset (ROI_Te), which contains 784 test images; while for the remaining 55 subsets, including 40099 images, are used as the source images dataset (ROI_Src), and ROI_Src is degraded to generate LR image dataset (ROI_LR). The source images $\mathbf{I}_{\mathrm{src}}$ in ROI_Src are directly used as HR images $\mathbf{I}_{\mathrm{HR}}$ in the training, which forms LHR image pairs dataset (ROI_Tr) with the images $\mathbf{I}_{\mathrm{LR}}$ in ROI_LR dataset one by one. This paper compares the performance of the newly proposed models including EDSR8-RGB [32], RCAN [21], and RS-ESRGAN [33], as well as the traditional model of BiCubic [42]. Bicubic directly uses ROI_Te for interpolation test without training; RCAN takes the images in ROI_Src as LR images, and generates HR images by BiCubic-interpolating every image to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

form a LHR image pair dataset; the models of EDSR8-RGB and RS-ESRGAN respectively refer to the models proposed in [32] and [33] to construct a dataset based on ROI_Src.

3. Methods

3.1. Structure of DKN-SR-GAN

This paper use DKN-SR-GAN to generate 2.5m resolution images $\mathbf{I}_{\mathrm{SR}}$ from 10m resolution source images $\mathbf{I}_{\mathrm{src}}$ of Sentinel-2 in two stages. In the first stage, KernelGAN is used to implement the estimation of the explicit degradation kernel of $\mathbf{I}_{\mathrm{src}}$ images, and then combined with injecting the degraded noise, the source images $\mathbf{I}_{\mathrm{src}}$ are degraded to LR images $\mathbf{I}_{\mathrm{LR}}$, which will combine with HR image $\mathbf{I}_{\mathrm{HR}}$ (equivalent to $\mathbf{I}_{\mathrm{src}}$) to construct LHR image pairs $(\mathbf{I}_{\mathrm{LR}}, \mathbf{I}_{\mathrm{HR}})$. In the second stage, the dataset $\{(\mathbf{I}_{\mathrm{LR}}, \mathbf{I}_{\mathrm{HR}})\}$ is used to train the super-resolution generative adversarial network (SR-GAN), which consists of a super resolution generator (SR-G), a super resolution discriminator (SR-D), and a super resolution perceptual feature extractor (SR-F). DKN-SR-GAN represents Sentinel-2 image SR model proposed in this paper, and the structure of DKN-SR-GAN is shown in Figure 2.

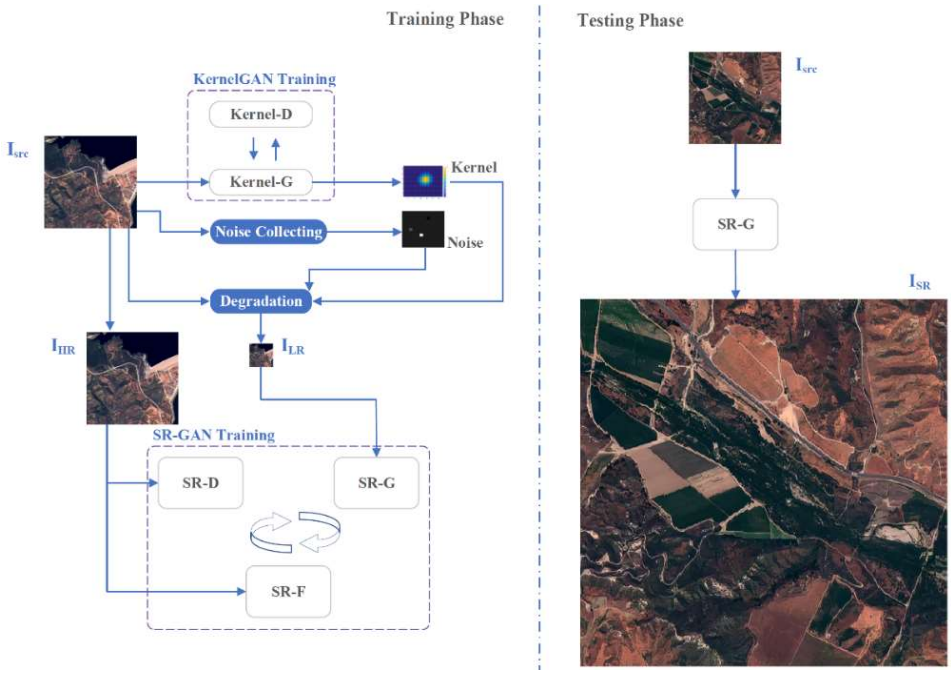

Figure 2. Structure of DKN-SR-GAN on Sentinel-2 remote sensing images

3.2. Degradation Kernel Estimation and Noise Injection

Here we introduce an image degradation model based on kernel estimation and noise injection. The natural pairing relationship between low and high resolution images can be approximately understood as the degradation relationship between HR images and LR images, and the degradation process can be expressed as:

$$\mathbf{I}_{LR} = (\mathbf{I}_{SRC} * \mathbf{k}^{s}) \downarrow_{s} + \mathbf{n} \quad (1)$$

Where, $\mathbf{k}^{s}$ and $\mathbf{n}$ represents degradation kernel and degraded noise respectively, and s represents scaling factor. The quality of degradation kernel and degraded noise determines the relevance between LHR image pairs and natural image pairs, as well as accuracy of the extracted mapping features between low and high resolution images, which further determines the quality of images generated by SR.

3.2.1. Degradation Kernel Estimation Based on KernelGAN

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

Here we first consider the noise-free degradation process, assuming that the noise-free LR image $\mathbf{I}_{LR\_cl}$ is the result of downsampled HR image $\mathbf{I}_{SRC}$ by using the degradation kernel through the scaling factor s:

$$\mathbf{I}_{LR\_cl} = (\mathbf{I}_{SRC} * \mathbf{k}^{s}) \downarrow_{S} \quad (2)$$

In this paper, KernelGAN is used to estimate the image degradation kernel $\mathbf{k}^{s}$, which is a blind SR degradation kernel estimation model based on Internal-GAN [43], and a completely unsupervised GAN requiring no extra training data except the image $\mathbf{I}_{SRC}$ itself [39]. KernelGAN uses only the images $\mathbf{I}_{SRC}$ for training to learn the distribution of internal pixel patches, with the goal to find the image-specific degradation kernel and to search for the best degradation kernel to retain the distribution of pixel patches on each scale of the image $\mathbf{I}_{SRC}$. More specifically, our goal is to "generate" downsampled images and to make the pixel patch distribution of the downsampled images as close to the images $\mathbf{I}_{SRC}$ as possible. The essence of the model is to extract the cross-scale recursive characteristics between LR and HR images through deep learning, and GAN in KernelGAN can be understood as the matching tool for pixel patch distribution. The implementation process of KernelGAN is shown in Figure 3, by training on a single input image to learn the distribution of internal pixel patches of the cropped patch. KernelGAN consists of a kernel generator (Kernel-G) and a kernel discriminator (Kernel-D). Both the kernel-G and the kernel-D are fully convolutional, which means that the network is applied to the pixel patch rather than the whole image. With the given input of the images $\mathbf{I}_{SRC}$, the kernel generator will learn to downsample then to $\mathbf{I}_{LR\_cl}$, whose goal is to make the discriminator indistinguishable from the input images $\mathbf{I}_{SRC}$ at the pixel patch level.

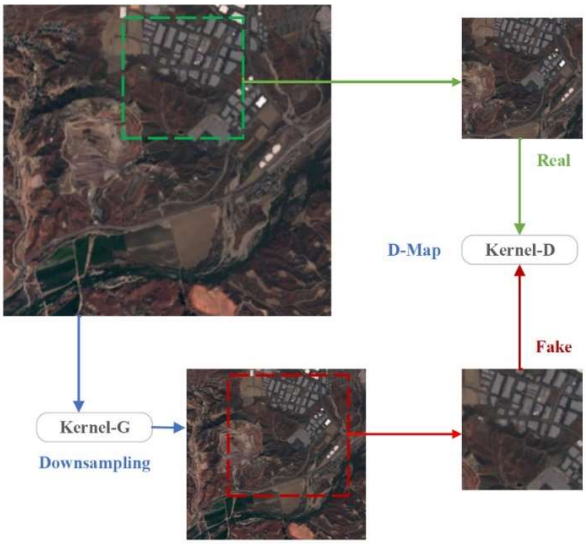

Figure 3. Structure of KernelGAN

The objective function of KernelGAN is defined as:

$$G^{*}(\mathbf{I}_{SRC}) = \arg\min_{G} \max_{D} \{\mathbb{E}_{x\sim \text{patches}(\mathbf{I}_{SRC})}[|D(\mathbf{I}_{SRC}) - 1| + |D(G(\mathbf{I}_{SRC}))|] + \mathcal{R}\} \quad (3)$$

Where, G represents generator, D represents discriminator. And $\mathcal{R}$ is the regularization term optimized by degradation kernel $\mathbf{k}^{s}$:

$$\mathcal{R} = \alpha_{s\_1}\mathcal{L}_{s\_1} + \alpha_{b}\mathcal{L}_{b} + \alpha_{sp}\mathcal{L}_{sp} + \alpha_{c}\mathcal{L}_{c} \quad (4)$$

Where, $\mathcal{L}_{s\_1}$、$\mathcal{L}_{b}$、$\mathcal{L}_{sp}$、$\mathcal{L}_{c}$ represent losses, and α_{s_1}、α_{b}、α_{sp}、α_{c} represent constant coefficients. In this paper, the constant coefficients are set according to experience as $\alpha_{s\_1} = 0.5, \alpha_{b} = 0.5, \alpha_{sp} = 5, \alpha_{c} = 1$. The losses are defined as following equations respectively:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

$$\mathcal{L}_{s\_1} = \left| 1 - \sum_{i,j} k_{i,j} \right| \tag{5}$$

Where, $k_{i,j}$ represents the parameter value of each point of the degradation kernel, and the goal of $\mathcal{L}_{s\_1}$ is that the sum of $\{k_{i,j}\}$ is 1.

$$\mathcal{L}_{\mathrm{b}} = \sum_{i,j} \left| k_{i,j} \cdot m_{i,j} \right| \tag{6}$$

The goal of $\mathcal{L}_{\mathrm{b}}$ is to punish the non-zero value near the boundary, and $m_{i,j}$ is the constant mask of weight, which increases exponentially with the distance from the center of $\{k_{i,j}\}$.

$$\mathcal{L}_{\mathrm{sp}} = \sum_{i,j} \left| k_{i,j} \right|^{1/2} \tag{7}$$

The goal of $\mathcal{L}_{\mathrm{sp}}$ is the sparsity of $k_{i,j}$ to avoid excess smoothness of interior kernel.

$$\mathcal{L}_{\mathrm{c}} = \left\| (x_0, y_0) - \frac{\sum_{i,j} k_{i,j} \cdot (i,j)}{\sum_{i,j} k_{i,j}} \right\|_2 \tag{8}$$

The goal of $\mathcal{L}_{\mathrm{c}}$ is to make the center of $\{k_{i,j}\}$ in the center of the interior kernel. (x_0, y_0) represents the indices of the center.

