HIGH QUALITY REMOTE SENSING IMAGE SUPER RESOLUTION USING DEEP MEMORY CONNECTED NETWORK

Wen-Jia Xu, Guang-Luan Xu, Yang Wang, Dao-Yu Lin,Jiu-Niu Wang, Yi-Rong Wu

Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing, China
School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China
Email:gluanxu@mail.ie.ac.cn

ABSTRACT

The saptial resolution of remote sensing image is crucial for many applications such as target detection and image classification. Single image super resolution is an effective way to exceed the natural limitation of remote sensors. In this letter, we propose a new algorithm named deep memory connected network (DMCN) based on convolutional neural network to resonstruct high quality super resolution images. Inspired by memory mechanism of brain, we build local and global memory connections to combine image detail with environmental information. To further reduce parameters and ease time consuming, downsampling units are utilized, which effectively shrink the spatial size of feature maps. We test DMCN on three remote sensing datasets with different spatial resolution. Experimental results indicate that our method yeilds promising improvements of both accuracy and visual performance over several state-of-the-arts.

***Index Terms*—** remote sensing image, super resolution, convolutional neural network, image fusion

1. INTRODUCTION

High-resolution (HR) images with more detail play an essential part in remote sensing applications such as image classification and target detection. However, due to hardware limitation and **large detection distance**, remote sensing images are more complex and blurry than ordinary images. For example, an image from the ImageNet dataset measuring 256×256 pixels may only depict a cat. While an equally sized image in GaoFen-1 satellite dataset we use in this paper may cover a small town with many buildings, streets and trees (shown in Fig. 1). Besides, remote sensing images have high intra-class variance and low inter-class variance, making it much harder for detection and classification.

In addition to enhancing physical imaging technology, many researchers aim to recover high resolution (HR) images from low-resolution (LR) ones, which is called image super-resolution (SR). Deep neural networks are natural candidates to tackle the challenges of SR in remote sensing. Liebel et al.[1] utilize a three layer convolutional neural network SRCNN [2] for multispectral satellite image super resolution. Lei et al. proposed a local-global combined network (LGCNet) to enhance remote sensing images [3].

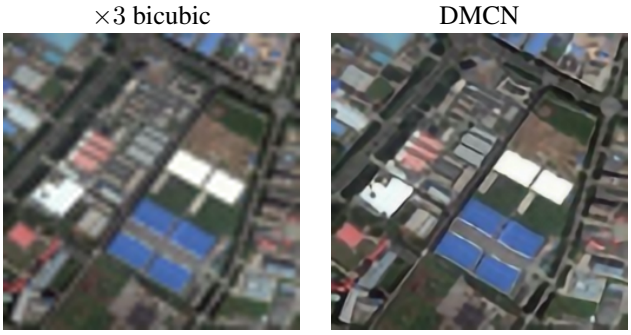


Fig. 1. The $\times 3$ super-resolution results of our method (DMCN) compared with $\times 3$ bicubic results.

However, networks such as SCRNN and LGCNet are very shallow (less than 10 layers), thus their receptive fields are small. When reconstructing HR images from remote sensing images with copious environmental information, the network capability are not satisfactory. Besides, these methods cannot reconstruct image details correctly under some circumstances, which may causes error for object detection.

In this paper, we propose a deep memory connected network (DMCN) with large receptive field and better reconstruction ability to tackle those problems in remote sensing image super resolution. The contributions of this work are as follows:

1. We build a deep network with a large receptive field, which achieves better reconstruction quality.

2. To combine local detail as well as global information learned in different neural layers, DMCN is elaborately designed with local and global memory connections.

