

## CONVOLUTIONAL NEURAL NETWORK BASED DEM SUPER RESOLUTION

Zixuan Chen, Xuwen Wang, Zekai Xu, Wenguang Hou\*

School of Life Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, China  
-(361528641, zuolunqiang, 1042422876@qq.com, houwenguang99@163.com)

### Commission III, WG III/2

**KEY WORDS:** DEM Super Resolution, Convolutional Neural Network, Reconstruction

#### ABSTRACT:

DEM super resolution is proposed in our previous publication to improve the resolution for a DEM on basis of some learning examples. Meanwhile, the nonlocal algorithm is introduced to deal with it and lots of experiments show that the strategy is feasible. In our publication, the learning examples are defined as the partial original DEM and their related high measurements due to this way can avoid the incompatibility between the data to be processed and the learning examples. To further extent the applications of this new strategy, the learning examples should be diverse and easy to obtain. Yet, it may cause the problem of incompatibility and unrobustness. To overcome it, we intend to investigate a convolutional neural network based method. The input of the convolutional neural network is a low resolution DEM and the output is expected to be its high resolution one. A three layers model will be adopted. The first layer is used to detect some features from the input, the second integrates the detected features to some compressed ones and the final step transforms the compressed features as a new DEM. According to this designed structure, some learning DEMs will be taken to train it. Specifically, the designed network will be optimized by minimizing the error of the output and its expected high resolution DEM. In practical applications, a testing DEM will be input to the convolutional neural network and a super resolution will be obtained. Many experiments show that the CNN based method can obtain better reconstructions than many classic interpolation methods.

#### 1. INTRODUCTION

DEM super resolution (Xu et al., 2015) aims to improve the resolution of a certain DEM based on the partial high resolution DEM without increasing the cost in obtaining high-accuracy equipment. It is a relatively new topic. Yet, super resolution is a classic and traditional problem in computer vision, which is applied to generate a visual-pleasing high resolution image through a single or several low resolution input image. Similarity to the image case, DEM can be simple treated as image while its planar coordinates and its height are considered as the position of the image pixel and the corresponding gray value. Then, the method in image super resolution can be used in the DEM one (Xu et al., 2015).

Image super resolution was first proposed by Tsai and Huang (Tsai et al., 1984). Since then, many image super resolution methods are proposed, which is usually classified into three types: interpolation-based, reconstruction-based and learning-based. The main idea of interpolation-based image super resolution is to estimate the unknown pixel in the high resolution grids by using an interpolate kernel or a base function. Owing to its low complexity, this kind of approach is usually relatively faster. However, the common weakness is that these methods tend to blur the high frequency details leading to an unclear edge in the output image. Later, some more powerful methods rely on statistical image prior are proposed. The reconstruction based one regards that the low resolution image is generated from the corresponding high resolution one though transformation, deformation, down sampling and noise disturbance. These approaches typically bring in a certain prior knowledge such as gradient profile prior (Sun et al., 2011) and natural image prior (Kim et al., 2010). The major advantage of these methods is that resultant image could be visually-pleasing while a good andfeat image prior information has already been acquired. The shortcoming, however, lies in the fact that an image prior is hard to obtain when the magnification ratio is large. Thus this group of super resolution methods is

usually unrobust. The learning-based methods try to add the high frequency information learned from the training data by referencing the similarities between the input examples. Chang (Chang et al., 2004) proposed the local linear embedding based super resolution method, the example-based method can roughly regarded as nearest neighbor based estimation, firstly pairs of low resolution and corresponding high resolution image are divided into overlapping patches, then some manifold learning method such as local linear embedding (Roweis et al., 2000) are used to modeled the mapping relationship among different parts. Finally, it computes the transition probability matrix between high-resolution patches and the related low-resolution data. Support vector machine based methods (Ni et al., 2007) first classify low-resolution image patches into different kinds and then in every class the output high resolution image are generated by referencing the high frequency information only learned by the same class. However, because the relationship is established by mapping the whole set of training data points, thus this high dimensional optimization problem is hard to approach an optimum solution. To transform the high dimensional problem to a relative low dimensional one, Yang (Yang et al., 2010) used a sparse coding formulation to establish a sparse dictionaries for the low- and high-resolution image via a common encoding. Besides, Zeyde improved the sparse coding methods by employing k-SVD and orthogonal matching pursuit, which leads to a faster training and inference (Zeyde et al., 2010). Recently, Dong proposed a super resolution method based on deep learning. The method directly learned an end-to-end mapping between low and high resolution image by using convolution neural network, which make the resultant image more close to the natural image (Dong et al., 2014).

