Super-Resolution Reconstruction of Images Based on Microarray Camera

Jiancheng Zou[1, *], Zhengzheng Li[1], Zhijun Guo[1] and Don Hong[2]

Abstract: In the field of images and imaging, super-resolution (SR) reconstruction of images is a technique that converts one or more low-resolution (LR) images into a high-resolution (HR) image. The classical two types of SR methods are mainly based on applying a single image or multiple images captured by a single camera. Microarray camera has the characteristics of small size, multi views, and the possibility of applying to portable devices. It has become a research hotspot in image processing. In this paper, we propose a SR reconstruction of images based on a microarray camera for sharpening and registration processing of array images. The array images are interpolated to obtain a HR image initially followed by a convolution neural network (CNN) procedure for enhancement. The convolution layers of our convolution neural network are 3×3 or 1×1 layers, of which the 1×1 layers are used to improve the network performance particularly. A bottleneck structure is applied to reduce the parameter numbers of the nonlinear mapping and to improve the nonlinear capability of the whole network. Finally, we use a 3×3 deconvolution layer to significantly reduce the number of parameters compared to the deconvolution layer of FSRCNN-s. The experiments show that the proposed method can not only ameliorate effectively the texture quality of the target image based on the array images information, but also further enhance the quality of the initial high resolution image by the improved CNN.

Keywords: Super-resolution reconstruction, microarray camera, convolution neural network.

1 Introduction

Super-resolution (SR) reconstruction is an image processing technique that converts one or more low-resolution (LR) images into high-resolution (HR) images. SR reconstruction of images has always been a hot topic in the field of image processing. Harris-Goodman spectrum extrapolation method [Harris (1964)], which realized SR in practical applications, became of great interest in SR research. Subsequently, Kim et al. [Kim and Su (1993)] extended the observation model and Gerchberg [Gerchberg (1974)] presented a SR method for continuous energy decline setting. Wadaka et al. [Wadaka and Sato

[1] College of Sciences, North China University of Technology, Beijing, 100144, China.

[2] Department of Mathematical Sciences, Middle Tennessee State University, Murfreesboro, TN 37132, USA.

[*] Corresponding Author: Jiancheng Zou. Email: zjc@ncut.edu.cn.

CMC. doi:10.32604/cmc.2019.05795

(1975)] proposed a superimposed sinusoidal template method. Nevertheless, the above-mentioned methods were used to obtain HR images from a single image in the frequency domain and the effects were not satisfactory. Until 1980, in order to solve the problem of low resolution in remote sensing image processing, Huang et al. [Huang and Tsai (1981)] first used multi-frame image sequences for SR reconstruction and got significant improvements. In 1986, Park et al. [Park, Min and Kang (2003)] proposed the maximum likelihood restoration method (Poisson-ML) based on Poisson distribution. Irani et al. [Irani and Peleg (1991)] developed a method to enhance resolution by image registration. Lecun et al. [Lecun, Bottou and Bengio (1998)] applied gradient-based learning for character recognition. Furthermore, Freeman et al. [Freeman, Jones and Pasztor (2002)] suggested sample-based SR reconstruction and Glasner et al. [Glasner, Bagon and Irani (2009)] combined traditional multi-image SR reconstruction with sample-based SR reconstruction.

With the development of deep learning, SR reconstruction technology has made breakthroughs. Krizhevshy et al. [Krizhevshy, Sutskever and Hinton (2012)] successfully classified images using deep convolution neural network (CNN). Meng et al. [Meng, Steven, Wang et al. (2018)] proposed a fusion steganographic algorithm based on faster regions with CNN features (Faster R-CNN). Cui et al. [Cui, Chang and Shan (2014)] presented deep network cascade (DNC) using small-scale parameters of each layer to enhance the resolution of images layer by layer. Osendorfer et al. [Osendorfer, Soyer and Smagt (2014)] proposed deconvolution network which learns the nonlinear mapping from LR space to HR space, and ultimately achieves the goal of improving the resolution. Dong et al. [Dong, Chen and He (2014); Dong, Chen and Tang (2016)] suggested that convolution neural network can be applied to SR reconstruction to generate HR image from a single LR image faster (SRCNN and FSRCNN-s). The reconstruction effect of this method is remarkable, but it cannot eliminate the noises in the LR image.

