

# HYPERSPECTRAL IMAGE SUPER-RESOLUTION USING GENERATIVE ADVERSARIAL NETWORK AND RESIDUAL LEARNING

Qian Huang<sup>1</sup>, Wei Li<sup>1,\*</sup>, Ting Hu<sup>2</sup>, and Ran Tao<sup>2</sup>

<sup>1</sup> College of Information Science & Technology, Beijing University of Chemical Technology

<sup>2</sup> School of Information and Electronics, Beijing Institute Technology, Beijing China

## ABSTRACT

Due to the limitation of image acquisition, hyperspectral remote sensing imagery is hard to reflect in both high spatial and spectral resolutions. Super-resolution (SR) is a technique which can improve the spatial resolution. Inspired by recent achievements in deep convolutional neural network (CNN) and generative adversarial network (GAN), a GAN based framework is proposed for hyperspectral image super-resolution. In the proposed method, residual learning is used to obtain a high metrics and spectral fidelity, and a shorter connection is set between the input layer and output layer. The gradient features from low-resolution (LR) image to high-resolution (HR) are utilized as auxiliary information to assist deep CNN to carry out counter training with discriminator. Experimental results demonstrate that the proposed SR algorithm achieves superior performance in spectral fidelity and spatial resolution compared with baseline methods.

**Index Terms**— Hyperspectral Imagery, Super-Resolution, Generative Adversarial Network.

## 1. INTRODUCTION

Hyperspectral sensor can capture wealthy spectral information with a larger number of narrow band wavelengths, which can potentially enhance many computer vision tasks, such as classification [1], target detection [2] and so on. However, due to the technical limitations of imaging equipment, hyperspectral images are usually unable to obtain high spatial and spectral resolution at the same time. The low spatial resolution results in the mixing of pixels, which is not conducive to the further promotion of hyperspectral image analysis. Therefore, it is necessary to develop effective super-resolution (SR) algorithm to obtain both high spatial and spectral resolution, which is beneficial to further image interpretation to improve its performance in the vision tasks.

Several techniques have been developed to obtain both spatial and spectral resolution in hyperspectral images. In

[3], multispectral images and traditional spectral unmixing analysis were considered to reconstruct hyperspectral images. And the other classical method is Pansharpening, which employs panchromatic image to provide the high spatial resolution information [4]. However, such spatial registered auxiliary sources are rare or sometimes impossible. Furthermore, many traditional methods are usually based on sparsity representation [5–7], while involving high computational cost. Thus, it is necessary to develop more efficient SR algorithm that may require no additional priors.

Recently, deep learning has brought revolutionary achievements in many applications [8]. Specially, SR for natural color images can be seen as a image transform task, which simplifies the SR algorithm framework and achieves a big success. In [9], deep convolutional neural network (CNN) has been used to realize an end-to-end learning from low-resolution (LR) image to high-resolution (HR), and the HR hyperspectral images can be inferred by LR hyperspectral images and CNN directly. In [10], authors proposed a new SRCNN framework by using both the spatial context between neighboring pixels and the spectral correlation in adjacent band images. However, although the SRCNN can obtain well quality-measure metrics, such as PSNR, SSIM or MRAE and so on, the perceptual quality is hard to get well. Therefore, SR generative adversarial network (SRGAN) and perceptual loss have been proposed to solve this problem [11]. Nevertheless, SRGAN can provide a high perceptual quality, but its PSNR metric is not as good as SRCNN or even worse than the typical bicubic interpolation. That is alright for natural color images SR, but for hyperspectral image SR, a lower PSNR metric usually means a distortion of spectral information.

Inspired by above observation, an novel GAN based framework is proposed for hyperspectral image SR, named as modified SRResGAN (MSRResGAN). In the proposed method, the SRCNN [10] is extended to make it deeper, with adding residual blocks; the generating network is similar to the SRGAN [11], but the pixel shuffle layers are removed and deep network is designed to learn the residual image; the deeper layer and residual learning can lead to higher metrics and satisfied spectral fidelity. In terms of the improvement of spatial perception quality, VGG loss is no longer used as