In our previous publication, According to the idea that there are many repeat or similar parts in a single DEM, we used a learning-based method to improve the resolution or accuracy on the basis of a low-resolution DEM and its partial high-resolution data. We firstly interpolated the low-resolution DEM to an expected

\*Corresponding author

resolution by bilinear or bicubic algorithms. Then, we divided the interpolated image into overlapping patches. To enlarge our learning dataset, we conducted a rotation for the learning partial DEM patches. After that, the similar patches are found for the test patch on the basis of nonlocal algorithm (Buades et al., 2005). Afterwards, we estimated the weights by minimizing the reconstruction error between the low-resolution patch and its similar patches in low-resolution domain. At last, we mapped the weighted corresponding high-resolution DEM patch to test patch and then averaged the super resolution result for all pixel in the overlapping regions. For this method, how to find similar patches would have a great influence on the resultant image, which cause a problem of incompatibility and unrobustness. In addition, the searching process is always time consuming. To solve these problems, we propose a convolutional neural network based method in this paper. It will adopt a three layers model to simulate the cognitive process of human. The main process are network training and super resolution application. A network is firstly trained on basis of lots of DEM data. Then, a high-resolution DEM will be output when the we input a low-resolution DEM to the trained network.

## 2. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network (CNN), which is a type of feed-forward artificial neural network, have recently shown a prosperous prospects in image processing and pattern recognition. In 1962, Hubel and Wiesel found that the animal visual cortex contains complex arrangements of cells which called receptive fields (Hubel et al., 1962). Then, Fukushima firstly introduced an architecture of the neocognitron which is a predecessor to convolutional networks (Fukushima et al., 1980). Since then, more attentions are focused on the theory and application of convolutional networks. The CNN includes the process of feature extraction and classification. Yet, it detects the features from the input data through implicitly learning instead of an explicit feature extraction. Furthermore, different from other deep learning network, convolution neural network is closer to the biological neural network. Its weight sharing architecture reduces the complexity of the network. So that, it can deal with high dimensional data well.

## 3. CNN-BASED DEM SUPER-RESOLUTION METHOD

We will take a low-resolution DEM data into account which is down sampled from its original one. It is firstly up-scaled to the desired size using bicubic interpolation denoted as  $Y$ . Our goal is to recover an new DEM data  $F(Y)$  from  $Y$  in the condition of that  $F(Y)$  should be similar to the ground truth high-resolution DEM data  $X$  as possible. For the ease of presentation, we still call  $Y$  a “low-resolution” DEM, although it has the same size as  $X$ . We wish to learn a mapping  $F$  using a convolutional neural network. This paper adopts a network consisting of three convolutional layer as Figure 1 (Dong et al., 2014). In the first layer, some features are extracted from the input low resolution data using a set of templates, which can be expressed as an operation  $F_1$ :

$$F_1(Y) = \max(0, W_1 * Y + B_1) \quad (1)$$

In Expression (1),  $W_1$  and  $B_1$  represent the feature detection templates and biases, respectively. Here, the size of  $W_1$  is  $1 * f_1 * f_1 * n_1$ ,  $f_1$  is the spatial size of a filter and  $n_1$  is the number of feature detection templates. Intuitively,  $W_1$  applies  $n_1$  convolutions on the input DEM data and each convolution has a kernel size  $f_1 * f_1$ .