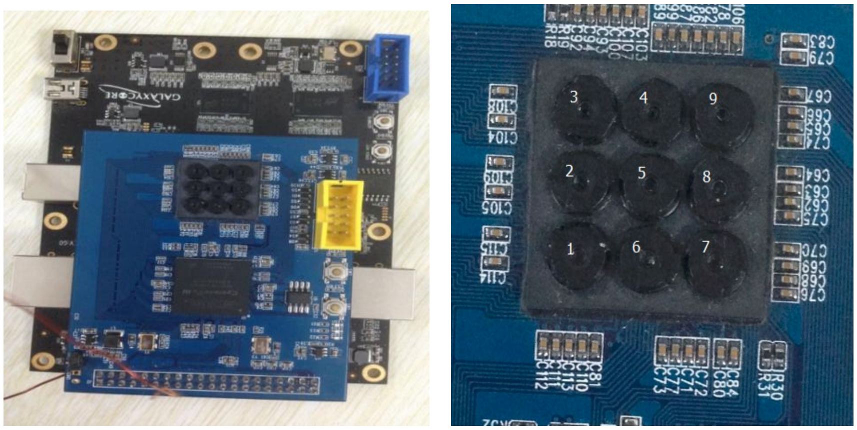
There is a way to improve the performance of SR using array images captured simultaneously by cameras. In recent years, binocular camera mode became popular. The main purpose of this camera is to imitate a few of the binocular visual characteristics of the human eye, and the camera will be adjusted to distance measurement and depth information acquisition. The array camera is usually composed of multiple lenses. The main difference from binocular lenses is that the array camera intends to mimic some of the visual features of the insect compound eye. This requires multiple cameras to be arrayed into an array form.

Currently, Stanford University, Pelican, Microsoft and other companies have invested in research on array cameras. Among them, Pelican introduced an array camera named PiCam [Kartik, Dan, Andrew et al. (2013)]. The camera is composed of sixteen sub-lenses arranged in 4×4 planes with the characteristics of ultra-thin thickness and high performance. Pelican has achieved good results in depth maps and SR reconstruction using array cameras. They use the disparity relationship between multi-scene images to obtain the depth map of the scene, and then use the depth information of the depth map to further synthesize high-quality HR images. In addition, Wilburn et al. [Wilburn, Joshi and Vaish (2004)] put forward a method of high-speed video capture using array cameras. By increasing the number of lens in array camera, the target of acquiring thousands of frames per second can be achieved. The array images used in this paper are captured by a microarray camera produced by ourselves. Unlike the array lenses produced by pelican, our array camera can output color images, which effectively cuts down the disadvantage of light and motion led by time delays and guarantees the efficiency of image registration. The space of the device in exchange for the advantages of the time, improve the speed and effectiveness of registration.

In this paper, we present a novel method of SR reconstruction of images based on microarray camera. First, array images from a microarray camera are performed in sharpening of the array images to enrich the texture details of the images. Then, the initial HR image is obtained by registration and interpolation of the array images. Finally, we propose an improved convolution neural network and use the network to reconstruct the initial HR image. Experiments show that the resolution and quality of the image are improved significantly.

2 3×3 microarray camera

In this section, a microarray camera is used for image acquisition. The 3×3 microarray camera is composed of nine micro-lenses arranged in a certain way as shown in Fig. 1. Compared with traditional cameras, this kind of array camera has the advantages of smaller size and more lenses, so the collected images have richer information. The microarray camera is about a dime in size as a whole and can capture nine different views of the same scene synchronously.



(a) (b)

Figure 1: Microarray camera for this paper. (a) The overall structure of the array camera. (b) The number of lens in microarray camera

In this paper, the image captured by Lens 5 is taken as the reference image. It has the most common areas with those images captured by the other lenses. In the case of low illumination, the camera is susceptible to illumination, resulting in noise on the captured images as shown in Fig. 2. The local detail of the reference image is displayed to show the noise effect as shown in Fig. 3.

