HRPGAN: A GAN-based Model to Generate High-resolution Remote Sensing Images

Hai Sun1, Ping Wang2, Yifan Chang3, Li Qi4, Hailei Wang3, Dan Xiao5, Cheng Zhong6, Xuelian Wu3, Wenbo Li3, Bingyu Sun7

1 China Telecom Shanghai Ideal Information Industry (Group) Co., Ltd, Shanghai, China
2 Department of Airborne Remote Sensing, National Disaster Reduction Center of China, Beijing, China
3 Institute of Technology Innovation, Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei, China
4 The Third Research Institute of the Ministry of Public Security, Shanghai, China
5 Guizhou Key Laboratory of Economics System Simulation, Guizhou University of Finance and Economics, Guiyang, China
6 Beijing Aerospace Flight Control Center, Beijing, China
7 Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China
Email: wbli@iim.ac.cn; 2462473897@qq.com

Abstract. Generative adversarial networks (GAN) has been mainly used in the generation of natural images such as MNIST, CIFAR10 as well as Imagenet datasets and achieves satisfying generation results. However, GAN always fails in generating high quality high-resolution remote sensing images because remote sensing images are large in size and have various ground objects. To address this issue, a novel framework called High-Resolution PatchGAN (HRPGAN) is introduced in this paper. The structure of HRPGAN follows PatchGAN, but the batch normalization layers are removed and the ReLU activation is replaced by the SELU activation. In addition, a new loss function consisting of the adversarial loss, perceptual reconstruction loss and regularization loss is used in HRPGAN. Experiment results show that the proposed HRPGAN model generates the more diverse and lifelike images in HR remote sensing generation than Bicubic method and TGAN model.

1. Introduction
High-Resolution (HR) images, which contain abundant detailed information, have been widely studied in recent years as a crucial technique for remote sensing applications [1], such as medicine, television imaging, satellite, aviation and so on. However, only few HR remote sensing images are available because they are time-consuming and expensive. Many methods to generate HR remote sensing images have been developed including interpolation methods [2, 3], statistic methods [5, 6], learning-based methods [6, 7], image reconstruction [8,9] and deep learning (DL) based on convolutional neural networks (CNNs) [10]. The generation results of traditional method are always overly smoothed, lack high-frequency detail, perceptually unsatisfying [11] while DL generates more lifelike images because it automatically learns deep-level feature representations from images. However, some generation results of remote sensing utilizing DL algorithm are still unsatisfying. In this paper, a new model based on Generative adversarial networks (GAN), which is one kinds of DL technique, is proposed to generate HR remote sensing images. GAN is a popular deep generative
model proposed by Goodfellow [12]. GAN learns a data distribution and realizes a model to sample from it as shown in Figure 1. GAN is composed of a generator (G) and a discriminator (D). In GAN, G takes a random noise vector as input and outputs an image. At the same time, the D differentiates whether the generated image is real. Although GAN can generate images automatically, it needs a lot of manual effort to adjust the parameters and still produces unrealistic results. PatchGAN [13] is proposed to generate more distinct images by adding skip connections between each layer in G and examining local image patches in D. PatchGAN shares low-level information with the input images and models high-frequency structure of images. There are main two reasons for PatchGAN is suitable to the task of recovering of HR remote sensing images. The first reason is that Low-Resolution (LR) and HR remote sensing images share the similar information in low-level features like colors and outlines. The second reason is that HR images are required to recover the high-level features like textures and edges from LR images.

Some researches indicated that the batch normalization (BN) layers should be removed in deblurring images because of the batch normalization layers would normalize and smooth the features and get rid of range flexibility [14]. It is also found that BN contributions a lot in the classification task but would result in blurriness in the area of image super-solution [15]. In addition, BN causes large memory consumption and the computational burden. Traditionally, ReLU activation is always combined with BN. The results indicate [16] that replacing BN and ReLU with SELU not only requires much less time to train but also has almost the same effect as BN and ReLU. Inspired by these findings, BN is removed and ReLU is replaced by SELU in HRPGAN.

