# LEARNING SPECTRAL AND SPATIAL FEATURES BASED ON GENERATIVE ADVERSARIAL NETWORK FOR HYPERSPECTRAL IMAGE SUPER-RESOLUTION

*Ruituo Jiang*[1,*], *Xu Li*[1,*], *Ang Gao*[1], *Lixin Li*[1], *Hongying Meng*[2], *Shigang Yue*[3], *Lei Zhang*[4,*]

[1] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China.
[2] Department of Electronic and Computer Engineering, Brunel University London, UB8 3PH UK.
[3] School of Computer Science, University of Lincoln, Lincoln, LN6 7TS UK.
[4] Shanghai Key Laboratory of Multidimensional Information Processing,
East China Normal University, Shanghai 200241, China.

**ABSTRACT**

Super-resolution (SR) of hyperspectral images (HSIs) aims to enhance the spatial/spectral resolution of hyperspectral imagery and the super-resolved results will benefit many remote sensing applications. A generative adversarial network for HSIs super-resolution (HSRGAN) is proposed in this paper. Specifically, HSRGAN constructs spectral and spatial blocks with residual network in generator to effectively learn spectral and spatial features from HSIs. Furthermore, a new loss function which combines the pixel-wise loss and adversarial loss together is designed to guide the generator to recover images approximating the original HSIs and with finer texture details. Quantitative and qualitative results demonstrate that the proposed HSRGAN is superior to the state of the art methods like SRCNN and SRGAN for HSIs spatial SR.

***Index Terms***— Hyperspectral images, super-resolution, generative adversarial network, residual network

## 1. INTRODUCTION

The development of hyperspectral technology has made the hyperspectral images irreplaceable in many civil and military fields such as environmental monitoring, geological exploration, agriculture, military applications, etc., which has profoundly affected the economic development, national defense construction and social life. However, due to the limitation in Signal-to-Noise Ratio (SNR) and time constraint, there is a tradeoff between spatial and spectral resolution in remote sensing. Consequently, hyperspectral sensing produces extremely high spectral resolution and low spatial resolution.

Single-image spatial SR is a signal processing technique, which can improve a low spatial resolution image to a high spatial resolution image without any other prior or auxiliary information. The basic method of SR is through a nonlinear interpolator, such as bilinear and bicubic interpolation [1], which directly exploits the information of neighboring pixels. However, these methods often lead to edge blur or ringing effect. Recently, deep learning based methods have been applied to the SR of color images and demonstrated great superiority. SR Convolutional Neural Network (SRCNN) [2] is a pioneering work for deep learning in SR reconstruction, which firstly uses bicubic interpolation to enlarge the low-resolution image to a target size and then fits the nonlinear mapping through a three-layer convolutional network. Efficient subpixel CNN [3] extracts features directly from a low-resolution image by convolutional layers and enlarges the image size by a sub-pixel convolutional layer. A generative adversarial network for super-resolution (SRGAN) [4] is proposed to reconstruct a more realistic image with finer texture details. Transferred generative adversarial network (TGAN) [5] is improved from SRGAN and trained in a transfer-learning fashion to cope with the insufficiency of remote sensing training dataset. However, HSIs have unique characteristics compared with nature images. The network pre-trained with non HSIs is not suitable or even misleading to learn useful features.

All of the above CNNs for the SR of color images can be directly applied to HSIs in a band-by-band or 3-band-group manner. However, spectral distortion is often induced in such straightforward extensions since the strong spectral correlation in contiguous bands is ignored. Therefore, in this paper, a new generative adversarial network for HSIs super-resolution (HSRGAN) is proposed which effectively extracts spectral and spatial features from hyperspectral data. The experimental results demonstrate that the proposed method makes improvement in terms of both the objective evaluation and the subjective perspective. The main contributions of this paper can be summarized as follows:

- For the first time, Generative Adversarial Network (GAN) is used for hyperspectral images SR.
- Spectral and spatial residual blocks are constructed in generator to explore both spectral correlation in adjacent band images and the spatial contexts between neighboring pixels so that the spectral distortion is alleviated.
- A new loss function which combines the pixel-wise loss and adversarial loss together is designed to guide the gener-

ator to reconstruct images approximating the original HSIs and with finer texture details.