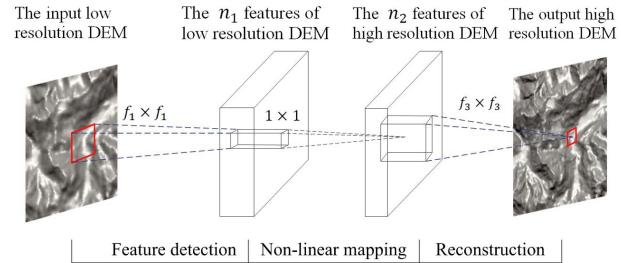


Figure 1. The CNN for super resolution

The output is composed of  $n_1$  feature maps.  $B_1$  is an  $n_1$  dimensional vector, whose each element is associated with a template. We apply the Rectified Linear Unit (ReLU,  $\max(0, x)$ ) as the activation function. In the second layer,  $n_1$  feature are mapped to  $n_2$  feature maps. This is equivalent to applying  $n_2$  templates which have a trivial spatial support  $n_1 * 1 * 1$ . The operation of the second layer is:

$$F_2(Y) = \max(0, w_2 * F_1(Y) + B_2) \quad (2)$$

Here,  $W_2$  is of size  $n_1 * 1 * 1 * n_2$  and  $B_2$  is of size  $n_2$  dimensional. In the last layer, a convolutional layer is established to produce the final high resolution image as:

$$F_3(Y) = w_3 * F_2(Y) + B_3 \quad (3)$$

In Expression (3),  $W_3$  is of a size  $n_2 * f_3 * f_3 * 1$  and  $B_3$  is a  $c$  dimensional vector. Interestingly, although the above three operations are motivated by different intuitions, they all lead to the same form as a convolutional layer. We put all three operations together and form a convolutional neural network. In this model, all the templates and biases are to be optimized. According to a series of experiments, we found that the  $f_1 = 9$ ,  $f_3 = 5$ ,  $n_1 = 64$ ,  $n_2 = 32$  may be a optimized parameters. On basis of these parameters, a fully trained convolutional neural network can be used to obtain a fine super resolution result. After determining the structure of the convolution neural network, the main issues for convolution neural network based DEM super-resolution method are as establishing training data, training the network and performing super resolution for a DEM.

### 3.1 Training data

Convolution neural network based DEM super-resolution method is a supervised learning method, i. e., our training set should contain the low-resolution DEM data and corresponding high-resolution DEM data (when only high-resolution DEM data, we can use down sampling method to obtain the corresponding low resolution DEM data). In our training set, the ground truth DEM data  $\{X_i\}$  are prepared as  $33 * 33$  grids block randomly cropped from the high resolution training DEM data. And the low resolution samples  $\{Y_i\}$  are prepared as  $21 * 21$  grids blocks randomly cropped from the low resolution training DEM data.

### 3.2 CNN Network training

Learning the end-to-end mapping function  $F$  requires the estimation of parameters  $\theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$ . This is achieved through minimizing the loss between the reconstructed DEM data  $F(Y; \theta)$  and the corresponding ground truth high resolution DEM data  $X$ . Given a set of high resolution DEM data

$\{X_i\}$  and their corresponding low resolution DEM data  $\{Y_i\}$ , we use Mean Squared Error (MSE) as the loss function:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - X_i\|^2 \quad (4)$$

Where,  $n$  is the number of training samples. The loss is minimized using stochastic gradient descent with the standard back propagation.

### 3.3 Application for DEM super resolution

After fully being trained, the convolutional neural network can express the mapping relationship of the low resolution DEM data to the high resolution DEM data. At this time, the low resolution DEM data will be input to the training of the network with a high resolution DEM as output.