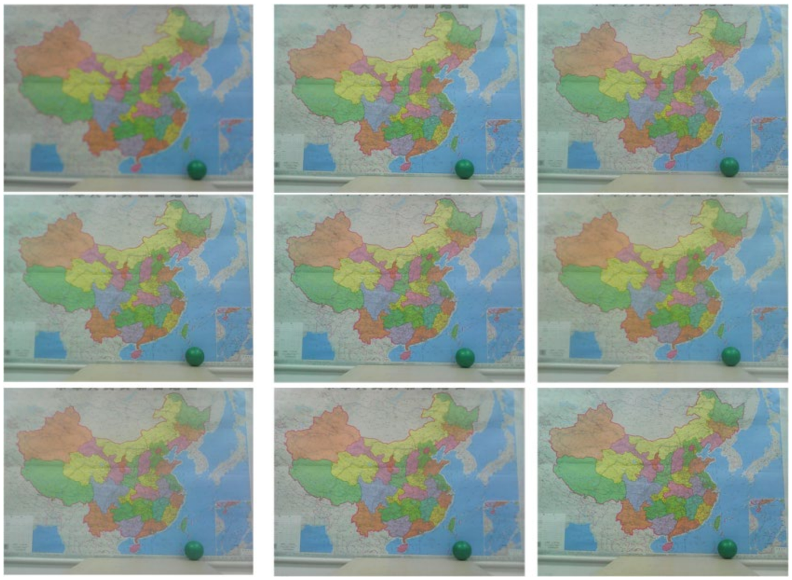


Figure 2: Array images captured by microarray camera

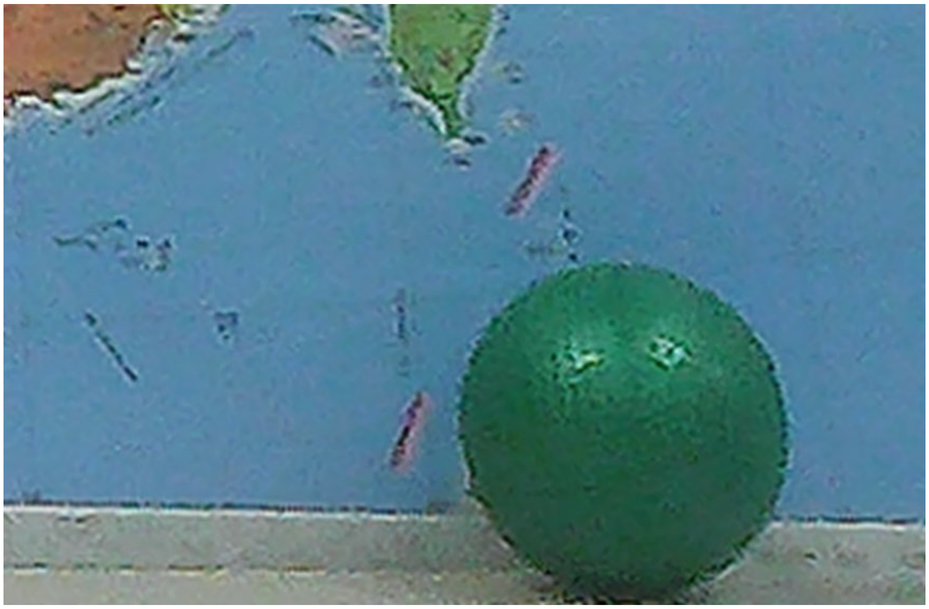


Figure 3: The local detail of the array image

3 Proposed method

3.1 Sharpening of the array images

Before the registration of nine array images, the geometric local adaptive sharpening (GLAS) is processed. The equation of GLAS [Zhu and Milanfar (2011)] is:

$$f(x_i) = \frac{\sum_{x \in w_l} S(x - x_i) Y(x)}{\sum_{x \in w_l} S(x - x_i)} \tag{1}$$

where $Y(x)$ is a low resolution image affected by noise and blur, and the S represents the kernel of GLAS.