Kernel-G can be regarded as an image downsampling model, which implements linear downsampling mainly through convolution layer, and the network contains no nonlinear activation unit. Nonlinear generator is not used here because it is possible for the nonlinear generator to generate physically unnecessary solutions for the optimization targets, for example, to generate an image that is not downsampled but contains effective pixel patches. In addition, because the single-layer convolution layer cannot converge accurately, we use the multi-layer structure of linear convolution layers as shown in Figure 4.

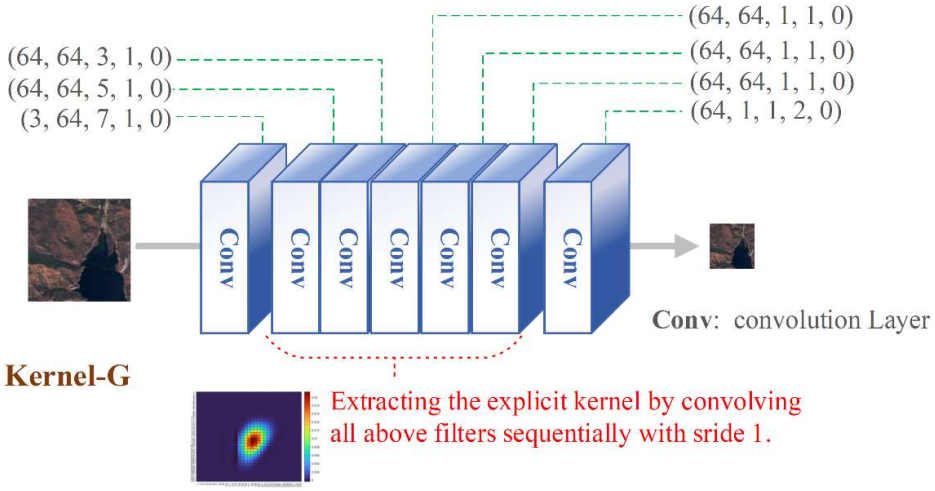

Figure 4. Network structure of Kernel generator consisting of multi-layer linear convolution layer

The goal of Kernel-D is to learn the distribution of pixel patches in the input images $\mathbf{I}_{SRC}$ and distinguish between the real patches and fake patches in the distribution. The real patches are cropped from the input images $\mathbf{I}_{SRC}$, while the fake patches are cropped from $\mathbf{I}_{LR\_cl}$ generated by the kernel-G. We use the full convolution pixel patch discriminator introduced in [44] to learn the pixel patch distribution of every single image as shown in Fig. 5.

The convolution layer used in the kernel-D does not perform pooling operations, so as to act on each pixel patch implicitly, and finally generate a hot map (D-map), of which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

each position corresponds to one cropped patch input. The hot map output by the kernel-D represents the possibility of each pixel extracting the surrounding pixel patches from the original pixel patch distribution, and used to distinguish the real patches from the fake patches. The loss is defined as the pixel-wise mean square error between the hot map and the label map. Label map refers to all 1 labels of the real patches and all 0 labels of the fake patches.

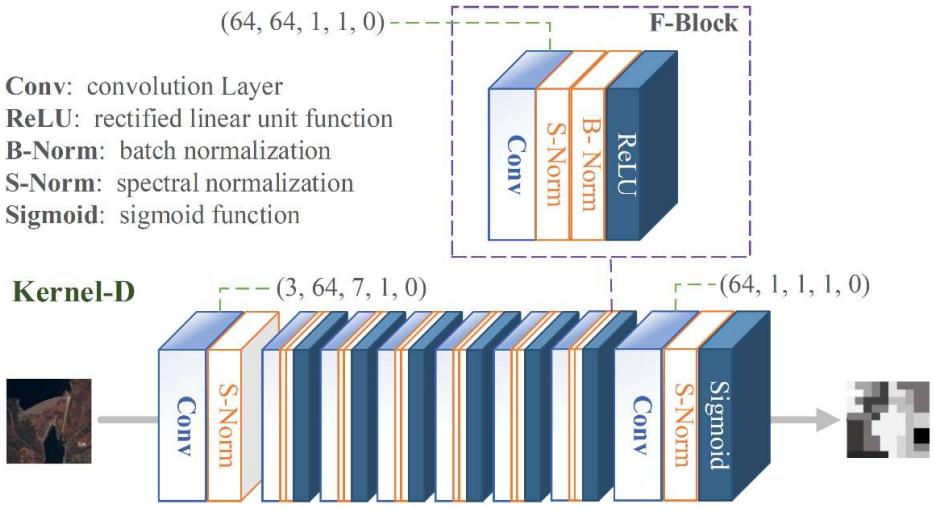

Figure 5. Discriminator network structure consisting of multi-layer non-pooled convolution layer

After the training of KernelGAN, we do not focus on the generator network, but convolute the convolution layers of the kernel-G with the stride of 1 successively to extract the explicit degradation kernel. Meanwhile, the training of KernelGAN is based on one single input image $\mathbf{I}_{SRC}$, which means that each input image trains one degradation kernel, and many degradation kernels generated by the training image set will be randomly selected and used in the subsequent steps. The graphical examples of some degradation kernels are shown in Figure 6.

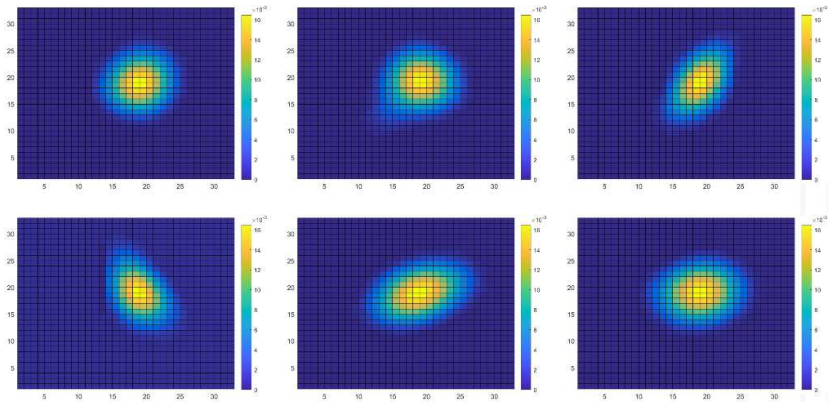

Figure 6. Graphical example of degradation kernel extracted after KernelGAN training

3.2.2. Generation and Injection of Noise

We explicitly inject noise into the downsampled images $\mathbf{I}_{LR\_cl}$ to generate realistic LR images $\mathbf{I}_{LR}$. In the process of image downsampling, the high-frequency information will be lost, so the distribution of noise will change at the same time. In order to ensure that the degraded images $\mathbf{I}_{LR}$ have a similar noise distribution as the source images $\mathbf{I}_{SRC}$,

we extract the noise mapping patches directly from the source images $\mathbf{I}_{SRC}$ in the training dataset. Due to the large variance of the patches with rick contents [38], and inspired by [40,45], when extracting noise mapping patches we control the variance within a specific range under the condition:

$$D(\mathbf{n}_i) < \sigma_{max} \tag{9}$$

Where, $D(\cdot)$ represents the variance function, and σ_{max} represents the maximum value of the variance. The noise mapping patches are extracted from images selected from the images of ROI_Src randomly, and a certain number of noise patches are extracted to construct the dataset (ROI_Noi). The noise mapping patches used for noise injection process are randomly selected from ROI_Noi.

To sum up, the process of generating LR images in ROI_LR from the source images in ROI_Src can be expressed as Equation (10), where i and j are randomly selected:

$$\mathbf{I}_{LR} = (\mathbf{I}_{SRC} * \mathbf{k}_i^s) \downarrow_s + \mathbf{n}_j \tag{10}$$

3.3. SR-GAN

SR-GAN consists of super resolution generator (SR-G), super resolution discriminator (SR-D) and perceptual feature extractor (SR-F). SR-G is designed on the basis of ESRGAN [26] model. Because ESRGAN discriminator may introduce more artifacts [38], SR-D is designed on the basis of PatchGAN [44] model. The perceptual feature extractor is designed on the basis of VGG-19 [46], so as to introduce the perceptual loss [47] to enhance the visual effect of low-frequency features of the images.