The loss function of GAN forces G to produce lifelike images that D can’t distinguish from the real images. Some researches [17] added the pixel-level loss to the loss function of GAN to make the generated images smoother and more similar to the real images in low-level features. However, pixel-level loss can’t minimize the difference between generated images and real images in high-frequency features such as textures and edges details [18] which are extremely necessary information in HR remote sensing images. Therefore, pixel-level loss and feature-level loss the original adversarial loss of GAN are combined in HRPGAN.

To sum up, HRPGAN is based on the structure of PatchGAN with the removal of BN and replacement of ReLU with SELU. In addition, low-frequency and high-frequency loss function are combined with the original adversarial loss of GAN.

The contributions of this paper are the following:

1. We proposed a new GAN model called HRPGAN to generate high-resolution remote sensing images based on PatchGAN with some adjustments which are removing the BN layers and replacing ReLU with SELU.
2. We added content loss and perceptual loss to the original GAN adversarial loss to cater to characteristic of abundant details information in remote sensing images.
3. HRPGAN achieves the highest scores in PSNR and SSIM as well as lowest scores in FID comparable with the traditional method Bicubic [19] and TGAN [1].
2. Related works
HR remote sensing images generation methodologies can be casted into two categories which refers to traditional methods and learning methods.

Traditional methods mainly include interpolation, statistic methods, image reconstruction and others methods. For example, Yang et al. utilized sparse-code to describe HR and LR images [20]. Li et al. used the sparsity prior of image statistics to generate HR image [21]. Pan et al. proposed compressive sensing and structural self-similarity method to recover super resolution of single-image [22]. Timofte et al proposed Anchored Neighborhood Regression (ANG) [23], which greatly reduced the generation time of high resolution images. Ponomaryov et al. generated high resolution images by combining discreet wavelet transfer and sparse representation [24].

In recent years, deep learning methods have achieved impressive results on HR remote sensing images generation. For example, Christian et al. proposed SRGAN based on GAN and added a perceptual loss function to the original loss function of GAN to recover super-resolution [11]. ESRGAN improved RSGAN, introducing the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit, and borrowing the idea from relativistic GAN to let the discriminator predict relative realness instead of the absolute value [25]. LGCNet (Local-Global Combined Network) was applied into the GAN to generate high quality remote sensing images based on the deep CNN [26].

Although deep learning methods outperform traditional methods in generating HR remote sensing images, they can’t still generate HR images with detailed low-frequency information and satisfying feature-level information. In this work, we attempt to generate more clear and lifelike remote sensing images.

3. Method
HRPGAN is proposed for generation of HR remote sensing images in this work. HRPGAN basically follows the structure of PatchGAN [13] but some are changes are made. Firstly, BN are omitted according to the findings that BN normalizes the features and gets rid of range flexibility [27], which are adverse to generation of super-resolution. Secondly, SELU activation are propagated through many network layers and will converge towards zero mean and unit variance [14], thus the RELU activation is replaced with SELU activation. Thirdly, we draw on the advantage of perceptual reconstruction loss and content loss in the loss function of HRPGAN [11].

3.1. The Structure of PHRGAN
3.1.1. The Structure of PatchGAN
For G, PatchGAN believes that for many images translation problems, there is a great deal of low-level information shared between the input and output, and it would be desirable to shuttle this information directly across the net. For recovering HR remote sensing images, the input LR images share the similar underlying structure with the output HR images. Therefore, the idea of skip connection in U-net [28] is also used in PatchGAN. The skip connection is shown in Figure 2. For D, PatchGAN restricts attention to the structure in local image patches to mode high-frequencies of images by penalizing structure at the scale of patches [13]. In remote sensing images, high-frequencies information is necessary and significant because remote sensing images contain abundant texture details and edges. In addition, PatchGAN reduces the running time, the number of parameters and storage significantly when the image is divided into patches, which is suitable to the remote sensing images large in size.
3.1.2. The Removal of Batch Normalization Layers
Batch normalization (BN) is first proposed by to reduce internal covariate shift and the dependence of gradients on the scale of the parameters or of their initial values, and thus accelerates the training of deep neural nets [29]. BN achieves satisfying performance on classification task. However, firstly these classification neural network models are always trained by the natural images such as Cifar10, LSUN and Imagenet datasets. BN normalizes the features and help images to be smoother. However, it is undesirable in HR remote sensing images because ground objects contain abundant textures and sharp edges. Secondly, remote sensing images are larger in size and have more ground objects than natural images, and thus need more parameters and training time. BN consumes large amount of time to fix the means and variances of layer inputs and thus is supposed to be removed. Therefore, BN is removed in HRPGAN as shown in Figure 3.