We can see that the texture information is more abundant after the sharpening of the array images as shown in Fig. 4.

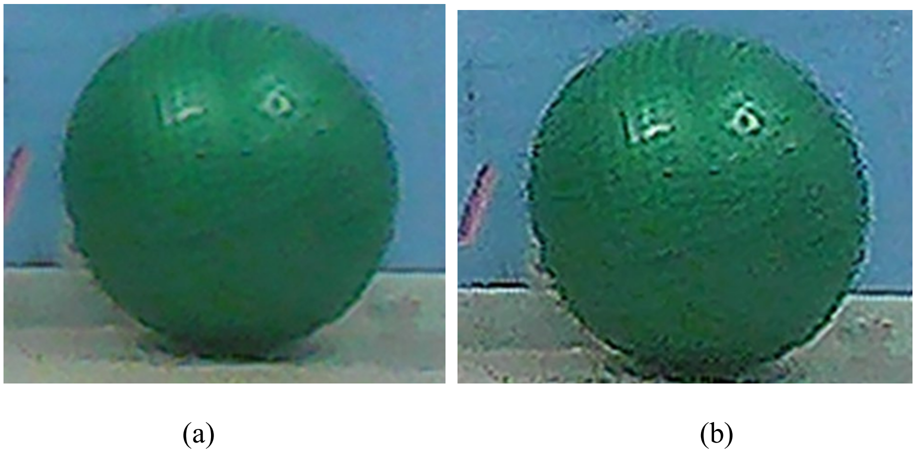


Figure 4: Comparison of local details in images. (a) The reference image. (b) The reference image after sharpening

3.2 Registration and fusion

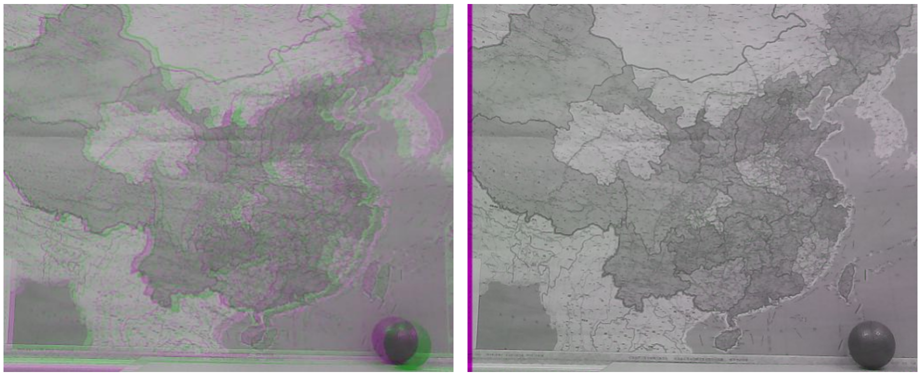
In this section, the other eight images are registered with the reference image respectively. We use multi-scale feature matching method [Sorgi] to detect the feature points of array images. In the scale space of this method, the image of each layer is generated by the blurring of the upper layer. In general, the Gauss function is used to convolve the image of the previous layer to get the next layer result.

$$Y_{l+1}(x,y) = Y_l(x,y) \otimes g_{\delta_\gamma}(x,y) \tag{2}$$

where $Y_l(x,y)$ is the image of layer l, and $g_{\delta_\gamma}(x,y)$ refers to the filtering window with standard deviation of δ .

In this paper, the multi-scale feature matching method is used to block the registration between the array images and the reference image. Because the parallax between each image and the reference image is small, the use of block registration can reduce the mismatching phenomenon.

The difference between the left-top image and the center image in the fused image is expressed in purple and green respectively. From the fusion image without registration as shown in Fig. 5(a), we can see that the disparity of the sphere is obvious. After registration, the fused image no longer appears double image as shown in Fig. 5(b), that is, the disparity offset decreases. Therefore, this method can effectively reduce the disparity shift between the other eight images and the reference image.