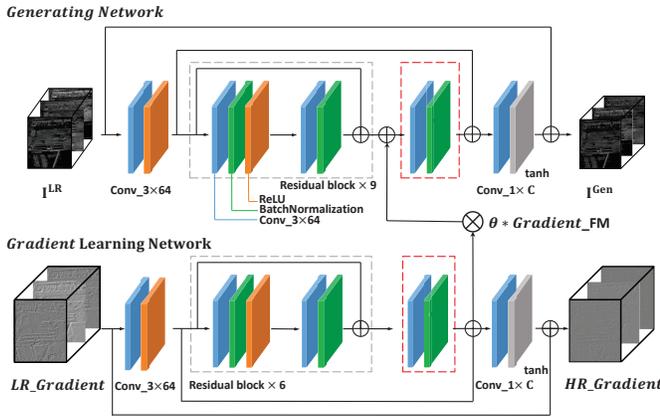
This research was supported by the National Natural Science Foundation of China (61571033,91638201), Beijing Natural Science Foundation (4172043), and partly by the Research Fund for Basic Researches in Central Universities under Grant PYBZ1831. (Email: liwei089@ieec.org).

perceptual loss to assist network recovery details, because it may produce some fake details in hyperspectral images. Instead, a gradient learning network is developed to train against the discriminator by transmitting the gradient features to the generating network and helping it recover details. Experiments on visible near infrared (VNIR) and short wave infrared (SWIR) hyperspectral remote sensing images are conducted the effectiveness of the proposed MSRResGAN.

## 2. PROPOSED SUPER-RESOLUTION METHOD

The proposed hyperspectral image SR method consists of two parts: generating network and discriminating network. The generating network is responsible for reconstructing a bicubic-interpolated LR hyperspectral image into a HR one, and the discriminating network makes a discrimination between the reconstructed image and the real image.

### 2.1. Generating Network in Proposed MSRResGAN



**Fig. 1.** The proposed MSRResGAN framework including generating network and gradient learning network.

Let  $I^{LR} \in \mathbb{R}^{h \times w \times c}$  and  $I^{HR} \in \mathbb{R}^{kh \times kw \times c}$  be the LR and HR hyperspectral cube, where  $k$  represent the upscale factor,  $h$ ,  $w$  and  $c$  represent the size of height, weight and bands of the image, respectively. The observed image  $I^{LR}$  can be seen as a blurred and downsample version of the  $I^{HR}$  [12]. In other words, the high-frequency components of HR image has been smoothed out by the blurr filtering. Thus, it can suppose that the low-frequency components is similar between the LR and HR image, the high-frequency components  $I^{res} \in \mathbb{R}^{kh \times kw \times c}$  is what LR image lacks can be obtained by subtracting LR images from HR image

$$I^{res} = I^{HR} - I^{LR} \uparrow k \quad (1)$$

where  $\uparrow$  represents the upsampling operation. Therefore, if the generating network can generate the residual image  $I^{res}$ , the HR image can be calculated easily.

The generating network is illustrated in Fig. 1. The network takes the bicubic interpolated LR hyperspectral image as input, but the remote sensing image usually has a large shape, we input the image into the model patch by patch and join output together finally. A very deep CNN is employed, which consists of 21 layers where except the first and the last are of the same configuration: 64 convolutional kernels of the size  $3 \times 3 \times 64$ , *BatchNormalization* [13]. Except the last layer, the activation function in other layres is *Relu*. The architecture contains 9 residual blocks and two shorter connections. The residual block enables networks to learn high level dimension features more effectively and reduce the possibility of network degradation due to deep layer [14]. The inner shortcut passes the features to the last feature extracting layer, which can help low level and high level features fusion.

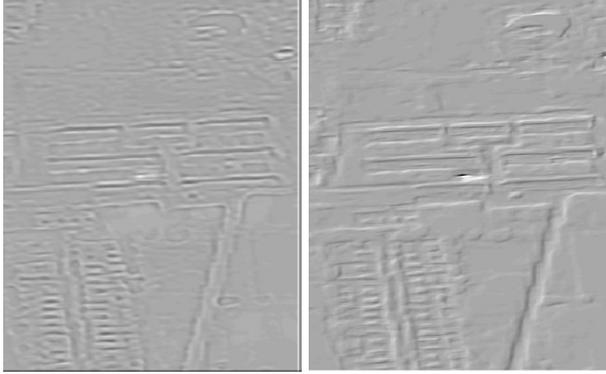
Particularly, in order to lead generating network to generate the residual image  $I^{res}$ , outermost shorter connection reuse the raw input image  $I^{LR}$  in output layer. Let  $X \in \mathbb{R}^{kh \times kw \times c}$  and  $I^{Gen}$  be the interpolated image from  $I^{LR}$  and generated image, we use the mean squared error (MSE) between actual output  $I^{Gen}$  and ideal output  $I^{HR}$  as the first cost function to optimize parameters in the generating network, which is