3.1.3. The utilization of SELU Activation
SELU activation is propagated through many network layers and will converge towards zero mean and unit variance and achieves lower training loss and stable training process than BN [14]. The SELU activation function is given by equation (1):

\[
SELU(x) = \begin{cases} 
\lambda x & x > 0 \\
\alpha \lambda e^x - \alpha \lambda & x \leq 0 
\end{cases}
\]

where 

\[
\lambda = 1.0507009873554804934193349852946 \\
\alpha = 1.6732632423543772848179429916717 
\]

Figure 2. The Skip Connection in CNN

Figure 3. (a) The CNN structure with BN, (b) The CNN structure without the BN
Finally, the structure of HRPGAN and G in HRPGAN are presented in Figure 4 and Figure 5, respectively.

Figure 4. The Structure of HRPGAN

Figure 5. The structure of G in HRPGAN

3.2. The New Loss Function

3.2.1. The Adversarial Loss

The original adversarial loss of GAN is equation (2).

\[
L_{al} = \min_G \max_D V(G, D) = E_{x \sim p_{data}}[\log D(x)] + E_{z \sim p_z(z)}[\log (1 - D(G(z)))]
\]  

(2)

Where, \(G\) is the generator; \(D\) is the discriminator; \(Z\) is the random noise vector fed into \(G\); \(X\) is the samples generated by \(G\). \(P_{data}\) is the true distribution of data, \(P_z(Z)\) is the distribution of the random noise. The adversarial loss tries to narrow the distribution between the fake images generated by \(G\) and original natural images.

3.2.2. The Perceptual Feature Loss

Perceptual loss [29] is an appropriate measure for features extracted from a pre-trained VGG. Perceptual loss minimizes the distance between two activated features, which increase the similarity between images in the high-level features. However, perceptual loss causes sparse activation and inconsistent reconstructed brightness of generated images, and overcame the drawbacks of the perceptual loss [25]. We employed the improved perceptual loss defined in equation (3).

\[
L_{pf} = \frac{1}{L \cdot W \cdot H} \sum_{j=1}^{L} \sum_{i=1}^{W} \sum_{h=1}^{H} (V(f_{\text{true}}^j) - V(f_{\text{fake}}^j))^2
\]  

(3)

Where, \(L\), \(W\), \(H\) represent the length, width and channels of images respectively. \(V\) represents a non-linear CNN transformation pre-trained by the fully connected layers.
3.2.3. The Pixel-wise Loss
The feature loss measures the perceptual similarity while the pixel-wise loss measures similarity in pixel space defined as equation (4).

$$L_{pw} = \frac{1}{LWH} \sum_{l=1}^{L} \sum_{w=1}^{W} \sum_{h=1}^{H} (I^{HR} - G(I^{LR}))^2$$  

(4)

Although pixel-wise loss may lose the high-frequency details of images such as textures, it can denoise the generated images and improve the images to be more visually smooth.

Our proposed loss function of HRPGAN given by equation (5) is combined with the adversarial loss, the feature loss and the pixel-wise loss.

$$L_{HRPGAN} = L_{al} + xL_{pf} + yL_{pw}$$

(5)

4. Experiment and Analysis

4.1. The Overview of Experiment
In this section, the detail information for model training and experiment results are presented.

Firstly, we down-sampled dataset to get the LR remote sensing images and then scaled the input images to the range of [-1, 1]. Batch-size was set as 64. ADAM [30] was selected to optimize our model with learning rate set as 0.0002. All of the parameters were initialized by a Gaussian distribution. We set $x=0.48\times10^{-1}$, $y=1.56\times10^{-6}$ in equation (5) based on multiple tests.