3. We utilize downsampling and upsampling units to build a hourglass structure, significantly reducing the memory foot-

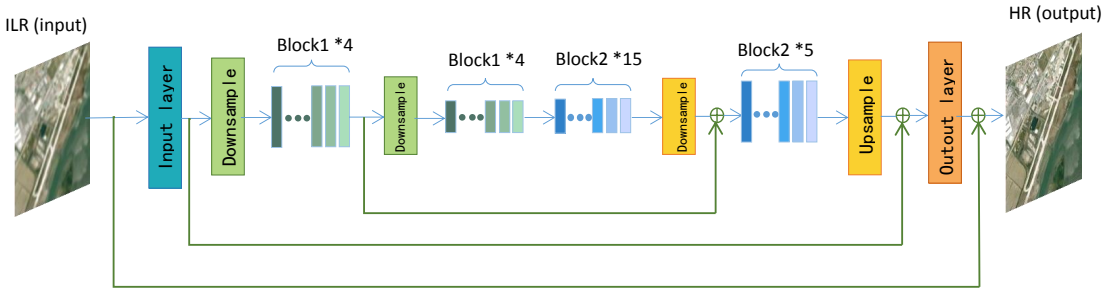
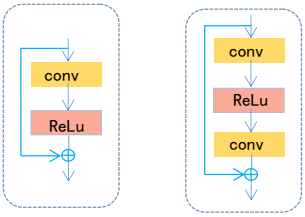


Fig. 2. The architecture of DMCN is symmetrical as a whole. The structure of Block1 and Block2 are shown in Fig. 3



(a) Block1 (b) Block2

Fig. 3. The structure of two blocks in DMCN. The convolutional layer in Block1 and Block2 are both Conv(64,3,64).

print and time consuming.

For training, three datasets with different spatial resolutions are used to test the robustness of our method. Experiment results show that DMCN outperforms the-state-of-arts.

2. METHOD

2.1. Neural Network structure

The overall structure of DMCN is illustrated in Fig. 2.. DMCN can be decomposed into four parts: input layer, downsampling unit, upsampling unit and output layer. We take an interpolated low resolution (ILR) image as input X, and learn an end-to-end mapping f between X and reconstructed HR image $\widehat{Y}$. A network with N convolutional layers can be denoted as follows,

$$f_N(X; W_N, b_N) = \sigma(W_N * f_{N-1}(X) + b_N) \qquad (1)$$

where W_i, b_i, and σ represents the filters, biases and the nonlinear function respectively. W_i is of size $c_i \times f_i \times f_i \times n_i$, where c_i is the number of input channels of the i_{th} convolutional layer, f_i is the spatial size of a filter, and n_i is the number of filters. A convolutional layer is denoted as $Conv(c_i, f_i, n_i)$.

Compared to SRCNN (3 layers) and LGCNet (7 layers), DMCN consists of 56 neural layers, which contributes to a large receptive field, providing more context to predict image detail.

2.2. Memory connection

In convolutional neural networks (CNN), the neurons of lower layers have small receptive field and focus more on local and detail information. Inspired by neural science study that human brain will protect previously acquired knowledge in neurons, we novelly propose different memory connections to combine network output with residual information: local memory connection in basic blocks, which is shown in Fig. 3 (the blue line), and global memory connection on the pipeline, shown in Fig. 2 (the green line). The function of memory connection f_c can be formulated as

$$f_c(H_{in}) = H_{in} + f_{conv}(H_{in}) \qquad (2)$$

Where H_{in} is the residual information, and f_{conv} denotes the convolutional layers between the connection.

Network with memory connections back-propagates gradients to former layers and accelerate the training process. We perform experiments in section 3 to verify these effects.

2.3. Downsampling unit and Upsampling unit

Before introducing the downsampling and upsamling units, we first investigate the time complexity of a convolutional network with N layers:

$$O_{time} = \sum_{i=1}^{N} c_i \cdot f_i^2 \cdot n_i \cdot m_i \qquad (3)$$

where m_i is the spatial size of the output feature map.

In DMCN, we propose a hourglass structure to shrink the spatial size of feature map. Our structure contains two downsampling units and two upsampling units (shown in Fig. 2). Every downsampling unit contains a convolutional layer with $stride = 2$, minishing the feature map by factor = 2. To rebuild feature map, we utilize upsampling unit with upscale factor = 2. With this hourglass structure, we significantly reduce time complexity while maintaining good performance.