$$\begin{aligned} L_{mse} &= \|I^{HR} - I^{Gen}\|_2^2 \\ &= \|I^{HR} - (X + G(X))\|_2^2 \\ &= \|I^{res} - G(X)\|_2^2 \end{aligned} \quad (2)$$

where  $G(X)$  is active output of the last convolutional layer in network. Thus, through a shorter connection from the input layer to the output layer, the MSE cost function will leads the network to learn to generate the residual image  $I^{res}$ . The second cost function will be introduced in next subsection.

Furthermore, in the SRGAN [11], although the VGG loss can help to further improve the quality of perception of the reconstruction image, but it is not conducive to hyperspectral image SR. Firstly, VGG has a fixed size of input channels, which is not conducive to multi-channel hyperspectral image restoration. Secondly, VGG loss and pixel shuffle upsampling layer sometimes makes the image produce some fake detail information, which is not suitable for remote sensing image. Instead, gradient information is employed to assist the network to further improve SR performance. Compare with VGG loss, gradient learning focuses on the high-frequency information, rather than the abstract and high-level features (e.g. semantic features) which may lead to generate the fake detail.

The gradient learning network is illustrated in Fig. 1, where the network is designed to learn a projection between the LR image gradient and HR image gradient. As shown in Fig. 2, it can be seen that the learned gradient image is more clear in its details by comparing the learned residual image. The last fusion features are shared in gradient learning network to the generating network, using a set of trainable



(a) Learned residual image (b) Learned gradient image

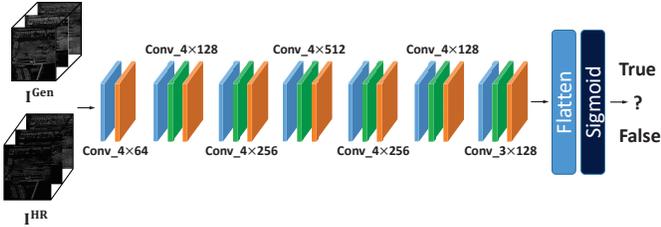
**Fig. 2.** An example of contrast between the learned residual image and the learned gradient image.

parameters to control the gradient features influence

$$F_i = F_i^{res} + \theta_i * F_i^{grad} \quad (3)$$

where  $F_i$  is the  $i$ th feature map, total 64 feature maps. The trainable parameters  $\theta$  can be defined by a set of convolutional kernels with size of  $1 \times 1 \times 64$ .

## 2.2. Discriminating Network in Proposed MSRResGAN



**Fig. 3.** The discriminating network architecture in the proposed MSRResGAN.

Inspired by the [11], the configuration of the discriminating network is outlined as shown in Fig. 3. The purpose of the adversarial modeling framework is to solve the adversarial min-max problem

$$\min_G \max_D \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})} [\log D(I^{HR})] + \mathbb{E}_{X \sim p_G(X)} [\log(1 - D(G(X) + X))], \quad (4)$$

where  $D(X)$  is output of the discriminating network. It should be able to accurately determine whether the input images is the real image or the generated image, while the generator should be able to generate the real image as much as possible. Therefore, for paired real HR image  $I^{HR}$  and interpolated image  $X$ , discriminating network is trained to maximize the probability of assigning the correct label both

$I^{HR}$  and  $I^{Gen}$ , which is the same as minimizing the following loss

$$L_D = \log(1 - D(I^{HR})) + \log(0 - D(G(X) + X)). \quad (5)$$

By cheating the discriminating network, generating net is trained to minimize  $\log(1 - D(G(X) + X))$ , and it is the second cost function to optimize parameters in the generating network. The complete loss function for generating network is

$$L_G = \|I^{res} - G(X)\|_2^2 + \lambda * \log(1 - D(G(X) + X)), \quad (6)$$

where  $\lambda$  is balancing parameter. The two networks are trained alternately. When the loss  $L_D$  and  $L_G$  converge, it can be seen that there is an equilibrium between the two network and the  $I^{Gen}$  is highly similar to real HR image.