(a) (b)

Figure 5: The comparison of fusion process. (a) The fusion result without registration of the left-top image and the center image in Fig. 2(b) The fusion result with registration of the left-top image and the center image in Fig. 2

3.3 Interpolation

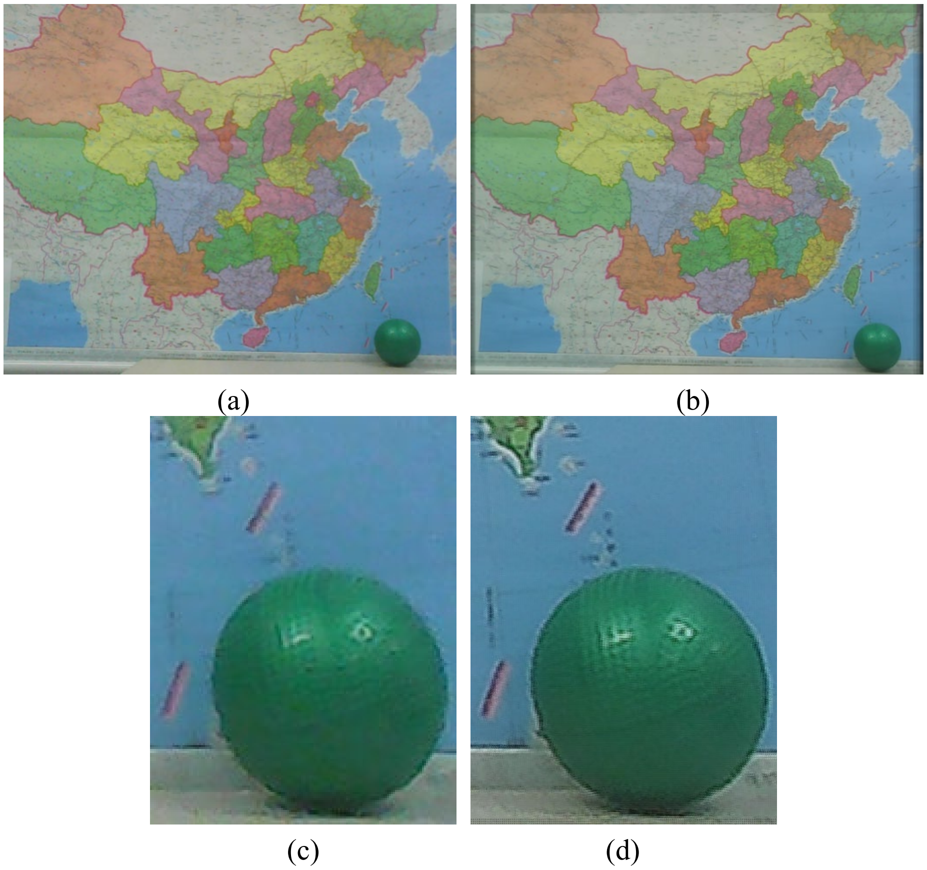
Interpolation processing is carried out on the registered array images, and additional information between images is used to enrich the texture information of the reference image. By this method, an initial HR image which has more texture details than the original image as shown in Fig. 6 can be generated.

To illustrate the improvement of interpolation, we use mean, standard deviation, smoothness and entropy to measure the difference between the reference image and interpolated image. Mean and standard deviation indicate the change of brightness and darkness of the image. Smoothness and entropy represent the smoothness and texture roughness of the image respectively. The smoother the image is, the smoothness tends to be 0; the higher the texture roughness is, the greater the entropy value is.

Table 1: Local texture contrast between the reference image and interpolated image

	The original of Lens 5	Interpolated image
Mean	107.06	106.12
Standard deviation	25.90	28.76
Smoothness	0.0102	0.0126
Entropy	6.2000	6.5434

As shown in Tab. 1, the brightness of the image after interpolation is slightly darker than that of the reference image. Smoothness and entropy enhancement indicate that the roughness of image texture is increased. Therefore, interpolation can effectively improve the texture details of images.



(a) (b)

(c) (d)

Figure 6: Comparison between the reference image and the interpolated image. (a) The reference image captured by Lens 5. (b) Interpolation result by nine array images. (c) Local details of (a). (d) Local details of (b)

3.4 Improved convolutional neural network

In this section, an improved convolution neural network is proposed which can achieve good results of magnification. The network has nine layers which can be divided into five main parts: feature extraction, size shrinkage, nonlinear mapping, dimension recovery, and deconvolution amplification.