Dataset	Scale	Bicubic PSNR/SSIM	SRCNN [2] PSNR/SSIM	VDSR [4] PSNR/SSIM	LGCNet [3] PSNR/SSIM	DMCN (ours) PSNR/SSIM
NWPU-RESISC45	×2	30.77/0.8172	29.37/0.7598	32.77/0.8778	32.86/0.8788	**33.07/0.8842**
	×3	27.86/0.6405	27.94/0.6545	29.28/0.7165	29.21/0.7163	**29.44/0.7251**
	×4	26.30/0.4970	26.52/0.5252	27.30/0.5549	27.35/0.5633	**27.52/0.5858**
UC Merced	×2	31.08/0.8316	31.06/0.8428	33.79/0.8909	33.80/0.8917	**34.19/0.8941**
	×3	27.59/0.6557	28.24/0.6998	29.63/0.7359	29.62/0.7350	**29.86/0.7454**
	×4	25.72/0.58	26.07/0.5439	27.31/0.5850	27.40/0.5963	**27.57/0.6150**
GaoFen1	×2	26.88/0.8585	26.98/0.8727	29.23/0.9155	29.14/0.9084	**29.26/0.9150**
	×3	23.30/0.7659	23.83/0.7264	24.65/0.7631	24.63/0.7640	**24.76/0.7658**
	×4	21.48/0.6032	21.78/0.5474	22.31/0.5879	22.23/0.5874	**22.38/0.6031**

Table 1. Evaluation of state-of-the-art SR methods on remote sensing datasets NWPU-RESISC45, UC Merced, and GaoFen1. We evaluated the average PSNR/SSIM for scale factor $\times 2$, $\times 3$ and $\times 4$. The **bold number** denotes the best performance.

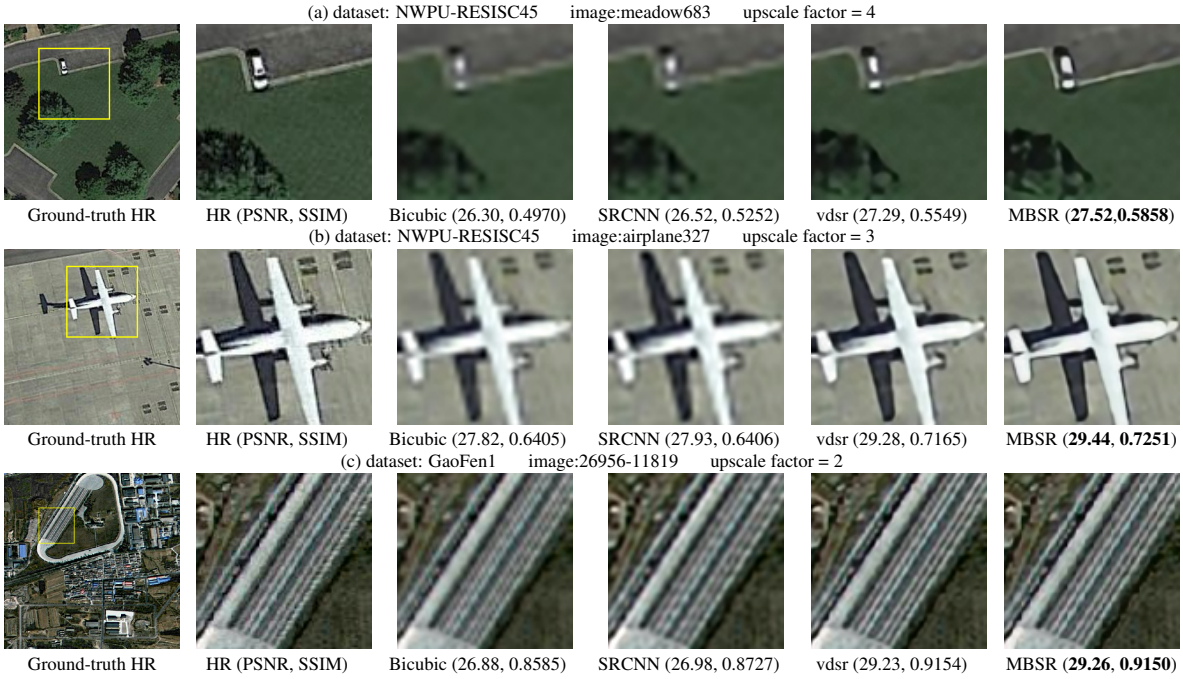


Fig. 4. Super resolution results of three datasets with upscale factor ranging from 2 to 4. In (a), the outline of the car is distinct in our result, while in other works it is blurry. In (b), the airplane in our result has clear edges. In (c), the stripe in ground truth is also observed in our result, while it is not clear in other results. In (d), our results has sharper and straight edges.

3. EXPERIMENTS AND RESULTS

3.1. Data

To verify the robustness of our method, we choose three datasets with different spatial resolutions for both training and testing.