The loss $\mathcal{L}_{\mathrm{SR}}$ of SR-GAN consists of three parts, including pixel-wise loss $\mathcal{L}_{\mathrm{x}}$ [26], perceptual loss $\mathcal{L}_{\mathrm{p}}$ and adversarial loss $\mathcal{L}_a$.

$$\mathcal{L}_{\mathrm{SR}} = \alpha_x \mathcal{L}_{\mathrm{x}} + \alpha_{\mathrm{p}} \mathcal{L}_{\mathrm{p}} + \alpha_{\mathrm{a}} \mathcal{L}_a \tag{11}$$

Where, $\alpha_x, \alpha_{\mathrm{p}}$ and α_{a} are constant coefficients, and the constant coefficients are set according to experience as $\alpha_x = 0.01$、$\alpha_{\mathrm{p}} = 1$、$\alpha_{\mathrm{a}} = 0.005$. The losses $\mathcal{L}_{\mathrm{x}}, \mathcal{L}_{\mathrm{p}}$ and $\mathcal{L}_a$ are defined as equation (12), (13) and (16).

$$\mathcal{L}_{\mathrm{x}} = \mathbb{E}_{\mathbf{I}_{LR}} \| G(\mathbf{I}_{LR}) - \mathbf{I}_{HR} \|_1 \tag{12}$$

Pixel-wise loss $\mathcal{L}_{\mathrm{x}}$ uses L1 distance to evaluate the pixel-wise content loss between the generated images $G(\mathbf{I}_{LR})$ and the real images $\mathbf{I}_{HR}$.

$$\mathcal{L}_{\mathrm{p}} = \lambda_f \mathcal{L}_f + \lambda_t \mathcal{L}_t \tag{13}$$

Perceptual loss $\mathcal{L}_{\mathrm{p}}$ evaluates the perceived differences in content and style among different images, and consists of feature reconstructing loss $\mathcal{L}_f$ related to content and style reconstructing loss $\mathcal{L}_t$, where λ_f and λ_t denotes constant coefficients, and $\mathcal{L}_f$ and $\mathcal{L}_t$ can be expressed as:

$$\mathcal{L}_f = \frac{1}{C_j H_j W_j} \| \phi_j(G(\mathbf{I}_{LR})) - \phi_j(\mathbf{I}_{HR}) \|_2^2 \tag{14}$$

$$\mathcal{L}_t = \left\| \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} [\phi_j(G(\mathbf{I}_{LR}))_{h,w,c} \phi_j(G(\mathbf{I}_{LR}))_{h,w,c'} - \phi_j(\mathbf{I}_{HR})_{h,w,c} \phi_j(\mathbf{I}_{HR})_{h,w,c'}] \right\|_F^2 \tag{15}$$

Where $\phi_j(I)$ represents the characteristic diagram obtained at level j of the convolution layer after the image $\mathbf{I}$ inputs SR-F, and the shape of the obtained characteristic diagram is $C_j \times H_j \times W_j$ (Channel × Height × Width) and $\|\cdot\|_F^2$ represents square Frobenius norm.

$$\mathcal{L}_a = \sum_{n=1}^{N} -D\big(G(\mathbf{I}_{LR})\big) \tag{16}$$

Adversarial loss $\mathcal{L}_a$ is used to enhance the texture details of the generated image to make it look more realistic.

The structure of SR-G is shown in Fig. 7. Based on ESRGAN model, and adopting RRDB [39] structure, it is trained in the constructed LHR image pairs $(\mathbf{I}_{LR}, \mathbf{I}_{HR})$ and the resolution of the generated images will be magnified x4.

Due to the discriminator in ESRGAN model may introduce more artifacts, this paper uses the patch discriminator instead of VGG-128 discriminator in ESRGAN model, and SR-D is designed based on PatchGAN [44] model. In addition, the patch discriminator is used instead of VGG-128 discriminator out of consideration for the following aspects:

 doi:10.20944/preprints202104.0556.v1

VGG-128 limits the size of the generated image to 128, which makes it inconvenient to conduct the multi-scale training; VGG-128 uses a fixed fully-connected layer, which makes the discriminator pay more attention to the global features and ignore the local features [34]. We use the patch discriminator with a fully convolutional structure and a fixed receptive field. Each output value of SR-D is only related to the patches in the local fixed region, so that we can optimize the local details. The average value of all local errors are used as the final error to guarantee the global consistency. The structure of SR-D is shown in Figure 8.

Based on VGG-19 [46] model, this paper introduces the perceptual feature extractor to extract the perceptual loss $\mathcal{L}_p$, that is, to extract the inactive features in VGG-19. The perceptual loss can enhance the low-frequency features of the images and make the images generated by the generator look more realistic. The structure of the perceptual feature extractor is shown in Figure 9.

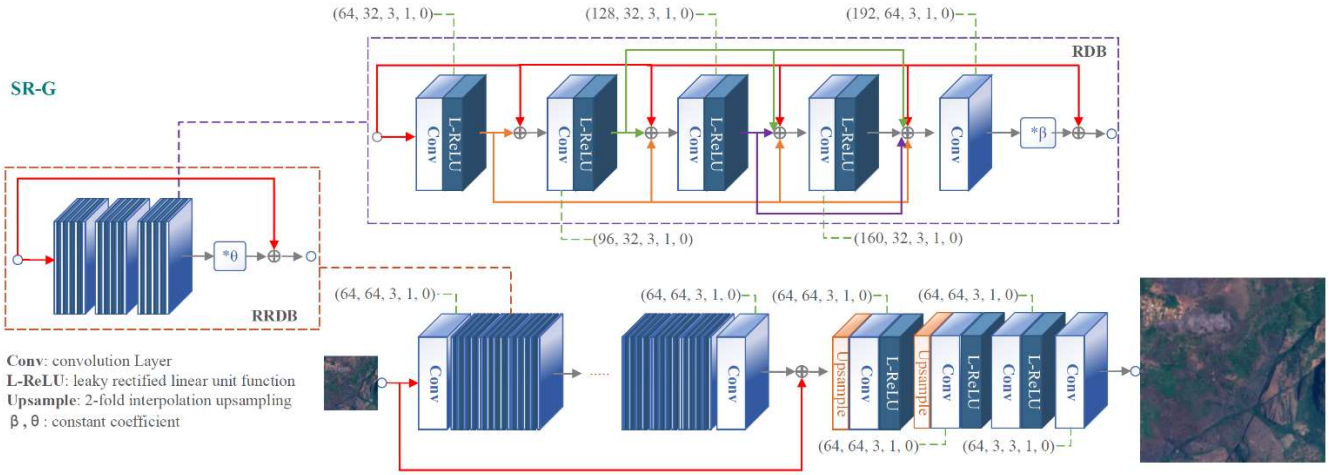

Figure 7. Structure of super-resolution generator(SR-G)

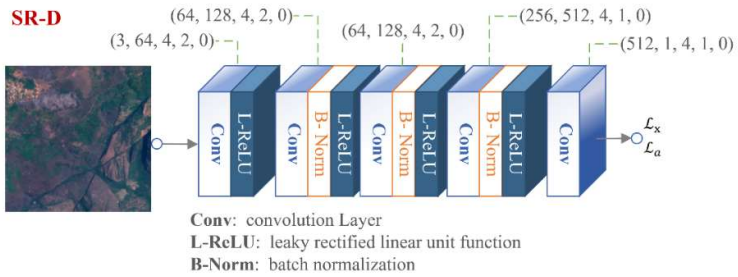

Figure 8. Structure of super-resolution discriminator (SR-D)

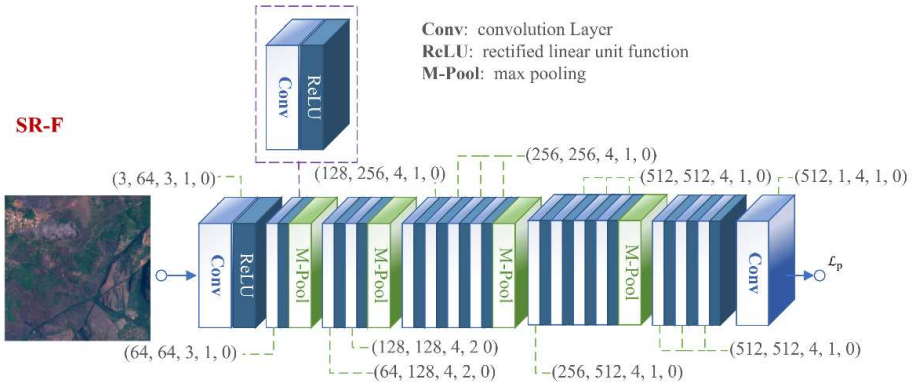

Figure 9. Structure of perceptual feature extractor(SR-F)

4. Experiments and results

4.1. Training Details

The proposed model DKN-SR-GAN and other compared models like EDSR8-RGB, RCAN, and RS-ESRGAN are run under Pytorch environment, using the modules provided by "sefibk/KernelGAN" project [39], "xinntao/BasicSR" project [48] and "Tencent/Real-SR" project [38] in the Github library. BiCubic can be obtained directly using Matlab functions to perform interpolation operations.

DKN-SR-GAN first generates a LHR image pair dataset (ROI_Tr) based on a training dataset (ROI_Src) for training and testing. We randomly select 2134 images from 40099 images of ROI_Src to generate a degraded kernel dataset (ROI_Ker) through KernelGAN training one by one, namely $\mathbf{k}_i^s \in \{\text{ROI\_Ker}\}, i \in \{1,2\cdots 2134\}$; and then randomly select 4972 images from 40099 images of ROI_Src to extract noise patches one by one to form a noise patch dataset (ROI_Noi), namely $\boldsymbol{n_j} \in \{\text{ROI\_Noi}\}, j \in \{1,2\cdots 4972\}$; finally, we use the degradation kernel and injected noise to perform degrading operations on the images in ROI_Src one by one. In the processing of each image, the degradation kernel and injected noise are randomly selected from ROI_Ker and ROI_Noi.

