LEARNING SPATIAL AND SPECTRAL FEATURES VIA 2D-1D GENERATIVE ADVERSARIAL NETWORK FOR HYPERSPECTRAL IMAGE SUPER-RESOLUTION

Ruituo Jiang[1,*], *Xu Li*[1,*], *Shaohui Mei*[1], *Lixin Li*[1], *Shigang Yue*[2], *Lei Zhang*[3,*]

[1] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China.
[2] School of Computer Science, University of Lincoln, Lincoln, LN6 7TS UK.
[3] East China Normal University, Shanghai 200241, China.

ABSTRACT

Three-dimensional (3D) convolutional networks have been proven to be able to explore spatial context and spectral information simultaneously for super-resolution (SR). However, such kind of network can't be practically designed very 'deep' due to the long training time and GPU memory limitations involved in 3D convolution. Instead, in this paper, spatial context and spectral information in hyperspectral images (HSIs) are explored using Two-dimensional (2D) and One-dimenional (1D) convolution, separately. Therefore, a novel 2D-1D generative adversarial network architecture (2D-1D-HSRGAN) is proposed for SR of HSIs. Specifically, the generator network consists of a spatial network and a spectral network, in which spatial network is trained with the least absolute deviations loss function to explore spatial context by 2D convolution and spectral network is trained with the spectral angle mapper (SAM) loss function to extract spectral information by 1D convolution. Experimental results over two real HSIs demonstrate that the proposed 2D-1D-HSRGAN clearly outperforms several state-of-the-art algorithms.

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

1. INTRODUCTION

Single-image spatial SR is a signal processing technique that can improve a low spatial resolution image to a high spatial resolution image without any other prior or auxiliary information. Similarly, SR of HSIs enhances the spatial resolution of hyperspectral imagery and the super-resolved results will benefit many remote sensing applications, such as classification, target detection, and identification, etc.

Recently, deep learning based SR methods have been applied to the natural color images and demonstrated to be of great superiority. SR Convolutional Neural Network (SRCNN) [1] 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 sub-pixel CNN [2] extracts features directly from a low-resolution image by convolutional layers and enlarges the image size by a sub-pixel convolutional layer. The Dense Convolutional Network (DenseNet) [3], which concatenates features of all layers by feeding the features of each layer to all subsequent layers in a dense block, has also been used for SR problem [4]. A generative adversarial network for super-resolution (SRGAN) [5] is proposed to reconstruct a more realistic image with finer texture details. While an enhanced super-resolution generative adversarial network (ESRGAN) [6] improves SRGAN from network architecture, adversarial loss and perceptual loss so as to achieve consistently better visual quality. All of these CNNs for the SR of color images can be directly applied to HSIs in a band-by-band or 3-band-group manner. Inevitably, obvious spectral distortions are often induced in such straightforward extensions since the strong spectral correlation existed in contiguous bands is ignored. Therefore, a 3D full CNN (3D-FCNN) [7] is constructed to extract the spatial and spectral information jointly by 3D convolution. However, it is still hard to extract effective features from rich and redundant spectral signatures in HSIs by ordinary 3D convolution operation, though the spectral distortion is suppressed. And also in practice, the networks with 3D convolution can't be designed very deep because of the long training time and GPU memory limitations. Consequently, a novel 2D-1D generative adversarial network (GAN) [8] architecture for hyperspectral images super-resolution (2D-1D-HSRGAN) is proposed. Specifically, spatial and spectral features are explored by 2D and 1D convolution separatively, being more effective than vanilla 3D convolution. The experimental results demonstrate that the proposed method makes improvement in terms of both the objective evaluation and the subjective perspective. In summary, the main contributions of this work can be summarized as follows:

- A novel 2D-1D GAN is proposed for SR of HSIs which consists of spatial and spectral networks in generator to effectively learn spatial and spectral features from HSIs.
- Spatial network is trained with the least absolute deviations loss function to explore spatial context by 2D convolution and spectral network is trained with the spectral angle mapper (SAM) loss function to extract spectral information by

1D convolution.