Peak signal to noise ratio (PSNR), structural similarity (SSIM) and Frechet Inception Distance (FID) [31] were selected as evaluation indexes. PSNR is one image quality evaluation based on error sensitivity. The larger PSNR is, the smaller the distortion of one image is. SSIM measures structural similarity between images based on their luminance, contrast and structure. The larger SSIM is, the more similar two images are. FID is an indicator specifically used in GAN to measure the distance between distributions of different images. FID is sensitive to both quality and diversity [32]. The generated images have samples the higher quality when the value of FID is lower. The baseline models are TGAN [1] and Bicubic [33]. TGAN also removes the BN and learns from transfer-learning method to recover HR remote sensing images.

4.2. Experiment Results and Analysis
The results of the proposed HRPGAN model for generated HR remote sensing images are presented in this section. The HR remote sensing images recovered by the proposed HRPGAN model have abundant textures, sharp edges as well as contrasting colors, which are comparable to the ground truth as shown in Figure 6.
In addition, the detail of training process is shown in Figure 7. With the increasing epochs, the outlines of the mountains gradually become clear and the gullies become distinct. From Table 1, the average PSNR of HRPGAN (29.33) is higher than Bicubic (24.97) and TGAN (27.81). When come to the evaluation index of SSIM, the proposed HRPGAN is also higher than the methods of Bicubic and TGAN. The FID of the proposed HRPGAN is 52.68 which is lower than both baseline models (Table 1). The three evaluation indexes indicate that the proposed HRPGAN model generates the more diverse and lifelike images in HR remote sensing generation. The reason why the proposed model outperforms TGAN mainly is that the more exacting D. The D in our model examines the large images in patches, which would force the G to generate images with more vivid details.

5. Conclusion
We proposed a novel model called HRPGAN to generate HR remote sensing images based on GAN in this paper. In HRPGAN, the structure of generator is based on the generator in PatchGAN but BN is removed and RELU is replaced by SELU activation. In addition, feature loss and pixel-wise loss are added to the original adversarial loss in order to generate images smoother and more similar to the ground truth in high-level features. Experiment results show that the proposed HRPGAN model outperforms the traditional method Bicubic and TAGN.
Figure 7. The training process of HRPGAN: 1 epoch, 2 epoch, 3 epoch, 5 epoch, 10 epoch, 15 epoch, 20 epoch, 25 epoch, 30 epoch, 35 epoch, 38 epoch and 40 epoch respectively. (from top to bottom, from left to right)

6. Acknowledgments
This paper is financially supported by the National Natural Science Foundation of China (No. 41871302, 61773360, 41701594) and the National Key Research and Development Program of China (No. 2018YFB1004600).
7. Reference

[1] Wang A, Wang Y, Song X, Iwahori Y 2019 Remote Sensing Image Super-Resolution Reconstruction based on Generative Adversarial Network Int. J. Perform. Eng. 15 1783-1791

[2] Ding H Y and Bian Z F 2008 Remote sensing image Super-resolution reconstruction based on interpolation Micro. Info. 24 292-293

[3] Liu Z J, Cai C, Peng X M, Zhou C P, Ding M Y 2004 A Novel Regularized Image Interpolation Algorithm Based on Genetic Algorithm J. Image Gna. 9 934-40

[4] Xu F, Tang M, Huang C F, Xu L Z 2011 Remote sensing image fusion using MAP super-resolution with enhanced edge New industr. 1 79-83

[5] Li L, Sui L C, Ding M T 2018 Improved Algorithm of Remote Sensing Images Super-Resolution Based on Nonparametric Bayesian Laser Optoelec. Prog. 55 032801

[6] Nasrollahi K and Moeslund T B 2014 Super-resolution: a comprehensive survey Mach. vision. app. 25 1423-68

[7] Li C T and Tian S F 2017 Super-resolution fusion method for remote sensing image based on dictionary learning Remote. Sen. Land. Res. 29 50-56

[8] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition Preprint arXiv/1409.1556