The network structural parameters of the kernel-G and the kernel-D and the constant coefficients of losses of KernelGAN have been mentioned above, so we will not repeat them here. In the training phase, both the generator and the discriminator adopt ADAM optimizer with the parameters $\beta_1 = 0.5$、$\beta_2 = 0.999$, the learning rates of the generator and the discriminator are both set to 0.0002, decrementing by x0.1 every 750 iterations, and the network is iteratively trained 3000 epochs.

SR-G uses "RRDBNet" model in "BasicSR" project, and SR-D uses "NlayerDiscriminator" model in "Real-SR" project. The network structural parameters and the constant coefficients of losses have been mentioned above, therefore we will not repeat them here. The image is magnified by 4 times, and during the training phase, both the generator and the discriminator adopt ADAM optimizer with the parameters $\beta_1 = 0.9, \beta_2 = 0.999$, the learning rates of the generator and the discriminator are both set to 0.0001, and the network is iteratively trained 60,000 epochs.

Many convolutional layers are used in KDN-SR-GAN, and these convolutional layers play a vital role. After many tests, it is known that the parameters of the convolutional layer in the network need to be set to the values shown in Table 1, to achieve the x4 resolution images by KDN-SR-GAN and obtain the image quality we want.

EDSR8-RGB, RCAN, and RS-ESRGAN models implement training and testing under the framework of BasicSR [48], and adopt the parameter setting schemes which have been proven to achieve better results in references [21,32,33], and the parameters used in the implementation are detailed in Table 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

Table 1. Setting of specific parameters for convolutional layers of KDN-SR-GAN

	in_ channels	out_ channels	kernel_ size	stride	padding	dilation	groups	bias	padding_ mode
Kernel-G	3	64	7	1	0	1	1	True	'zeros'
	64	64	5	1	0	1	1	True	'zeros'
	64	64	3	1	0	1	1	True	'zeros'
	64	64	1	1	0	1	1	True	'zeros'
	64	64	1	1	0	1	1	True	'zeros'
	64	64	1	1	0	1	1	True	'zeros'
	64	1	1	2	0	1	1	True	'zeros'
Kernel-D	3	64	7	1	0	1	1	True	'zeros'
	64	64	1	1	0	1	1	True	'zeros'
	64	1	1	1	0	1	1	True	'zeros'
SR-G	64	32	3	1	0	1	1	True	'zeros'
	96	32	3	1	0	1	1	True	'zeros'
	128	32	3	1	0	1	1	True	'zeros'
	160	32	3	1	0	1	1	True	'zeros'
	192	64	3	1	0	1	1	True	'zeros'
	64	64	3	1	0	1	1	True	'zeros'
	64	3	3	1	0	1	1	True	'zeros'
SR-D	3	64	4	2	0	1	1	True	'zeros'
	64	128	4	2	0	1	1	True	'zeros'
	128	256	4	2	0	1	1	True	'zeros'
	256	512	4	1	0	1	1	True	'zeros'
	512	1	4	1	0	1	1	True	'zeros'
SR-F	3	64	3	1	0	1	1	True	'zeros'
	64	64	3	1	0	1	1	True	'zeros'
	64	128	4	2	0	1	1	True	'zeros'
	128	128	4	2	0	1	1	True	'zeros'
	128	256	4	1	0	1	1	True	'zeros'
	256	256	4	1	0	1	1	True	'zeros'
	256	512	4	1	0	1	1	True	'zeros'
	512	512	4	1	0	1	1	True	'zeros'
	512	1	4	1	0	1	1	True	'zeros'

Because the source images used are already the highest resolution (10m) images of Sentinel-2, there are no real ground truth images (2.5m resolution) that can be compared with the generated images in reality, and some image quality assessment metrics commonly used, such as, PSNR, SSIM, etc., are no longer applicable in this scene. Therefore, this paper adopts non-reference image quality assessment (NR-IQA) metrics, including NIQE [49], BRISQUE [50] and PIQE [51]. NIQE is a fully-blind image quality assessment model, and it establishes a "quality awareness" statistical feature set based on a simple and effective statistical model under a natural scene in the spatial domain, and only uses the measurable deviations of the statistical regularity observed in natural images for training. BRISQUE is a general non-reference image quality assessment model based on natural scene statistics in the spatial domain. BRISQU does not calculate the distortion-specific features, but uses the scene statistics of locally normalized luminance coefficients to quantify the possible "natural" losses. Without any training data, PIQE quantifies distortion, and relies on extracting local features to evaluate the image quality. The evaluation values of NIQE, BRISQUE and PIQE can be calculated by the corresponding functions niqe, brisque and piqe in Matlab, and the output results of the three functions are all within the

 doi:10.20944/preprints202104.0556.v1

range of [0, 100], where the lower number indicates high perceptual quality and the higher number indicates low perceptual quality.

Table 2. Setting of specific parameters for the models implemented in the framework of BasicSR

Model	EDSR8-RGB	RCAN	RS-ESRGAN
Network	**network_g:** type: EDSR num_in_ch: 3 num_out_ch: 3 num_feat: 256 num_block: 32 upscale: 4 res_scale: 0.1 img_range: 255. rgb_mean: [0.4488, 0.4371, 0.4040]	**network_g:** type: RCAN num_in_ch: 3 num_out_ch: 3 num_feat: 64 num_group: 10 num_block: 20 squeeze_factor: 16 upscale: 4 res_scale: 1 img_range: 255. rgb_mean: [0.4488, 0.4371, 0.4040]	**network_g:** type: RRDBNet num_in_ch: 3 num_out_ch: 3 num_feat: 64 num_block: 23 **network_d:** type: VGGStyleDis-criminator128 num_in_ch: 3 num_feat: 64
Traing	**optim_g:** type: Adam learning rate: 0.0001 weight_decay: 0 betas: [0.9, 0.99] **scheduler:** type: MultiStepLR milestones: [200000] gamma: 0.5 total_iter: 300000	**optim_g:** type: Adam learning rate: 0.0001 weight_decay: 0 betas: [0.9, 0.99] **scheduler:** type: MultiStepLR milestones: [200000] gamma: 0.5 total_iter: 300000	**optim_g:** type: Adam learning rate: 0.0001 weight_decay: 0 betas: [0.9, 0.99] optim_d: type: Adam learning rate: 0.0001 weight_decay: 0 betas: [0.9, 0.99] **scheduler:** type: MultiStepLR milestones: [50000, 100000, 200000, 300000] gamma: 0.5 total_iter: 400000

Because the source images used are already the highest resolution (10m) images of Sentinel-2, there are no real ground truth images (2.5m resolution) that can be compared with the generated images in reality, and some image quality assessment metrics commonly used, such as, PSNR, SSIM, etc., are no longer applicable in this scene. Therefore, this paper adopts non-reference image quality assessment (NR-IQA) metrics, including NIQE [49], BRISQUE [50] and PIQE [51]. NIQE is a fully-blind image quality assessment model, and it establishes a "quality awareness" statistical feature set based on a simple and effective statistical model under a natural scene in the spatial domain, and only uses the measurable deviations of the statistical regularity observed in natural images for training. BRISQUE is a general non-reference image quality assessment model based on natural scene statistics in the spatial domain. BRISQU does not calculate the distortion-specific features, but uses the scene statistics of locally normalized luminance coefficients to quantify the possible "natural" losses. Without any training data, PIQE quantifies distortion, and relies on extracting local features to evaluate the image quality. The evaluation values of NIQE, BRISQUE and PIQE can be calculated by the corresponding functions niqe, brisque and piqe in Matlab, and the output results of the three functions are all within the

 doi:10.20944/preprints202104.0556.v1

range of [0, 100], where the lower number indicates high perceptual quality and the higher number indicates low perceptual quality.

This paper randomly selects one sub-dataset "ROIs1158_spring_106" in ROIs1158 as the testing dataset (ROI_Te) containing 784 images. The remote sensing images in ROI_Te are collected from the ground areas as shown in Figure 10. In the figure, we marked 8 regions with strong geographic features, and the x4 generated images of these regions will be listed subsequently to visually compare the differences among those models.

Figure 10. Ground map corresponding to sub-dataset "Rois1158_spring_106"

Use BiCubic, EDSR8-RGB, RCAN, RS-ESRGAN and DKN-SR-GAN models to process 784 images in ROI_Te respectively to generate x4 HR images, and use Matlab to calculate evaluation values of NIQE, BRISQUE and PIQE one by one for the images. The histograms are drawn according to the distribution of evaluation values, as shown in Figure 11, Figure 12, and Figure 13, and the mean and extreme values based on the evaluation values are provided in Table 3. It can be seen from the histograms and Table 3 that our proposed DKN-SR-GAN model is superior to other models in a variety of non-reference image quality assessment metrics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

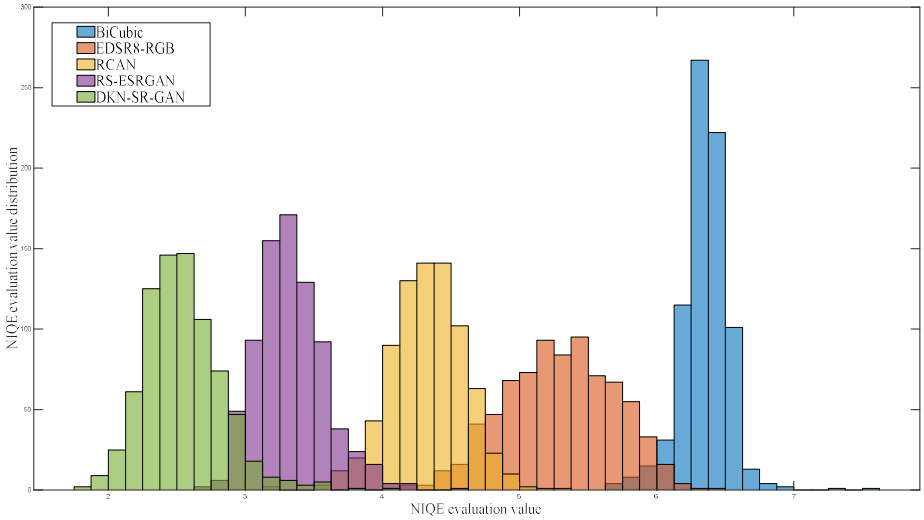

Figure 11. Distribution of evaluation values of non-reference image quality assessment metric NIQE