2. METHODOLOGY

2.1. Adversarial network structure

The general idea of GAN in SR task is that it aims to train a generator to reconstruct high-resolution images from low-resolution images for fooling a discriminator that is trained to distinguish super-resolved images from real ones. The generator structure of the proposed 2D-1D-HSRGAN is illustrated in Fig. 1. It contains a spatial network with 2D convolution for spatial feature extraction and a spectral network with 1D convolution for spectral reconstruction. In addition, padding is used to prevent shrink in the size of the image in all convolutional layers in generator.

The core part of the spatial network in the proposed GAN is basic blocks with identical layout. This basic block combines multi-level residual network and dense connections as shown in Fig. 2 which is inspired by [6]. Specifically, it has a residual-in-residual structure, where residual learning is used in different levels and the dense connections are used to improve network capacity. The basic block constructs two dense blocks in a residual manner and each dense block contains three convolutional layers followed by LeakyReLU activation (α = 0.2) [10]. Each convolutional layer in dense block has access to all the subsequent layers and passes on information that needs to be preserved. One the whole, in spatial network, the first two convolutional layers are used to extract low-level features and reduce dimensionality, then the special basic blocks are adopted to further extract high-level features, after that low-level features and high-level features are fused in a convolutional layer. Finally, a sub-pixel convolutional layer proposed by [2] is adopted to increase the resolution of the input images. After spatial features extraction, a spectral network which contains three 1D convolutional layers is constructed for spectral reconstruction.

The architecture of discriminator is almost the same as that in [6] except that the layers before the third convolutional layer are removed. In standard GAN, the discriminator simply estimates the probability that one input image is real. While in Relativistic average Discriminator (RaD) proposed by [11], it predicts the probability that the given real image x_r is relatively more realistic than the fake one x_f. The RaD

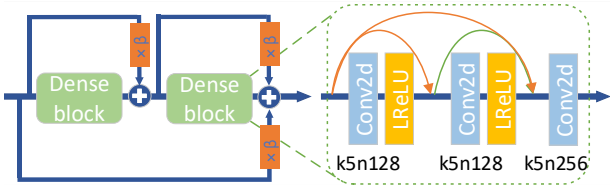


Fig. 2. The 'Basic block' in the 'Spatial network' of the proposed model and β is the residual scaling parameter [9] of 0.2.

is an improved version of discriminator in standard GAN and thus it is adopted in this paper. Consequently the discriminator loss is defined as:

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

in which

$$\bar{D}(x) = \begin{cases} sigmoid(C(x) - E_{x_f}C(x_f)) \text{ if x is real} \\ sigmoid(C(x) - E_{x_r}C(x_r)) \text{ if x is fake} \end{cases} \quad (2)$$

where $C(x)$ is the non-transformed discriminator output and $E_x[\cdot]$ represents the operation of taking average. $x_f = G(I^{LR})$ represents the super-resolved HSI and I^{LR} is an input low-resolution HSI, x_r is I^{HR} which represents a corresponding high-resolution HSI of I^{LR}. The adversarial loss for generator is in a symmetrical form:

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

It is observed that the proposed generator is guided by both super-resolved HSI and ground-truth HSI in adversarial training while only super-resolved image plays a part in standard GAN.

2.2. Loss function

In the training process, the spatial network and spectral network are first trained separately and then fine-tuned in the proposed GAN framework. As a result, the training process of the proposed 2D-1D-HSRGAN is divided into three stages: 1). training spatial network with the least absolute deviations loss to explore spatial context, 2). training spectral network with the SAM loss to extract spectral information, 3). training the whole GAN with a new loss that is defined as the summation of the least absolute deviations loss, SAM loss and adversarial loss.

The least absolute deviations measured by ℓ_1-norm is more robust to outliers than traditional pixel-wise loss Mean-Squared-Error (MSE) measured by ℓ_2-norm. Therefore, the ℓ_1-norm based loss function is adopted to train the spatial network of GAN for SR:

$$\mathcal{L}_1 = \frac{1}{HWD}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{D}\left|I_{i,j,k}^{HR} - G(I^{LR})_{i,j,k}\right| \quad (4)$$

where H is the height of I^{HR}, W is the width of I^{HR} and D is the number of spectral bands.