## 2. METHODOLOGY

### 2.1. Adversarial network structure

The general idea of GAN is that it aims to train a generator to reconstruct high-resolution images for fooling a discriminator that is trained to distinguish generative images from real ones. The generator structure of HSRGAN is illustrated in Fig. 1. Spectral and spatial residual blocks are the core part of generator network. HSIs usually have hundreds of contiguous bands with abundant spectral signatures and spatial contexts. General 2D convolution that is straightly applied to HSIs in a band-by-band or 3-band-group manner will result in spectral distortion. Though ordinary 3D convolution operation can explore spectral and spatial features together so that the spectral distortion is suppressed, it is still hard to extract effective features from rich and redundant spectral signatures in HSIs. Therefore this special structure containing spectral and spatial residual blocks is designed to learn both spectral correlation in adjacent band images and the spatial contexts between neighboring pixels. The structure of basic residual block is shown in Fig. 2. It contains two convolutional layers followed by a Parametric Rectified Linear Unit (PReLU) layer as the activation function to adaptively learn the parameters of the rectifiers [6]. In spectral residual block, the convolutional kernel size is set as $1 \times 1 \times 9$ so that the parameters to be learned is a 1D vectors, which can be seen as a special case of 3D convolutional kernels. It benefits to extract spectral features from abundant and redundant spectral signatures in contiguous bands. In spatial residual block, the convolutional kernels with a size of $3 \times 3 \times d$ are applied to extract the spatial contexts between neighboring pixels and further fine-tune spectral distortion, where $d$ depends on the depth of input HSIs. Other convolutional layers in the generator network are set as $3 \times 3 \times 3$ as usual. All the convolution operation is more efficient because it is performed on low resolution images instead of interpolated images. In addition, padding is used to prevent shrink in the size of the image in these convolutional layers. Finally, a sub-pixel convolutional layer proposed by [3] is adopted to increase the resolution of the input images. The architecture of discriminator network is almost same as SRGAN except that 3D convolution operation is adopted rather than 2D convolution.

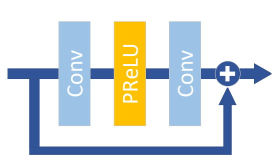



**Fig. 2**. The "Basic block" in the proposed model.

### 2.2. Loss function

The loss function designed in HSRGAN is to guide the generator to reconstruct images approximating the original HSIs and with finer texture details. It is defined as the sum of a pixel-wise loss and an adversarial loss.

Traditional pixel-wise loss Mean-Squared-Error (MSE) measured by $\mathcal{L}_2$-norm is much larger in the case of outliers compared to the least absolute deviations measured by $\mathcal{L}_1$-norm. As a consequence, MSE based loss function may try to adjust the model according to these outlier values. On the contrary, $\mathcal{L}_1$-norm based loss function is more robust to outliers, which is especially beneficial for training a network. In this paper, the $\mathcal{L}_1$-norm based loss function is adopted as the pixel-wise loss:

$$\mathcal{L}_1 = \frac{1}{HWD} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{D} \left| I_{i,j,k}^{HR} - G(I^{LR})_{i,j,k} \right| \quad (1)$$

where $I^{LR}$ represents a input low-resolution image and $I^{HR}$ represents a corresponding high-resolution image with the size of $H \times W \times D$, $G(I^{LR})$ represents the super-resolved result.

In standard GAN, such as SRGAN, the discriminator simply estimates the probability that one input image is real. However, the Relativistic average Discriminator (RaD) proposed by [7] is proven to be able to fix and improve standard GAN. It predicts the probability that the given real image $x_r$ is relatively more realistic than fake one $x_f$. This forms a class of models named Relativistic average GANs (RaGAN) that is adopted in this paper. Then the discriminator loss is defined as:

$$\mathcal{L}_D^{RaGAN} = -E_{x_r}[\log(\bar{D}(x_r))] - E_{x_f}[\log(1 - \bar{D}(x_f))] \quad (2)$$

The adversarial loss for generator is in a symmetrical form:

$$\mathcal{L}_G^{RaGAN} = -E_{x_r}[\log(1 - \bar{D}(x_r))] - E_{x_f}[\log(\bar{D}(x_f))] \quad (3)$$

where

$$\bar{D}(x) = \begin{cases} sigmoid(C(x) - E_{x_f} C(x_f)) \; if \; x \; is \; real \\ sigmoid(C(x) - E_{x_r} C(x_r)) \; if \; x \; is \; fake \end{cases} \quad (4)$$

where $C(x)$ is the non-transformed discriminator output and $E_x[\cdot]$ represents the operation of taking average for all fake or real data. $x_f = G(I^{LR})$ and $x_r$ is $I^{HR}$. It can be seen that the $L_G^{RaGAN}$ is influenced by gradients from both $x_f$ and $x_r$. In other words, our generator can be guided by both super-resolved image and real image in adversarial training, while only super-resolved image plays a part in SRGAN.