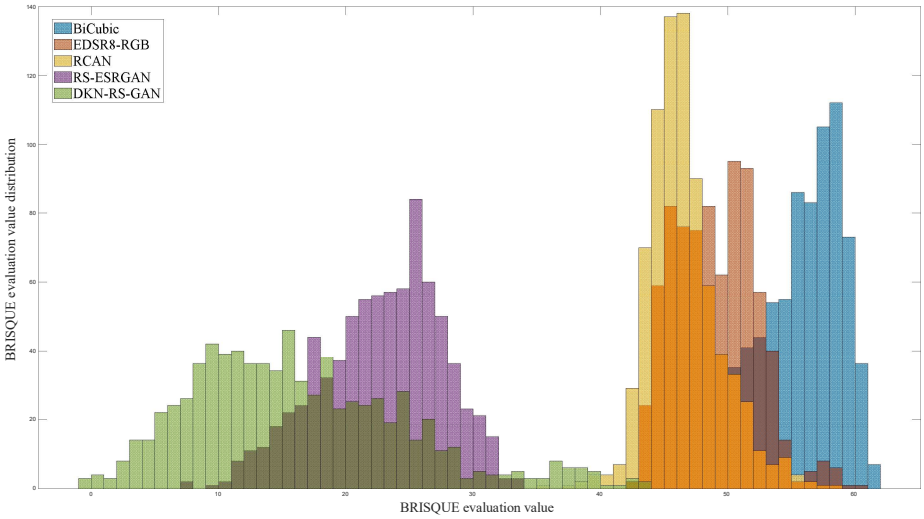

Figure 12. Distribution of evaluation values of non-reference image quality assessment metric BRISQUE

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

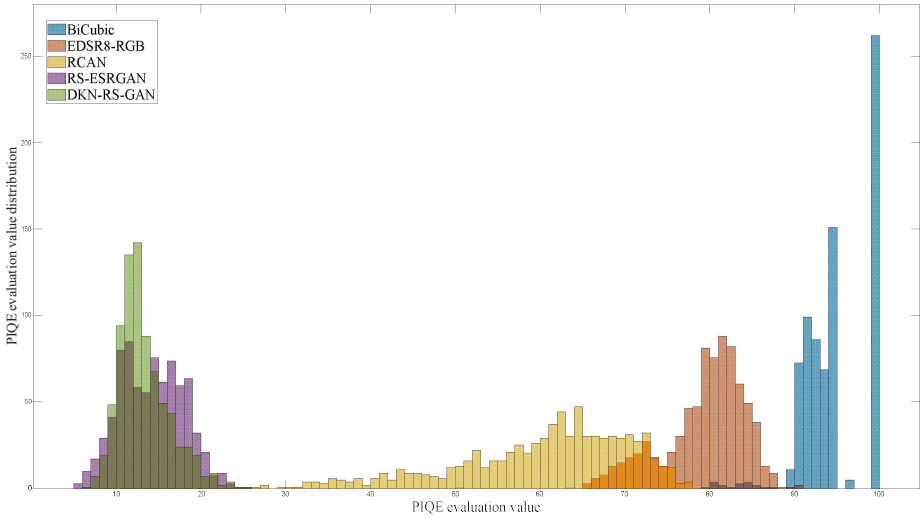

Figure 13. Distribution of evaluation values of non-reference image quality assessment metric PIQE

Table 2. Statistics of NIQE, BRISQUE and PIQE evaluation values

	BiCubic	EDSR8-RGB	RCAN	RS-ESRGAN	DKN-SR-GAN
NIQE mean	**6.349**	**5.296**	**4.329**	**3.337**	**2.544**
NIQE max	7.607	6.381	5.260	4.608	4.120
NIQE min	5.639	4.086	3.180	2.678	1.816
BRISQUE mean	**55.662**	**49.041**	**46.564**	**22.786**	**16.408**
BRISQUE max	61.535	60.014	58.167	33.340	43.405
BRISQUE min	44.464	42.699	35.306	7.886	3.424
PIQE mean	**94.635**	**79.374**	**60.333**	**14.186**	**13.231**
PIQE max	100.000	33.340	77.966	25.524	25.293
PIQE min	50.000	65.191	25.925	5.707	6.900

Figures 14-21 show the generated images of 8 regions with strong geographic features selected in "ROIs1158_spring_106" to visually compare the differences between different models. Through the comparison of the images of various terrains in Figures 14-21, it can be obviously seen that the images processed by traditional BiCubic method are bleariest and smoothest due to the inherent deficiencies of the interpolation algorithm. EDSR8-RGB, RCAN and RS-ESRGAN models cannot correctly distinguish the noises with sharp edges, resulting in blurred results, and even indistinguishable for houses and roads. As shown in our DKN-SR-GAN results, the dividing lines among the objects and the backgrounds such as roads, bridges and houses, are much clearer, which indicates that the noise estimated by noise injection is closer to the real noise. Compared with EDSR8-RGB, RCAN and RS-ESRGAN models, our DKN-SR-GAN results are clearer and have no ambiguity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

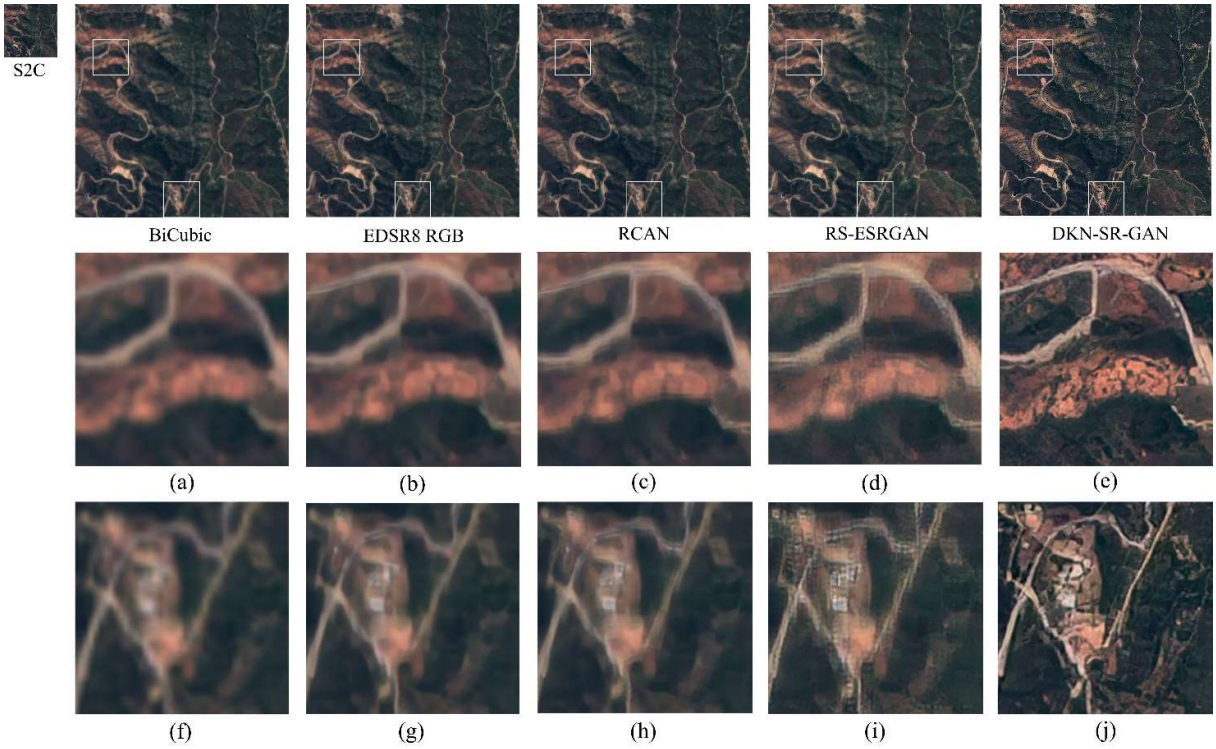

Figure 14. Comparison of visual effects of the generated images of the region containing mountain-road terrain