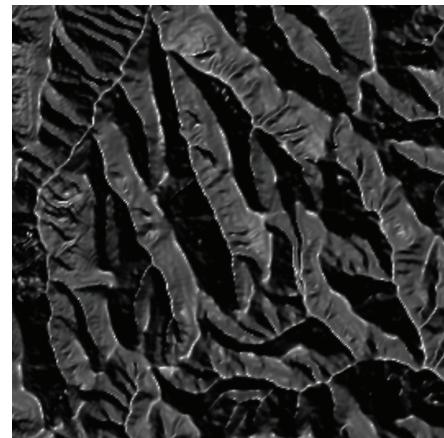
## 4. EXPERIMENTS AND ANALYSIS

Two experiments have been done to validate the effectiveness of the proposed method. The first data is a smooth DEM and another is the data of a mountains region which all have been used in (Xu et al., 2015). We firstly down-sampled them to 1/4, 1/9 and 1/16 pyramid images. Then, these pyramid images are interpolated to the original resolution on basis of the bicubic algorithm. At the same time, we trained three CNN networks according to above mentioned method. Three CNN networks corresponds to 1/4, 1/9 and 1/16 reconstructions. Afterwards, the interpolated pyramid images will be input to the corresponding network and the reconstruction image can be obtained. Figure 2 shows the 1/4 down sampled DEM, the bicubic interpolation and the CNN-based reconstruction. Compared with the down sampled DEM the CNN-based reconstruction and bicubic interpolation are better Yet, the difference between them are visually minor. Figure 3 shows the corresponding data of another DEM. Figure 3(b) is the bicubic interpolation and Figure 3(c) is the corresponding CNN based reconstruction. It can be found that more details are appeared in Figure 3(c). The details are reconstructed by the CNN which are learned from the training data.

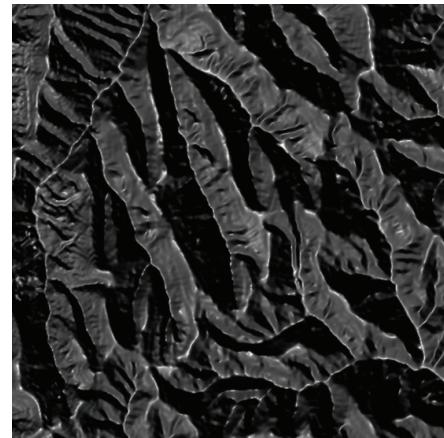
Data	Methods	The down sampling ratio		
		2	3	4
Data1	Bicubic	1.990	4.479	7.868
	CNN-based	1.612	4.048	7.521
Data2	Bicubic	3.227	6.231	9.888
	CNN-based	2.627	5.823	9.701

Table 1. The mean square error of reconstructions

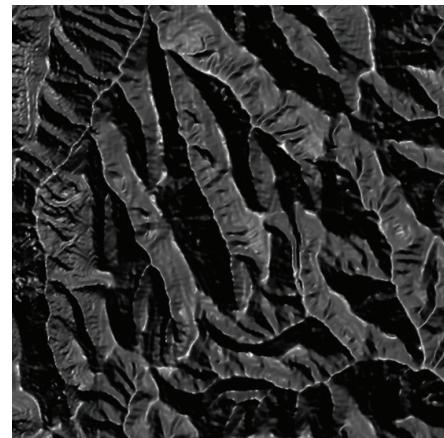
Meanwhile, we perform quantitative evaluation for above two data by comparing the super resolution results with the original DEM. MSE is taken as the criterion. Table 1 shows the comparisons. From it, we can find that the CNN based method can obtain better results than the classic bicubic interpolation. Moreover, the preponderance is about 10 percent in cases of 1/4 and 1/9 down sampling. It should be mentioned that many experiments have been done which shows that the CNN based method can obtain better reconstructions in most situations. In particular, the superior is relatively obvious in cases of DEM with more details.