The SAM loss is designed to minimize the spectral angle between the reconstructed spectra and its corresponding ground truth spectra, which is defined as:

$$\mathcal{L}_{SAM} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\arccos\left(\frac{\langle z_{i,j}, \hat{z}_{i,j}\rangle}{\|z_{i,j}\|_2\|\hat{z}_{i,j}\|_2}\right) \quad (5)$$

where $\hat{z}_{i,j}$ denotes the spectral vector at the i-th row and j-th column in super-resolved HSI and $z_{i,j}$ represents its corresponding ground truth spectral vector with the same spatial position, $\langle\cdot,\cdot\rangle$ denotes dot product of two vectors and $\|\cdot\|_2$ represents the l_2 norm of a vector.

Overall, the total loss to fine-tune the generator of the pro-

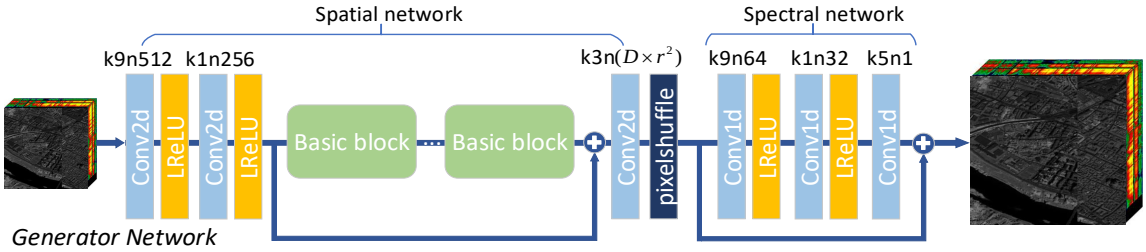


Fig. 1. The generator network architecture of the proposed 2D-1D-HSRGAN with corresponding kernel size (k), number of channels (n) indicated for each convolutional layer.

Algorithm 1 The training procedure of 2D-1D-HSRGAN

Initialize: He initialization, as described in [12], is employed to initialize all networks in 2D-1D-HSRGAN. m is the batch size.

Train the spatial network N_{spa} firstly
while the spatial network hasn't converged yet **do**
  sample a batch $\{x^{(i)}\}_{i=1}^{m}$ from I^{LR}.
  sample a batch $\{y^{(i)}\}_{i=1}^{m}$ from I^{HR}.
  update spatial network by minimizing the $\mathcal{L}_1$ loss between the pixel in $N_{spa}(x^{(i)})$ and $y^{(i)}$ according to Eq. (4).
end while
save the spatial network model.
Train the spectral network N_{spe} then
while the spectral network hasn't converged yet **do**
  sample a batch $\{x^{(i)}\}_{i=1}^{m}$ from I^{LR}.
  sample a batch $\{y^{(i)}\}_{i=1}^{m}$ from I^{HR}.
  $x^{(i)}$ is fed into the spatial network model to obtain $z^{(i)}$.
  update spectral network by minimizing the $\mathcal{L}_{SAM}$ loss between the spectra in $N_{spe}(z^{(i)})$ and $y^{(i)}$ according to Eq. (5).
end while
save the spectral network model.
Train the 2D-1D-HSRGAN finally
Initialize the generator with the spatial network model and the spectral network model that are well trained before.
while the generator G hasn't converged yet **do**
  sample a batch $\{x^{(i)}\}_{i=1}^{m}$ from I^{LR}.
  sample a batch $\{y^{(i)}\}_{i=1}^{m}$ from I^{HR}.
  update generator by minimizing the $\mathcal{L}_G$ loss between the $G(x^{(i)})$ and $y^{(i)}$ according to Eq. (6).
  sample a batch $\{\hat{y}^{(i)}\}_{i=1}^{m}$ from $G(I^{LR})$.
  sample a batch $\{y^{(i)}\}_{i=1}^{m}$ from I^{HR}.
  update discriminator by minimizing the $\mathcal{L}_D^{RaD}$ loss between the $\hat{y}^{(i)}$ and $y^{(i)}$ according to Eq. (1).
end while
save the generator model.

posed GAN for SR of HSIs is formulated as:

$$\mathcal{L}_G = \mathcal{L}_1 + \mathcal{L}_{SAM} + \mathcal{L}_G^{RaD} \quad (6)$$

Pre-training with $\mathcal{L}_1$ and $\mathcal{L}_{SAM}$ loss can avoid undesired local optima for the generator so that the GAN-based method will construct more visually pleasing results. In summary, the training details for the proposed 2D-1D-HSRGAN is presented in 3.2.