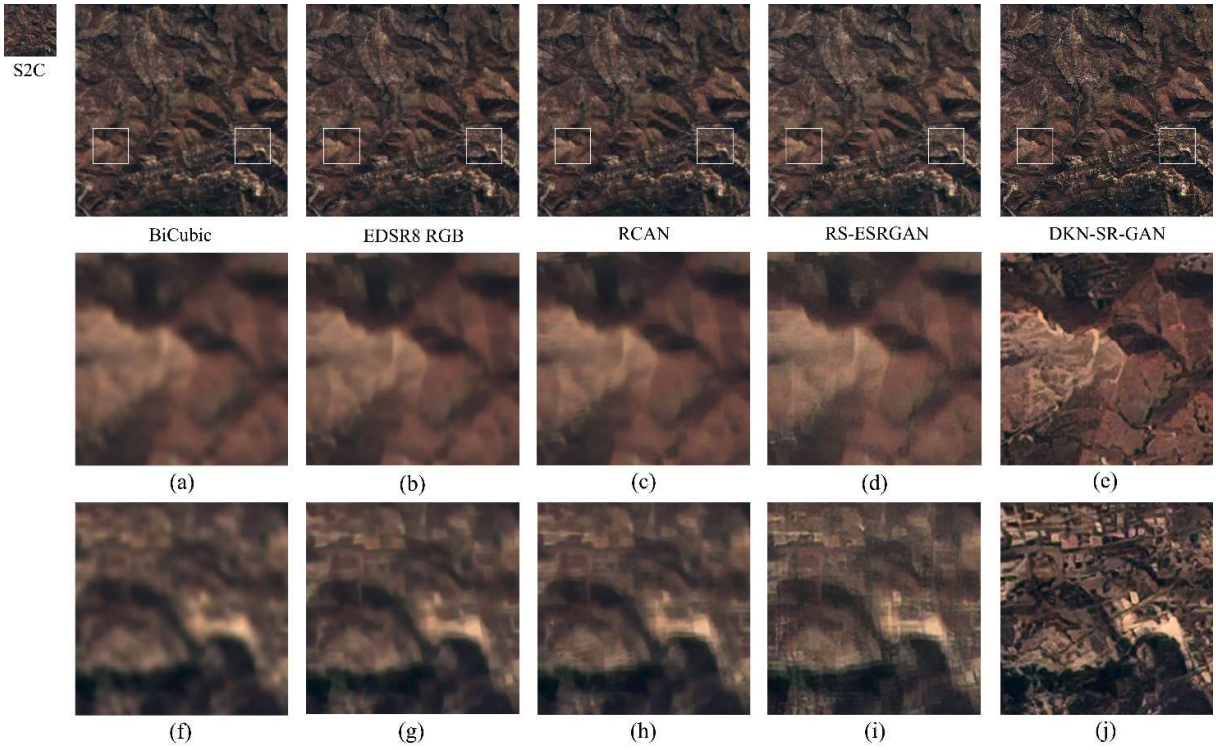

Figure 15. Comparison of visual effects of the generated images of the region with hilly terrain

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

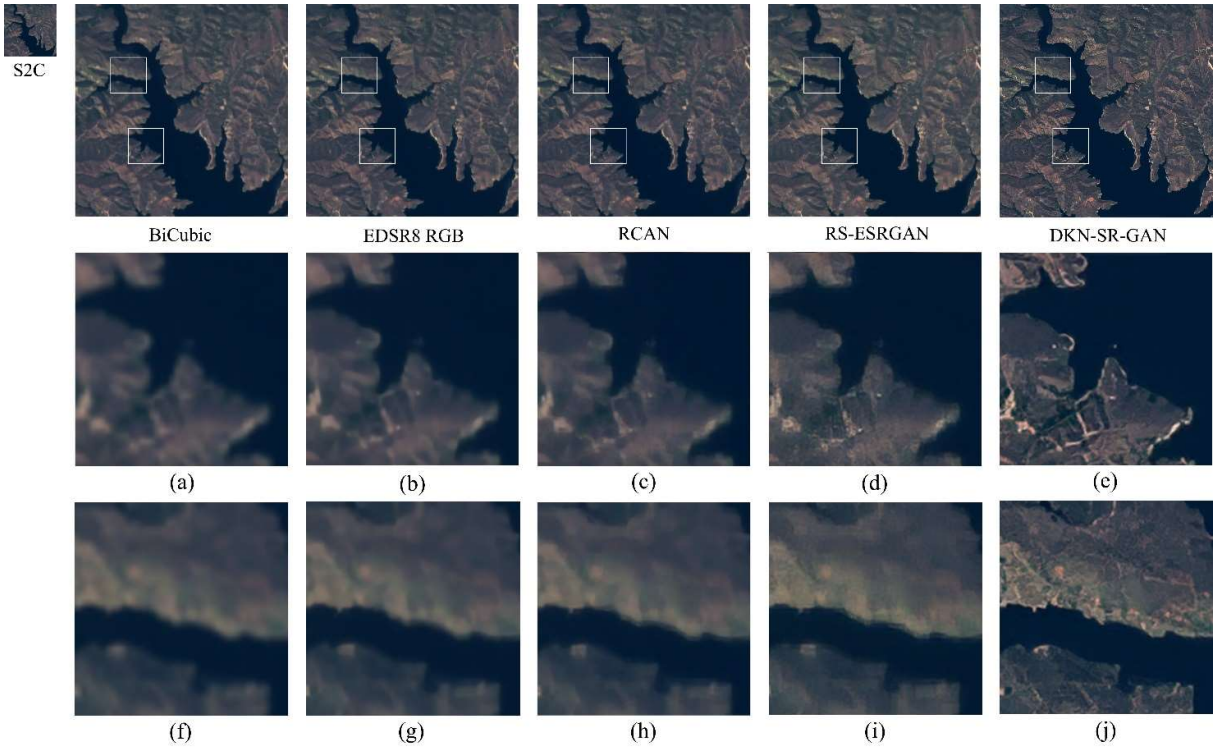

Figure 16. Comparison of visual effects of the generated images of the region containing surface water terrain

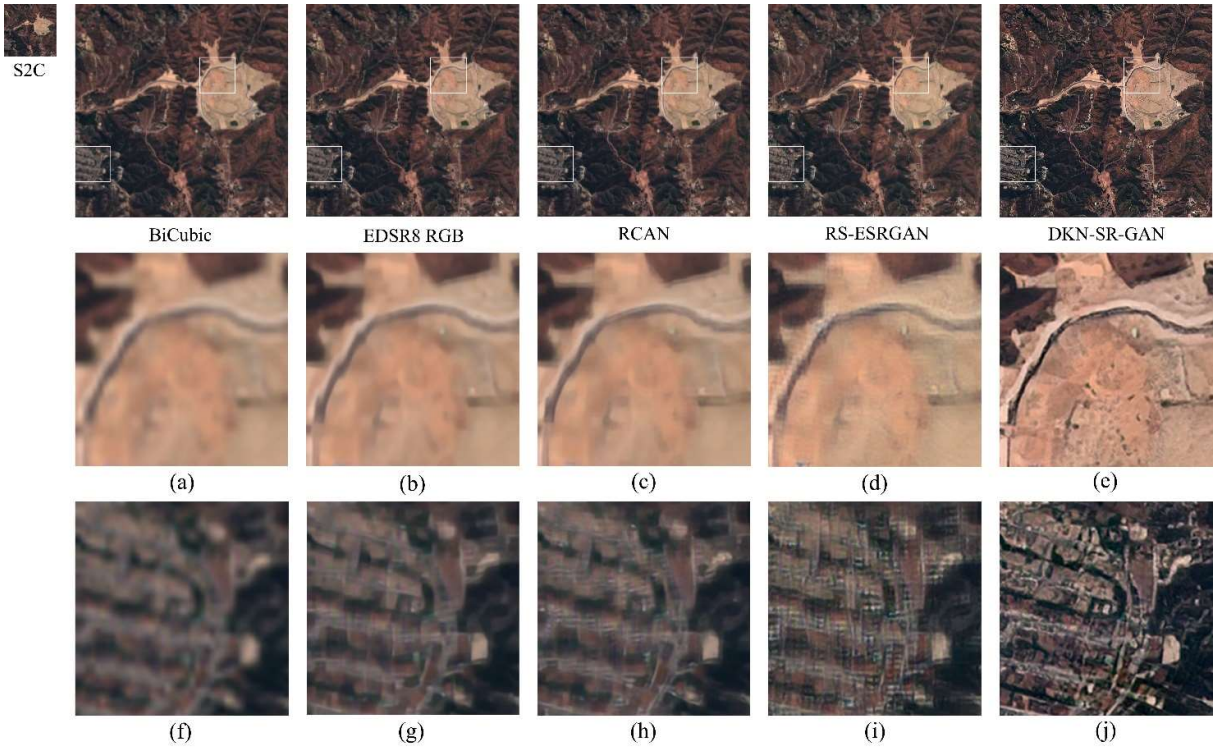

Figure 17. Comparison of visual effects of the generated images of the region containing dry river beds and residential houses

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

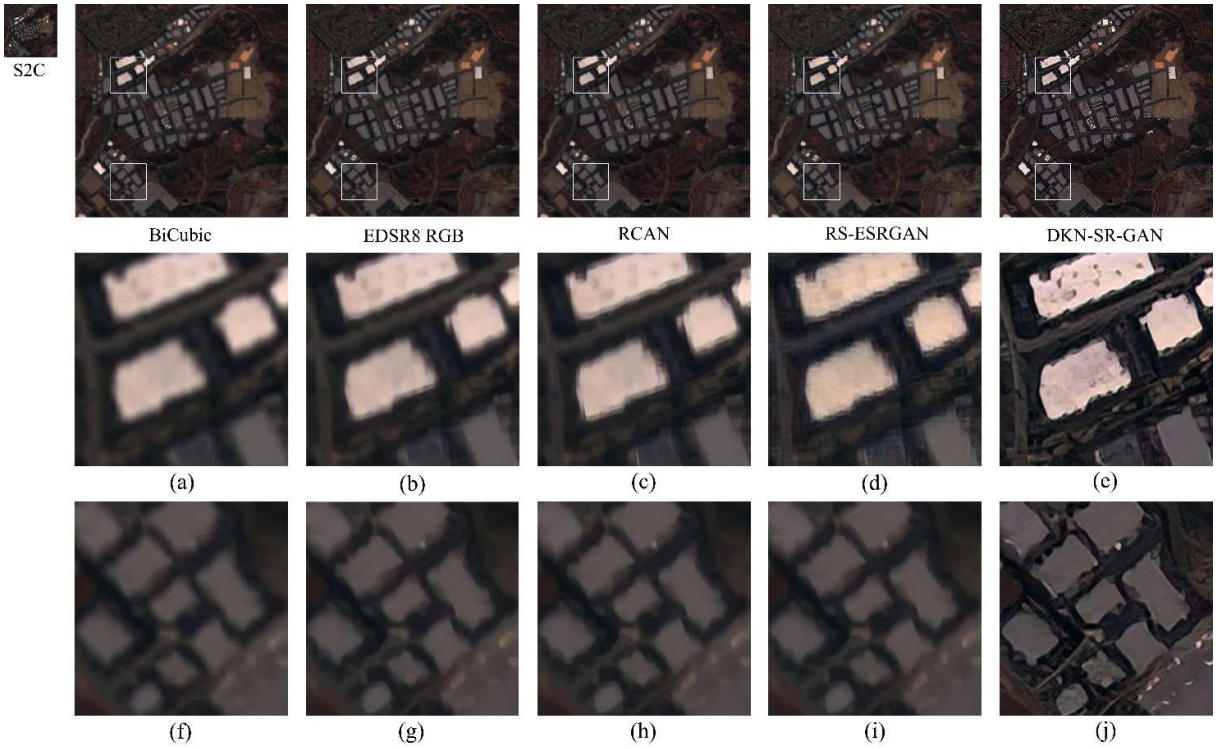

Figure 18. Comparison of visual effects of the generated images of the region containing factories