[9] Hars W J and Venkateswaran N 2016 An Example-based Super-Resolution Algorithm for Multi-Spectral Remote Sensing Images Sci. World. J. 65 1-12

[10] LeCun Y, Bengio Y, Hinton G 2015 Deep learning Nature 7553 436-444

[11] Isola P, Zhu J Y, Zhou T H, Efros A 2017 Image-to-image translation with conditional adversarial networks Proceedings of the IEEE Conf. Com. Vision Pattern Reco. p 6401-6408

[12] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B 2014 Generative adversarial nets Adv. Neu. Info. Proc. Sys. p 2672-80

[13] Isola P, Zhu J Y, Zhou T H, Efros A 2017 Image-to-image translation with conditional adversarial networks Proceedings of the IEEE Conf. Com. Vision Pattern Reco. p 1125-34

[14] Nah S, Hyun Kim T, Mu Lee K 2017 Deep multi-scale convolutional neural network for dynamic scene deblurring Proceedings of the IEEE Conf. Com. Vision Pattern Reco. p 3883-91

[15] Lei S, Shi Z, Zou Z 2017 Super-resolution for remote sensing images via local–global combined network IEEE Geo. Remote. Sens. Let. 14 1243-47

[16] Xu S H, Mu X D, Chai D, Zhang X M 2018 Remote sensing image scene classification based on generative adversarial networks Remote. Sens. Let. 9 617-626

[17] Dosovitskiy A, Brox T 2016 Generating images with perceptual similarity metrics based on deep networks Adv. Neu. Info. Proc. Sys. p 658-666

[18] Ledig C, Theis L, Huszár F, Caballero J 2017 Photo-realistic single image super-resolution using a generative adversarial network Proceedings of the IEEE Conf. Com. Vision Pattern Reco. p 4681-90

[19] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B 2014 Generative adversarial nets Adv. Neu. Info. Proc. Sys. p 2672-80

[20] Keys R 1981 J. Cubic convolution interpolation for digital image processing IEEE Trans. Acou. Speech. Signal. Pro. 29 1153-60

[21] Yang J, Wright J, Huang T, Ma Y 2008 Image super-resolution as sparse representation of raw image patches IEEE Conf. Com. Vision Pattern Reco. p 1-8

[22] Li F, Jia X P, Fraser D, Lambert A 2009 Super resolution for remote sensing images based on a universal hidden Markov tree model IEEE Trans. Geo. Remote. Sens. 48 1270-78

[23] Pan Z X, Yu J, Huang H J, Hu S X 2013 Super-resolution based on compressive sensing and structural self-similarity for remote sensing images IEEE Trans. Geo. Rem. Sens. 51 4864-76

[24] Chavez-Roman H and Ponomaryov V 2014 Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation IEEE Trans. Geo. Rem. Sens. 11 1777-81

[25] Wang X, Yu K, Wu S X, Gu J J 2018 Esrgan: Enhanced super-resolution generative adversarial networks Proc. Euro. Conf. Comp. Vision (ECCV)
[26] Lei S, Shi Z, Zou Z 2017 Super-resolution for remote sensing images via local–global combined network IEEE Trans. Geo. Rem. Sens. 14 1243-47
[27] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift. Preprint arXiv/1502.03167
[28] Ronneberger O, Fischer P, Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation In MICCAI (Springer International Publishing) p 234-241
[29] Johnson J, Alahi A, Li F F 2016 Perceptual losses for real-time style transfer and super-resolution Euro. Conf. Comp. Vision. (Springer/Cham) p 694-711
[30] Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng A Y 2011 Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning (Granada) p 12-17
[31] Lecun Y, Bottou L, Bengio Y, Haffner P 1998 Gradient-based learning applied to document recognition Proce. IEEE 86 2278-2324
[32] Keys R 1981 Cubic convolution interpolation for digital image processing IEEE Trans. Acou. Speech. Signal. Pro. 29 1153-60
[33] Xu S H, Mu X D, Chai D, Zhang X M. Remote sensing image scene classification based on generative adversarial networks Rem. Sens. Let. 9 617-26