3. EXPERIMENTS

3.1. Dataset and training details

In this experiment, Pavia Center and Cuprite datasets are selected, which are acquired by two well-known hyperspectral sensors, namely ROSIS and AVIRIS. The Pavia Center dataset owns 102 spectral bands containing 1096×715 effective pixels, while the Cuprite has 202 effective spectral bands of 512×614 pixels. For quantitative assessment, these two original datasets are used as the ground-truth I^{HR}. The low-resolution HSIs I^{LR} are simulated from I^{HR} by using Gaussian low-pass spatial filtering with a down-sampled factor of 2 and variance of 0.72. For these two datasets, a 150×150 sub-region is selected to validate the performance of our proposed model, while the remaining pixels are used for training. The input sub-images with a size of $64 \times 64 \times D$ for the proposed model are cropped by using a 64×64 spatial window sliding on the simulated I^{LR}. Their corresponding $128 \times 128 \times D$ sub-images are also cropped from I^{HR} as ground-truth.

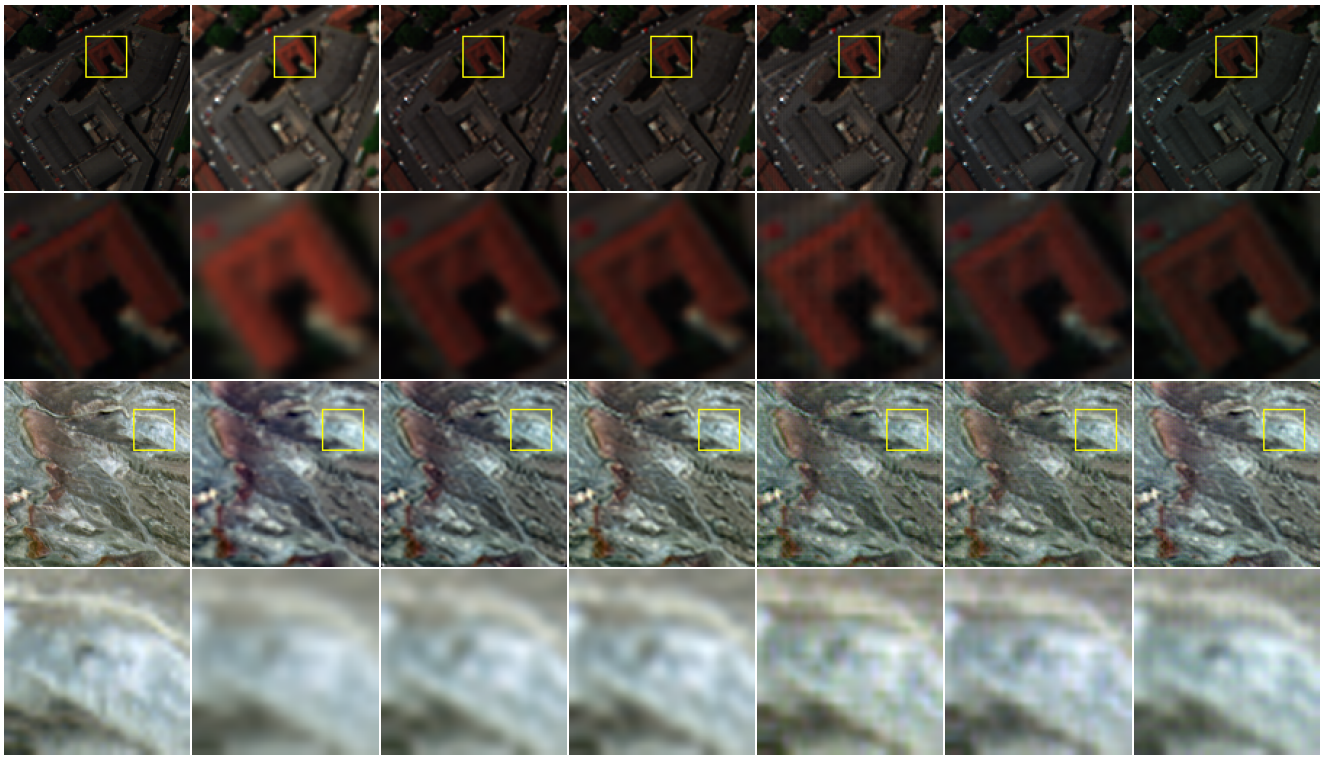
The implementation is based on Pytorch framework and accelerated with a single NVIDIA 1080Ti GPU. Adam optimizer [13] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ is employed for all the networks and the Back Propagation (BP) strategy is adopted to alternately update the generator and discriminator network with a learning rate of 0.0001 until the model converges.