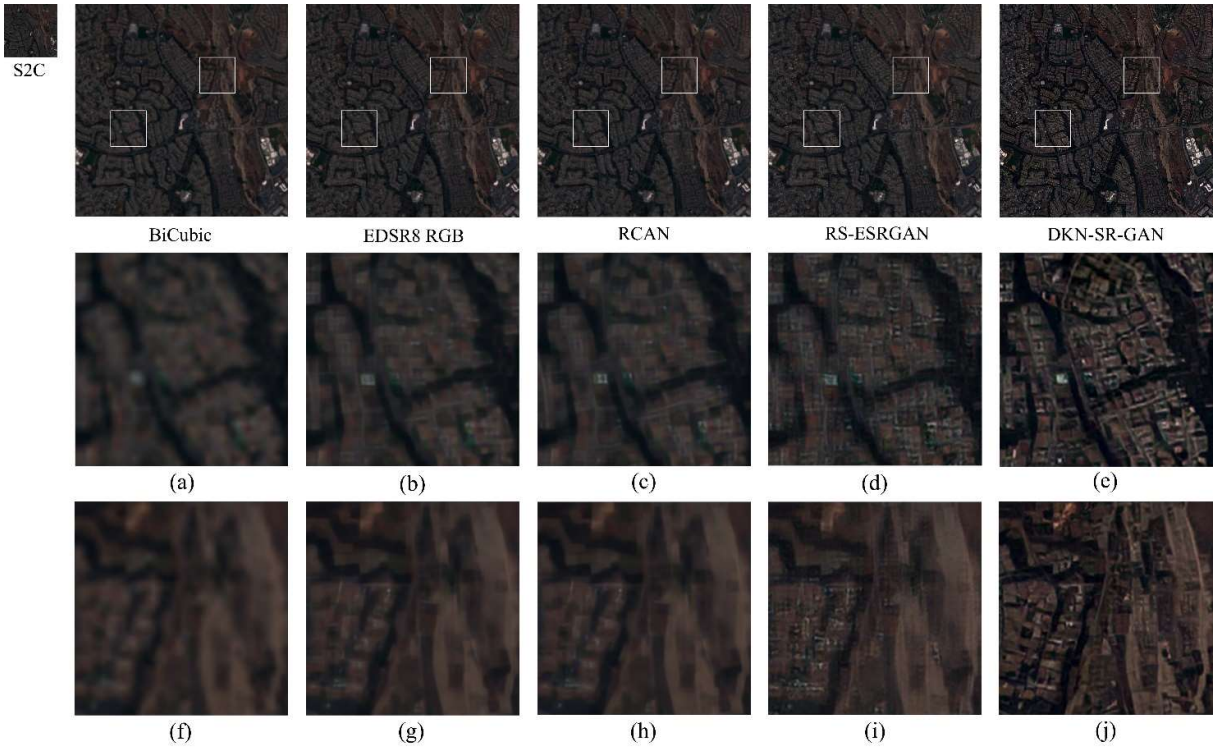

Figure 19. Comparison of visual effects of the generated images of the region containing residential houses

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021 doi:10.20944/preprints202104.0556.v1

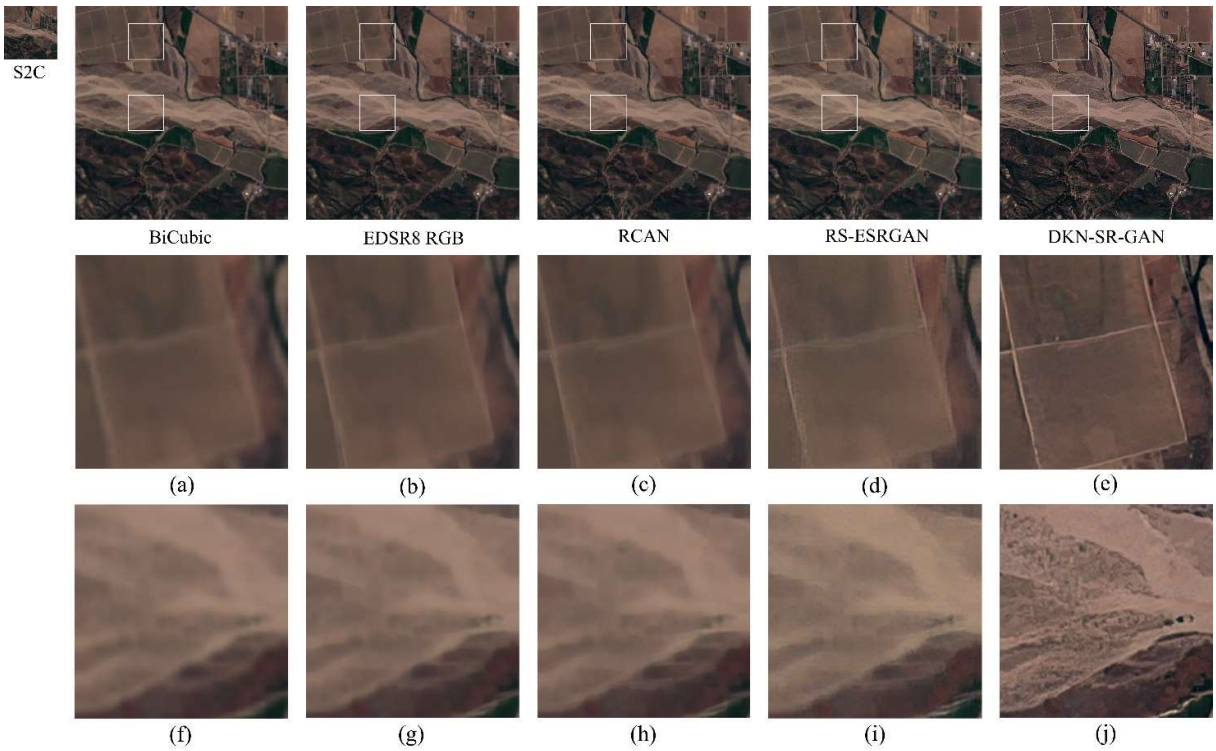

Figure 20. Comparison of visual effects of the generated images of the region containing farmlands and sandy terrain

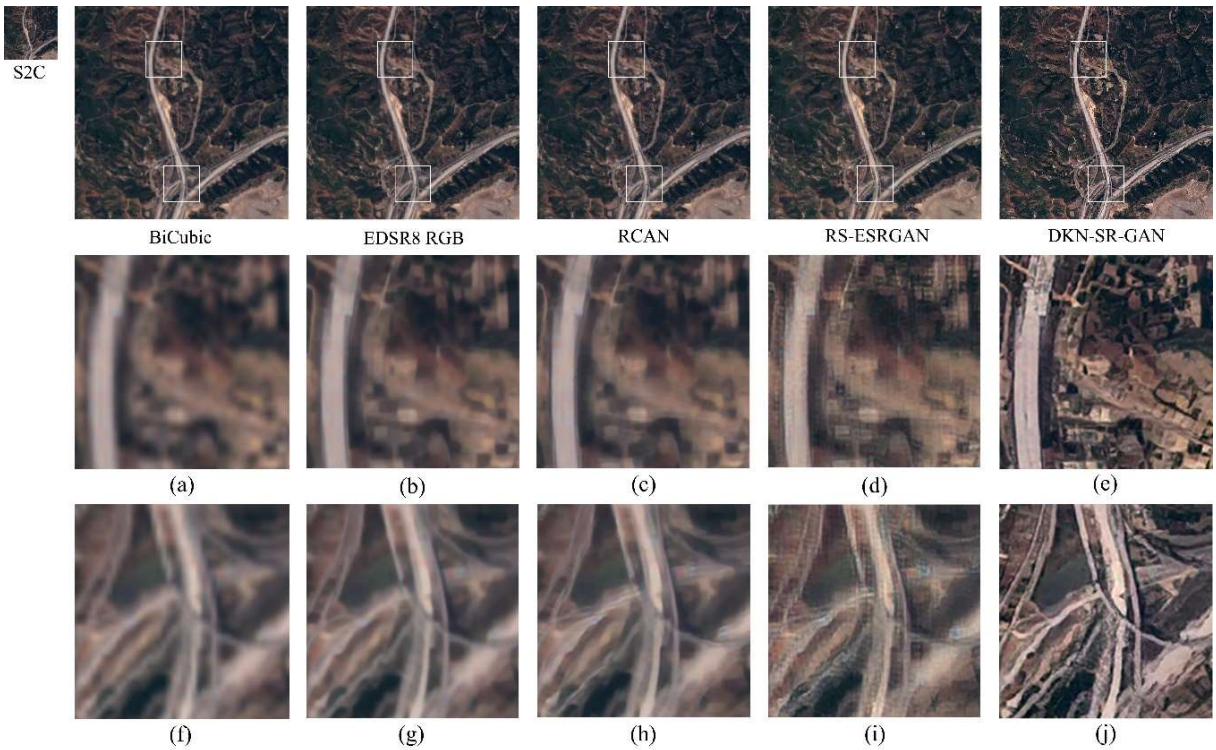

Figure 21. Comparison of visual effects of the generated images of the region containing overpasses