(a) Down sampled DEM



(b) Bicubic interpolation



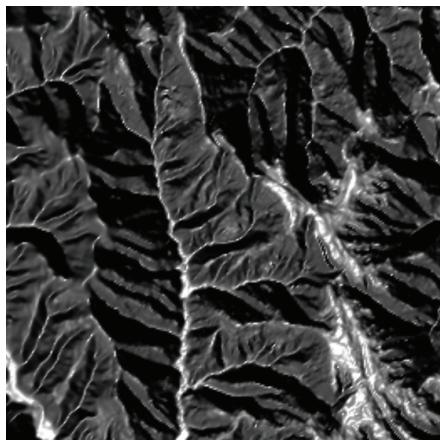
(c) CNN based super resolution

Figure 2. Down sampled DEM, bicubic and CNN reconstruction

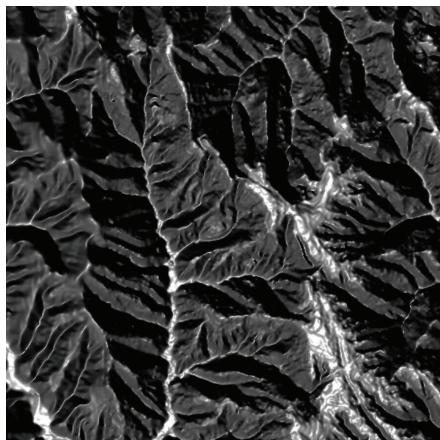
## 5. CONCLUSION

This paper intends to propose a new method to improve the resolution of a DEM on basis of the CNN. It obtain better results than the classic bicubic method and shows better robustness than our previous proposed nonlocal similarity based DEM super resolution. CNN simulates the process of human beings in reconstructing the high resolution data from the related low resolution ones. Many details can be reconstructed from the CNN which are learned from the training data, i. e., the structure and the learning process make it be superior to the traditional reconstruction

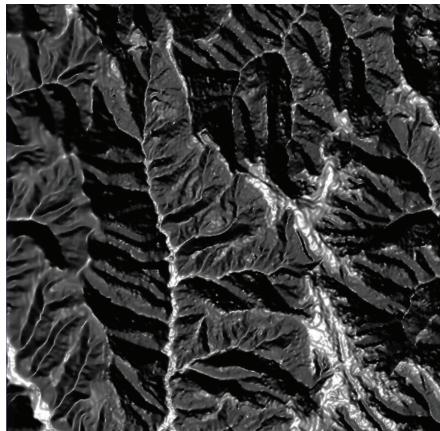
methods. How to improve the accuracy and accelerate the training process will be our future works.



(a) Down sampled DEM



(b) Bicubic interpolation



(c) CNN reconstruction

Figure 3. Bicubic interpoation and CNN reconstruction of a mountain DEM

## REFERENCES

- Buades A, Coll B, Morel J M. A non-local algorithm for image denoising. In: *Computer Vision and Pattern Recognition*, 2005, 2: 60-65.
- Chang, H., Yeung, D., Xiong, Y. Super-resolution through neighbor embedding. In: *Graphical Models and Image Processing*, 2004, 275-282.
- Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution. *Computer Vision ECCV 2014 Springer International Publishing*, 2014: 184-199.
- Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 1980, 36(4): 193-202.
- Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 1962, 160(1): 106-154.
- Kim K I, Kwon Y. Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans.Pattern Analysis and Machine Intelligence*, 2010, 32(6): 1127-1133.
- Ni K S, Nguyen T Q. Image super resolution using support vector regression. *IEEE Trans.Image Processing*, 2007, 16(6): 1596-1610.
- Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(5500): 2323-2326.
- Sun J, Xu Z, Shum H Y. Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE Trans.Image Processing*, 2011, 20(6): 1529-1542.
- Tsai R Y, Huang T S. Multiframe image restoration and registration. *Advances in computer vision and Image Processing*, 1984, 1(2): 317-339.
- Xu Z, Wang X, Chen Z, et al. Nonlocal similarity based DEM super resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 110: 48-54.
- Yang J, Wright J, Huang T S, et al. Image super-resolution via a sparse representation. *IEEE Trans.Image Processing*, 2010, 19(11): 2861-2873.
- Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. *Curves and Surfaces. Springer Berlin Heidelberg*, 2010: 711-730.

*Revised April 2016*

## ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (Grant Number : 41371426) and Key Laboratory for National Geographic Census and Monitoring, National Administration of Surveying, Mapping and Geoinformation (Grant Number : 2014NGCM09).