The first eight layers of the network use convolution layers for feature information processing, and the last layer uses deconvolution layer for image enlargement. All the step values of the convolution layer are set to 1, while the step size of the deconvolution is set to 3. The method in this section is consistent with FSRCNN-s in the selection of activation function: in order to avoid the "feature loss" problem caused by ReLU, PReLU is selected as activation function after each convolution layer. PReLU can be expressed as:

$$f(x_i) = \max(x_i, 0) + a_i \min(0, x_i) \tag{3}$$

Where x_i be the input of the activation function on the i channel, and a_i is the coefficient on negative half axis.

In feature extraction of the input images, we apply a 3×3 convolution layer to the network and obtain the feature information of LR image block through this layer. Because the 3×3 size of this layer is smaller than that 5×5 size of the first layer in FSRCNN-s, the first layer of this method has fewer network complexity parameters. And $Conv_1(1,3,1,0,30)$ is represented as the first layer.

We use the second convolution layer to reduce the dimension of feature extraction: the 3×3 convolution layer reduces the dimension of vector feature information of LR image from 30 to 6. This layer is used $Conv_2(30,3,1,0,6)$ for presentation. After second layer, a 1×1 convolution layer is added to the convolution neural network. Although the nonlinear capacity of the network will promote as the number of convolution layers increases (i.e., the number of activation functions), arbitrary addition of convolution layers will lead to the difficulty of convergence. At the same time, this will also lead to the complexity of the reconstruction process. In order to avoid the problem of increasing network complexity and keep the spatial dimension of feature information unchanged as much as possible [Lin, Chen and Yan (2015)], we need to select smaller convolution layers, such as 1×1 convolution layer which only improves $6 \times 6 \times 1$ for the complexity of the overall convolution network. This layer is called $Conv_3(6,1,1,0,6)$.

In the nonlinear mapping, we use the "bottleneck" [He, Zhang and Ren (2015)] construction structure (three-tier stacking structure) to promote the nonlinear capacity of the network to construct complex functions. The "bottleneck" construction structure can be broken down into three parts: $1 \times 1, 3 \times 3, 1 \times 1$ three convolution layers. Among them, the second level output dimension of "bottleneck" construction structure is reduced by half than that of the previous level. These three layers can reduce feature dimension, transform feature information and restore feature dimension respectively. It should be noted that the filling values of the two 1×1 convolution layers are both set to 1, so the width and height of the generated image blocks will be expanded after the fourth and the sixth "bottleneck" structures are constructed. The three layers are represented by using $Conv_4(6,1,1,1,3)$, $Conv_5(3,3,1,0,3)$, and $Conv_3(3,1,1,1,6)$, respectively.

The structure and function of the seventh layer are similar to those of the third layer. The purpose of this layer is to increase the nonlinear capacity of the network through the 1×1 convolution layer. The eighth layer of the network is for dimension recovery. The 3×3 layer restores the image feature information dimension of the previous layers from 6 to 30. The two layers are expressed as $Conv_7(6,1,1,0,6)$, $Conv_8(6,3,1,0,30)$.

The last layer uses the deconvolution layer to magnify the image feature information and generate the corresponding HR image. The layer is used $Deconv_9(30,3,3,1,1)$ for presentation. Because this layer uses a 3×3 deconvolution layer instead of 9×9