5. Conclusion

In this paper, based on the latest and widely-recognized GAN technologies such as KernelGAN, ESRGAN, PatchGAN and so on, we introduce the degradation kernel estimation and noise injection, to perform SR for Sentinel-2 satellite remote sensing images, and improve the original images with the highest resolution of 10m to 2.5m. Through the combination of the degradation kernel and injected noise, we obtain LR images in the same domain as the real images, and get the near-natural LHR image pairs. On the basis of near-natural LHR image pairs, we use GAN combined of ESRGAN-type generator, PatchGAN-type discriminator and VGG-19-type feature extractor, use the perceptual loss, and focus on the visual characteristics of the images, so that our results have clearer details and better perceptual effects. Compared with state-of-the-art SR models of Sentinel-2 such as EDSR8-RGB，RCAN and RS-ESRGAN, the main difference of our model lies in the construction of LHR image pairs for the training datasets. In the scene training with natural LHR image pairs, there is no significant difference in the effect for SR images obtained by those models; however, in the scene with only LR images and no HR prior information, compared with RCAN which constructs the image pairs through BiCubic, with EDSR8-RGB and RS-ESRGAN which use WorldView satellite HR images to construct the image pairs, KDN-SR-GAN have obvious advantages in the quantitative comparison of the non-reference image quality assessment and the intuitive visual effects.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; software, Y.L.; validation, Y.L.; formal analysis, Y.L.; resources, Y.L. and B.L.; writing—original draft preparation, Y.L. and B.L.; writing—review and editing, B.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Guangdong Province, China (Grant No. 2018A030313346); General University Key Field Special Project of Guangdong Province, China (Grant No. 2020ZDZX3078); General University Characteristic Innovation Project of Guangdong Province, China (Grant No. 2018KTSCX250); Research Fund Program of Guangdong Key Laboratory of Aerospace Communication and Networking Technology (Grant No. 2018B030322004).

Data Availability Statement: The x4 images by models KDN-SR-GAN, BiCubic, EDSR8-RGB，RCAN and RS-ESRGAN are available online at Baidu Wangpan(code: mbah). And all data, models, and code generated or used during the study will be available at GitHub soon.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Verbyla, D.L. *Satellite remote sensing of natural resources*; CRC Press: 1995; Volume 4.
2. Ye, B.; Tian, S.; Ge, J.; Sun, Y. Assessment of WorldView-3 data for lithological mapping. *Remote Sensing* **2017**, *9*, 1132.
3. Williams, D.L.; Goward, S.; Arvidson, T. Landsat. *Photogrammetric Engineering & Remote Sensing* **2006**, *72*, 1171-1178.
4. Sentinel, E. User Handbook. *ESA Standard Document 64*.
5. Gargiulo, M.; Mazza, A.; Gaetano, R.; Ruello, G.; Scarpa, G. Fast Super-Resolution of 20 m Sentinel-2 Bands Using Convolutional Neural Networks. *Remote Sensing* **2019**, *11*, 2635, doi:10.3390/rs11222635.
6. Wang, Q.; Shi, W.; Li, Z.; Atkinson, P.M. Fusion of Sentinel-2 images. *Remote sensing of environment* **2016**, *187*, 241-252.
7. Yang, X.; Zhao, S.; Qin, X.; Zhao, N.; Liang, L. Mapping of urban surface water bodies from Sentinel-2 MSI imagery at 10 m resolution via NDWI-based image sharpening. *Remote Sensing* **2017**, *9*, 596.
8. Gargiulo, M.; Mazza, A.; Gaetano, R.; Ruello, G.; Scarpa, G. A CNN-based fusion method for super-resolution of Sentinel-2 data. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018; pp. 4713-4716.
9. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. Sentinel-2 image fusion using a deep residual network. *Remote Sensing* **2018**, *10*, 1290.
10. Wang, Q.; Atkinson, P.M. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment* **2018**, *204*, 31-42.
11. Wang, J.; Huang, B.; Zhang, H.K.; Ma, P. Sentinel-2A image fusion using a machine learning approach. *IEEE Transactions on Geoscience and Remote Sensing* **2019**, *57*, 9589-9601.

12. Rumora, L.; Gašparović, M.; Miler, M.; Medak, D. Quality assessment of fusing Sentinel-2 and WorldView-4 imagery on Sentinel-2 spectral band values: a case study of Zagreb, Croatia. *International Journal of Image and Data Fusion* **2020**, *11*, 77-96.
13. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE transactions on image processing* **2010**, *19*, 2861-2873.
14. Gou, S.; Liu, S.; Yang, S.; Jiao, L. Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 4784-4792.
15. Pan, Z.; Yu, J.; Huang, H.; Hu, S.; Zhang, A.; Ma, H.; Sun, W. Super-resolution based on compressive sensing and structural self-similarity for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 4864-4876.
16. Zhang, Y.; Du, Y.; Ling, F.; Fang, S.; Li, X. Example-based super-resolution land cover mapping using support vector regression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 1271-1283.
17. Li, J.; Yuan, Q.; Shen, H.; Meng, X.; Zhang, L. Hyperspectral image super-resolution by spectral mixture analysis and spatial–spectral group sparsity. *IEEE Geoscience and Remote Sensing Letters* **2016**, *13*, 1250-1254.
18. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2020**.
19. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European conference on computer vision, 2014; pp. 184-199.
20. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016; pp. 1646-1654.
21. 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 Proceedings of the European conference on computer vision (ECCV), 2018; pp. 286-301.
22. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 3147-3155.
23. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 7132-7141.
24. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv preprint arXiv:1406.2661* **2014**.
25. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 4681-4690.
26. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018; pp. 0-0.
27. Ma, W.; Pan, Z.; Guo, J.; Lei, B. Super-resolution of remote sensing images based on transferred generative adversarial network. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018; pp. 1148-1151.
28. Haut, J.M.; Paoletti, M.E.; Fernández-Beltran, R.; Plaza, J.; Plaza, A.; Li, J. Remote sensing single-image superresolution based on a deep compendium model. *IEEE Geoscience and Remote Sensing Letters* **2019**, *16*, 1432-1436.
29. Haut, J.M.; Fernandez-Beltran, R.; Paoletti, M.E.; Plaza, J.; Plaza, A. Remote sensing image superresolution using deep residual channel attention. *IEEE Transactions on Geoscience and Remote Sensing* **2019**, *57*, 9277-9289.
30. Lei, S.; Shi, Z.; Zou, Z. Super-resolution for remote sensing images via local–global combined network. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 1243-1247.
31. Gong, Y.; Liao, P.; Zhang, X.; Zhang, L.; Chen, G.; Zhu, K.; Tan, X.; Lv, Z. Enlighten-GAN for Super Resolution Reconstruction in Mid-Resolution Remote Sensing Images. *Remote Sensing* **2021**, *13*, 1104, doi:10.3390/rs13061104.
32. Galar, M.; Sesma, R.; Ayala, C.; Albizua, L.; Aranda, C. Super-Resolution of Sentinel-2 Images Using Convolutional Neural Networks and Real Ground Truth Data. *Remote Sensing* **2020**, *12*, doi:10.3390/rs12182941.
33. Salgueiro Romero, L.; Marcello, J.; Vilaplana, V. Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks. *Remote Sensing* **2020**, *12*, doi:10.3390/rs12152424.
34. Zhang, W.; Liu, Y.; Dong, C.; Qiao, Y. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019; pp. 3096-3105.
35. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-enhanced GAN for remote sensing image superresolution. *IEEE Transactions on Geoscience and Remote Sensing* **2019**, *57*, 5799-5812.
36. Wang, Z.; Jiang, K.; Yi, P.; Han, Z.; He, Z. Ultra-dense GAN for satellite imagery super-resolution. *Neurocomputing* **2020**, *398*, 328-337.
37. Zhu, X.; Zhang, L.; Zhang, L.; Liu, X.; Shen, Y.; Zhao, S. GAN-based image super-resolution with a novel quality loss. *Mathematical Problems in Engineering* **2020**, *2020*.
38. Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F. Real-world super-resolution via kernel estimation and noise injection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020; pp. 466-467.

39. Bell-Kligler, S.; Shocher, A.; Irani, M. Blind super-resolution kernel estimation using an internal-gan. *arXiv preprint arXiv:1909.06581* **2019**.
40. Chen, J.; Chen, J.; Chao, H.; Yang, M. Image blind denoising with generative adversarial network based noise modeling. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; pp. 3155-3164.
41. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS - A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2019**, *IV-2/W7*, 153-160, doi:10.5194/isprs-annals-IV-2-W7-153-2019.
42. Wang, H.-p.; Zhou, L.-l.; Zhang, J. Region-based Bicubic image interpolation algorithm. *Computer Engineering* **2010**, *19*, 216-218.
43. Shocher, A.; Bagon, S.; Isola, P.; Irani, M. Ingan: Capturing and remapping the" dna" of a natural image. *arXiv preprint arXiv:1812.00231* **2018**.
44. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 1125-1134.
45. Zhou, R.; Susstrunk, S. Kernel modeling super-resolution on real low-resolution images. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019; pp. 2433-2443.
46. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
47. 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, 2016; pp. 694-711.
48. Wang, X.; Yu, K.; Chan, K.C.K.; Dong, C.; Loy, C.C. BasicSR. **2020**.
49. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters* **2012**, *20*, 209-212.
50. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing* **2012**, *21*, 4695-4708.
51. Venkatanath, N.; Praneeth, D.; Bh, M.C.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In Proceedings of the 2015 Twenty First National Conference on Communications (NCC), 2015; pp. 1-6.