The UC Merced land-use dataset [5] is composed of 2100 land-use scene images measuring 256×256 pixels with high spatial resolution $(0.3m/pixel)$. NWPU-RESISC45 dataset [6] is a public benchmark created by Northwestern Polytechnical University, with spatial resolution varing from $30m$ to $0.2m$ per pixel. Further more, we also use 200 multispectral images from GaoFen-1 satellite. The three visible bands of the multispectral image $(2m/pixel)$ are extracted and stacked into pseudo-RGB image. We randomly select 80% of the dataset for training and the others for testing, to verify the robustness of our model for different spatial resolution.

Given an input LR image X, we optimize parameters $\Theta = \{W_i, b_i\}$ by minimizing the loss function between the ground truth HR image Y and reconstructed image $\widehat{Y} = f(X)$. The loss function of DMCN is:

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} |f(X_i;\Theta) - Y_i| \tag{4}$$

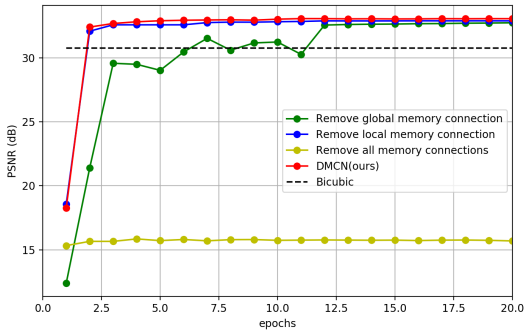


Fig. 5. The comparition of networks with or without memory connection

3.2. Training

In the training phase, the ground truth images $\{X_i\}$ are split into 48×48 sub-images with no overlap. Training uses a mini-batch size of 128. Our learning rate is initially set to 5×10^{-4} and dicreased every ten epochs by factor 10. We train the model with ADAM optimizer by setting $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^8, weight\_decay = 10^{-4}$.

3.3. Comparion with the State-of-the-arts

We evaluate the performance of DMCN on three datasets with upscale factor $\times2$, $\times3$ and $\times4$. Our method is compared with other methods including bicubic interpolation, the classic CNN-based SRCNN [2], LGCNet [3], and VDSR [4] (state-of-the-arts). In this paper, we use peak signal-to-noise ratio (PSNR) [dB] and sturctural similarity index measure (SSIM) as criteria to evaluate the performance of our network. The results are shown in Tab. 1. DMCN outperforms these methods with the highest PSNR and SSIM. Fig. 4 gives the reconstruction results. Compared with other methods, DMCN reconstructs detailed texture that are similar to the ground truth images, providing noticable improvements.

3.4. The Effect of Memory Connection

To evaluate the effect of memory connections, we disable them in turn and show the results in Fig. 5. Network with all the memory connections converges fast and gets the best performance. When we remove global and local memory connections in turn, the results decay. Network without memory connections cannot even converge.

3.5. Evaluation of Downsampling Unit and Upsampling Unit

We perform experiments to evaluate the effect of network with or without downsampling unit and upsampling unit. The result is shown in Tab. 2. Without diminishing accuracy, downsampling unit reduces memory footprint by 53.4%, and reduces testing time by 67.6%.

Table 2. Evaluate the effect of downsampling unit. Dis_D_U represents network without downsampling unit.

Model	Memory(MB)	Time(Sec)	PSNR
Dis_D_U	8265	0.037	34.17
DMCN(ours)	3849	0.012	34.19

4. CONCLUSIONS AND FUTURE WORK

In this letter, we proposes a novel network named DMCN for remote sensing image super resolution. DMCN focuses on the residuals produced at different stage and use memory connection to combine image detail with environmental information. To further reduce time complexity and memory footprint, we use downsampling unit to shrink the spatial size of feature map. Experiments shows that DMCN outperforms state-of-the-arts by a large margin in terms of visual quality and accuracy.

5. REFERENCES

[1] L Liebel and M Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *International Archives of the Photogrammetry Remote Sensing & S*, vol. XLI-B3, pp. 883–890, 2016.

[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*. Springer, 2014, pp. 184–199.

[3] Sen Lei, Zhenwei Shi, and Zhengxia Zou, "Super-resolution for remote sensing images via local-global combined network," *IEEE Geoscience and Remote Sensing Letters*, 2017.

[4] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[5] Yi Yang and Shawn Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010, pp. 270–279.

[6] Gong Cheng, Junwei Han, and Xiaoqiang Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, 2017.