## 3. EXPERIMENTAL ANALYSIS

Two paired of HR/LR hyperspectral datasets with CASI sensor are employed in the following experiments. There are acquired over the Zhangye city in China while different in range of spectrum. One is in the range of VNIR and the other is in the range of SWIR. The HR VNIR and SWIR dataset has  $1m$  spatial resolution, and contains size of  $6000 \times 1000 \times 48$  and  $6000 \times 1000 \times 87$ , respectively. And the LR VNIR and SWIR dataset has  $5m$  spatial resolution.

All the experiments are implemented with Matlab, Python, Tensorlayer[15] and Tensorflow<sup>1</sup>. We cut the dataset in two parts with the same size of  $3000 \times 1000 \times c$ , one is used for training and the other one is used for testing. The training dataset is divided into 690 patches with size of  $64 \times 64 \times c$  ( $c$  means the spectral bands). During training, *Adam* is used to optimize the network parameters, and the base learning rate is 0.001, learning decay is 0.5. Firstly, gradient learning network is trained, and the  $\lambda$  in  $L_G$  is set as  $\lambda = 0.0001$ .

The baseline methods for comparison with the proposed MSRResGAN are Bicubic SR, SRCNN [10], SRGAN [11]. For fair comparison, we also include modified SRCNN with residual learning (denoted as MSRResNet). Four metrics are used to quantitatively evaluate the performance of SR of hyperspectral images: the mean of relative absolute error (MRAE), the spectral information divergence (SID), the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). PSNR (dB) and SSIM are the widely used to evaluate the performance of reconstruction, the higher the value, the better the effect. The SID is an information theoretic measure which assesses similarity making use of the probabilistic discrepancy between the spectra under consideration [16], and the lower the value, the better the spectral fidelity.

Table 1 provides the performance of comparison between different algorithms. Meanwhile, the experimental results of

<sup>1</sup><https://www.tensorflow.org>

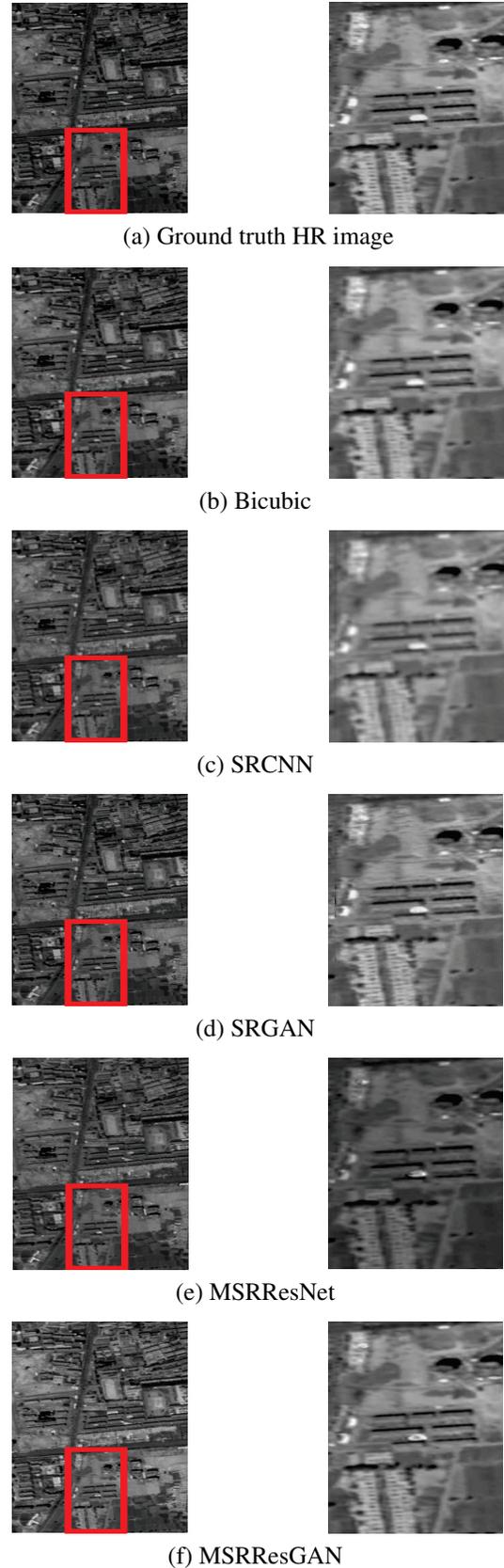
all the algorithms for SR of SWIR image are shown in Fig. 4. As enlarged in the red rectangular, although SRCNN is superior to SRGAN in terms of indicators, SRCNN does not perform well in details recovery. On the contrary, SRGAN can restore the details better but it has lower metrics. Meanwhile, it can be found some fake details in the SRGAN reconstructed image by carefully comparing the ground truth image. Finally, it can be seen that the MSRResNet has significant improvement in the metrics. After joining the gradient learning network and adversarial learning, it can be found that the proposed MSRResGAN has further obvious improvement in the perceived quality and spectral fidelity when compared to MSRResNet. In general, the experiments results confirm the proposed MSRResGAN obtains the best balance between perceptual quality and spectral fidelity.