Overall, the total loss for the generator is formulated as:

$$\mathcal{L}_G = L_1 + L_G^{RaGAN} \quad (5)$$

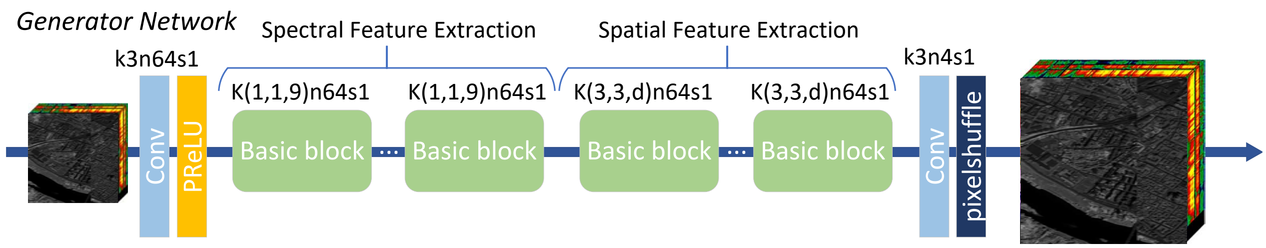



**Fig. 1**. The generator network of the proposed HSRGAN.

where $L_1$ is the pixel-wise loss that calculates the $\mathcal{L}_1$-norm distance between the super-resolved image $G(I^{LR})$ and the ground-truth $I^{HR}$ to make them more like. $L_G^{RaGAN}$ is the adversarial loss for generator resulting more realistic image with finer texture details.

## 3. EXPERIMENTS

### 3.1. Dataset and training details

In this experiment, Pavia Center dataset acquired by famous hyperspectral sensor ROSIS is selected. It has 102 spectral bands and contains $1096 \times 715$ effective pixels. For quantitative assessment, the original dataset is used as the ground-truth $I^{HR}$. Then the input $I^{LR}$ is simulated from $I^{HR}$ by using Gaussian low-pass spatial filtering with a down-sampled factor of 2. In this dataset, a $150 \times 150$ sub-region is selected to validate the performance of our proposed model, while the remaining pixels are used for training. In order to generate inputs for the proposed HSRGAN, sub-images with a size of $64 \times 64 \times 102$ are cropped by using a $64 \times 64$ spatial window sliding on the simulated $I^{LR}$. Their corresponding $128 \times 128 \times 102$ $I^{HR}$ are also cropped as ground-truth.

The implementation is based on Pytorch framework and accelerated with a single NVIDIA 1080Ti GPU. He initialization [6] and Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ are employed for networks, where Back Propagation (BP) strategy is adopted to alternately update the generator and discriminator network with a learning rate of 0.0002 until the model converges.