3.2. Results and Discussions

The performance of the proposed 2D-1D-HSRGAN is evaluated by comparing with several state-of-the-art SR methods including bicubic interpolation [14], SRCNN [1], 3DFCN [7], SRGAN [5], and ESRGAN [6]. Three quantitative 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 experimental results of these SR algorithms over the two HSI datasets are listed in Table 1, in which the best values are marked in bold. Obviously, the proposed 2D-

Table 1. Comparative results of different methods over Pavia Centre dataset and Cuprite dataset.

Dataset	Algorithm	Bicubic	SRCNN	3DFCN	SRGAN	ESRGAN	Proposed
Pavia Centre	MPSNR ($+\infty$)	31.833	33.480	33.812	33.356	33.837	**35.622**
	MSSIM (1)	0.901	0.937	0.942	0.938	0.943	**0.960**
	SAM (0)	4.149	4.037	3.991	4.311	4.639	**3.825**
Cuprite	MPSNR ($+\infty$)	33.302	34.315	35.114	32.809	33.703	**35.872**
	MSSIM (1)	0.945	0.956	0.962	0.946	0.952	**0.972**
	SAM (0)	1.277	1.308	1.444	1.764	1.601	**0.595**



(a) Ground-Truth (b) Bicubic (c) SRCNN (d) 3DFCN (e) SRGAN (f) ESRGAN (g) Proposed

Fig. 3. Sample results reconstructed for Pavia Center and Cuprite datasets by different methods. Band 15, 30, 60 and band 5, 15, 25 are displayed as blue, green, 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 in row 2 and row 4, respectively.

1D-HSRGAN achieves the best performance over these two datasets among all the compared methods, with the highest MPSNR and MSSIM values and lowest SAM values.

Fig. 3 also presents sample results of these HSI SR algorithms. To facilitate the comparison of subjective quality, a subscene in yellow square are zoomed up for better observing. It is also confirmed that the proposed model greatly improve the quality of the super-resolved results compared to other methods. It reconstructs clear and sharper results in terms of both overall concept style and texture details, moreover, the spectral distortion is alleviated.

4. CONCLUSION

A novel 2D-1D generative adversarial network architecture is proposed for SR of HSIs, which can effectively explore spatial context by 2D convolution and extract spectral information by 1D convolution. Experimental results over Pavia Center and Cuprite datasets demonstrate that the proposed method can produce high quality super-resolved results and outperforms the state-of-the-art methods.

5. ACKNOWLEDGEMENTS

This research has received funding from the European Union Horizon 2020-ULTRACEPT (778062), the Seed Foundation of Innovation and Creation for Graduate Students in NPU(ZZ2019164), the National Natural Science Foundation of China (61671383), 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(No. 2019KEY001).

6. REFERENCES

[1] 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.

[2] 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.

[3] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261–2269.

[4] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao, "Image super-resolution using dense skip connections," in *IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 4809–4817.

[5] 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.

[6] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, and Xiaoou Tang, "Esrgan: Enhanced super-resolution generative adversarial networks," *arXiv: Computer Vision and Pattern Recognition*, 2018.

[7] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, pp. 1139, 2017.

[8] Ian J Goodfellow, Jean Pougetabadie, Mehdi Mirza, Bing Xu, David Wardefarley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio, "Generative adversarial nets," pp. 2672–2680, 2014.

[9] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.

[10] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng, "Rectifier nonlinearities improve neural network acoustic models," in *International Conference on Machine Learning*, Atlanta, GA, USA, 2013, pp. 3–9.

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

[12] 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.

[13] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, San Diego, CA, USA, 2015.

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