**Table 1.** Performance Comparison of Super-Resolution on Real SWIR and VNIR Hyperspectral Image Data.

SWIR Data				
metrics	MRAE	SID	PSNR(dB)	SSIM
<b>Bicubic</b>	0.1256	31.4386	43.7679	0.9016
<b>SRCNN</b>	0.0812	10.6691	49.4285	0.9872
<b>SRGAN</b>	0.0677	10.8887	48.9640	0.9832
<b>MSRResNet</b>	0.0492	4.6164	51.6164	0.9953
<b>MSRResGAN</b>	<b>0.0372</b>	<b>2.8824</b>	<b>52.6751</b>	<b>0.9959</b>
VNIR Data				
metrics	MRAE	SID	PSNR(dB)	SSIM
<b>Bicubic</b>	0.1661	27.3275	46.5987	0.9929
<b>SRCNN</b>	0.1388	21.7286	47.3267	<b>0.9952</b>
<b>SRGAN</b>	0.2385	40.5758	44.9196	0.9567
<b>MSRResNet</b>	0.1257	18.3422	47.9100	0.9949
<b>MSRResGAN</b>	<b>0.1205</b>	<b>17.2523</b>	<b>48.1036</b>	0.9949

#### 4. CONCLUSIONS

In this paper, a GAN based framework was proposed for hyperspectral remote sensing image super-resolution. The proposed method employed residual learning to learn the difference between LR image and HR image, which makes it easier to learn effective features and obtain higher metrics, resulting in a satisfied spectral fidelity. Furthermore, gradient learning network was utilized to provide gradient features for generating network and adversarial learning strategy, which can further improve the performance of deep network and perceptual quality of generated image. Experimental results on two real hyperspectral datasets demonstrated that the proposed MSRResGAN could obtain superior performance with both spatial resolution and spectral fidelity when compared with state-of-the-art methods.



**Fig. 4.** Illustration of 40-th band in SWIR hyperspectral image on experimental results of different algorithms.

## 5. REFERENCES

- [1] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [2] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 597–601, 2017.
- [3] Mohamed Amine Bendoumi, Mingyi He, and Shaohui Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6574–6583, 2014.
- [4] Laetitia Loncan, Luis B Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al., "Hyperspectral pansharpening: A review," *arXiv preprint arXiv:1504.04531*, 2015.
- [5] Jakub Bieniarz, Rupert Müller, Xiao Xiang Zhu, and Peter Reinartz, "Hyperspectral image resolution enhancement based on joint sparsity spectral unmixing," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. IEEE, 2014, pp. 2645–2648.
- [6] Elham Kordi Ghasrodashti, Azam Karami, Rob Heylen, and Paul Scheunders, "Spatial resolution enhancement of hyperspectral images using spectral unmixing and bayesian sparse representation," *Remote Sensing*, vol. 9, no. 6, pp. 541, 2017.
- [7] Chimam Kwan, Bence Budavari, Minh Dao, and Jin Zhou, "New sparsity based pansharpening algorithms for hyperspectral images," in *Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017 IEEE 8th Annual*. IEEE, 2017, pp. 88–93.
- [8] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 27, no. 6, pp. 2623–2634, 2018.
- [9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [10] Shaohui Mei, Xin Yuan, Jingyu Ji, Shuai Wan, Junhui Hou, and Qian Du, "Hyperspectral image super-resolution via convolutional neural network," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4297–4301.
- [11] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network.," in *CVPR*, 2017, vol. 2, p. 4.
- [12] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [13] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [15] Hao Dong, Akara Supratak, Luo Mai, Fangde Liu, Axel Oehmichen, Simiao Yu, and Yike Guo, "TensorLayer: A Versatile Library for Efficient Deep Learning Development," *ACM Multimedia*, 2017.
- [16] Chein-I Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1927–1932, 2000.