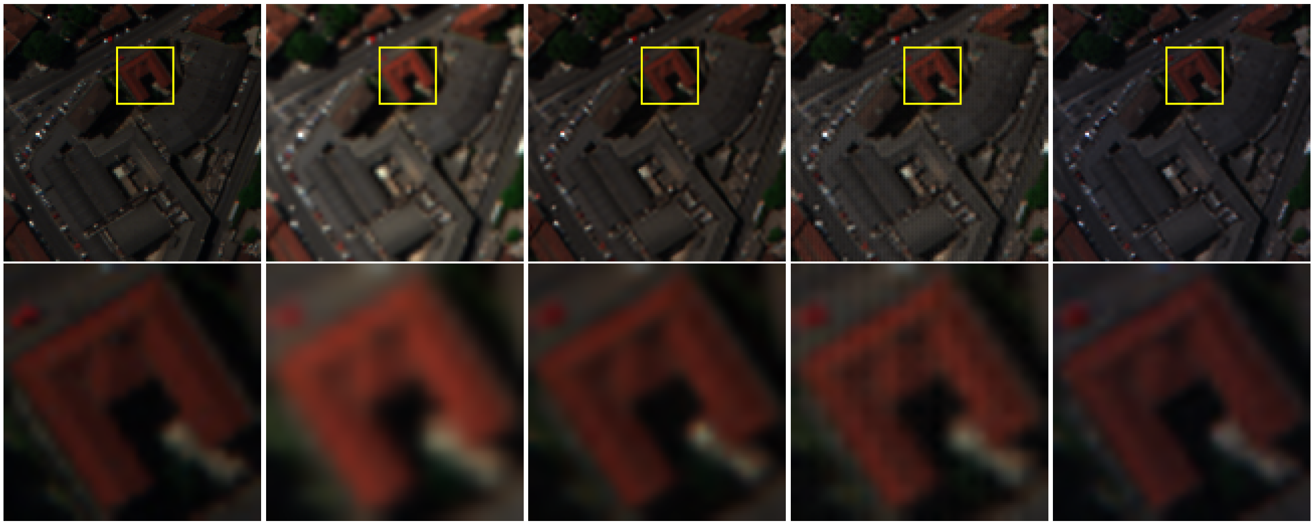
### 3.2. Comparison

The performance of the proposed HSRGAN is evaluated on the Pavia Center dataset and compared with other methods including bicubic interpolation, SRCNN and SRGAN. In order to evaluate the performance of HSRGAN quantitatively, three metrics are used to evaluate the quality of the super-resolved results, including mean peak signal-to-noise ratio (MPSNR), mean structural similarity index (MSSIM), and spectral angle mapper (SAM). The comparative results are listed in Table 1, in which the best values are marked in bold. As one can see, The performance of SRGAN with perceptual loss is slightly worse than SRCNN with MSE loss and outperforms traditional bicubic interpolation on objective measures. However, the proposed HSRGAN achieves the best performance among all the compared methods, with the highest MPSNR and MSSIM values and lowest SAM values. Specifically, the MPSNR of HSRGAN is 1.967dB, 1.548dB and 3.781dB higher than SRGAN, SRCNN and Bicubic respectively. The MSSIM of HSRGAN is 0.011, 0.001 and 0.025 higher than SRGAN, SRCNN and Bicubic respectively. The SAM of HSRGAN is 0.204, 0.09 and 0.324 lower than SRGAN, SRCNN and Bicubic respectively.

**Table 1**. Comparative results of different methods over Pavia Centre dataset. For MPSNR and MSSIM, the bigger, the better. For SAM, the smaller, the better.

| Dataset | Algorithm (Ideal Values) | MPSNR ($+\infty$) | MSSIM (1) | SAM (0) |
|---|---|---|---|---|
| Pavia Centre | Bicubic | 31.615 | 0.937 | 4.376 |
| | SRCNN | 33.848 | 0.961 | 4.142 |
| | SRGAN | 33.429 | 0.951 | 4.256 |
| | HSRGAN | **35.396** | **0.962** | **4.052** |

To facilitate the comparison of subjective quality, the super-resolved results are shown in Fig. 3 and a subscene in yellow square are zoomed up for better observing. It is obviously seen that the experimental results of the classical super-resolution method with bicubic interpolation has severe spectral distortion and edge blurring. SRCNN mitigates spectral distortion compared to traditional interpolation, but there exists excessive texture smoothing. As the first super-resolution method based on GAN, SRGAN produces sharper results, which alleviates the over smoothing problem of SRCNN, however, it fails to reconstruct texture details when directly applied on HSIs because the network was originally designed for natural images. Compared to these methods, the proposed HSRGAN reconstructs clear and sharper results in terms of both overall concept style and texture details, greatly improving the quality of the super-resolved results.



(a) Ground-Truth (b) Bicubic (c) SRCNN (d) SRGAN (e) HSRGAN

**Fig. 3**. Super-resolved results reconstructed for Pavia Center dataset by different methods. Band 15, 30, 60 are displayed as blue, green, and red respectively to show the composite color images. In order to observe more clearly, the part of each result with yellow square is zoomed up and shown correspondingly under the whole result.

## 4. CONCLUSION

This paper presents a novel super-resolution method for hyperspectral images which considers learning both spectral and spatial features based on generative adversarial network. Experimental results on Pavia Center dataset demonstrate that the proposed method can produce high quality super-resolved results and outperforms the state-of-the-art methods.

## 5. ACKNOWLEDGEMENTS

This work is partially supported by the European Union Horizon 2020-ULTRACEPT (778062), Shanghai Aerospace Science and Technology Innovation Fund (SAST2017049 and SAST2016034), the Seed Foundation of Innovation and Creation for Graduate Students in NPU(ZZ2019164), the National Key R&D Program of China (No.2018YFE0101000) and Open Fund of Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University(Grant No.2019KEY001).

## 6. REFERENCES

[1] Xin Li and Michael T Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–1527, 2001.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*, Zurich, Switzerland, 2014, Springer, pp. 184–199.

[3] Wenzhe Shi, Jose Caballero, Ferenc Huszár, et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 1874–1883.

[4] Christian Ledig, Lucas Theis, Ferenc Huszár, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 105–114.

[5] Wen Ma, Zongxu Pan, Jiayi Guo, and Bin Lei, "Super-resolution of remote sensing images based on transferred generative adversarial network," in *2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 2018, IEEE, pp. 1148–1151.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 1026–1034.

[7] Alexia Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.