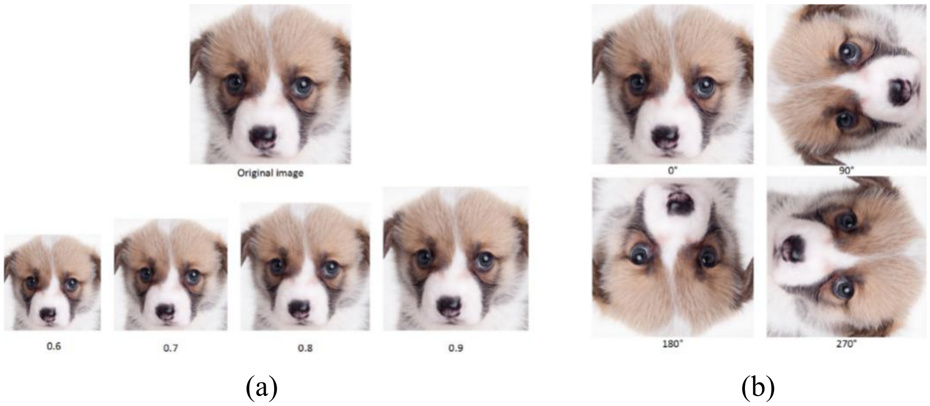
deconvolution layer in FSRCNN-s, the main parameters of network complexity in this layer are greatly reduced:

$$32\times 9^2\times 1 - 30\times 3^2\times 1 = 2592 - 270 = 2322$$

4 Experimental results

4.1 Convolution neural network configuration and training

In order to prove the superiority of the convolutional neural network proposed in this paper, the same data set as FSRCNN-s is used for network training in experimental comparison. In this section, 91-image and General-100 image data sets are used as training images, and Set5 data sets are selected as test images. Furthermore, the above training images are enhanced by scaling and rotating the images in data sets as shown in Fig. 7.

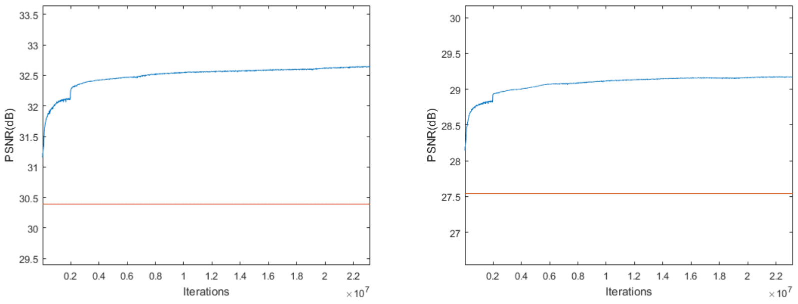


(a) (b)

Figure 7: Data enhancement is done in two ways. (a) The original image is scaled in 0.6, 0.7, 0.8, 0.9 ratio. (b) The original image rotates at 90 degrees, 180 degrees and 270 degrees

After the data enhancement process of training images, the grayscale images of these images are taken as original HR images. After that, the original HR gray-scale images are scaled 2 or 3 times, and then the size is restored by bi-cubic interpolation method. We use these restored images with obvious texture blurring as LR images. Finally, the LR and the original HR images are segmented by $10^2/19^2$ and $7^2/19^2$ ratio respectively under the conditions of double and triple enlargement.

The method of training convolution neural network is Caffe [Jia, Yang and Shelhamer (2014)]. The deconvolution layer initialization is set to Gauss type and 0.001 standard deviation value. When training convolution neural network [He, Zhang and Ren (2015)], only 91-image data sets are used for training. The initial learning rate of convolution layer is set to 0.001, and that of deconvolution layer is set to 0.0001. When the network is nearly saturated, 91-image and General-100 enhanced data sets are used as training data. At this point, we need to change the learning rate of convolution layer to 0.0005 and the deconvolution layer to 0.00005.



(a) (b)

Figure 8: With the increase of iteration times, the change of average PSNR. (a) Set5 data sets as test images. (b) Set14 data sets as test images

In the initial stage of training convolution neural network with three times amplification condition, only 91-image data set is used for training. When the number of iterations reaches 2 million, PSNR grows slowly and shows a smooth trend. After that, the training data sets are replaced by 91-image and General-100 enhanced data sets. The PSNR values of Set5 and Set14 test images show a short-term sharp increase. After that, the PSNR values gradually promote with the increase of iteration times until they approach the smooth state again.

4.2 Convolution neural network reconstruction experiment

In this section, Set5, Set14 and array images were reconstructed with bi-cubic interpolation, SRCNN, FSRCNN-s and improved convolutional neural network respectively. The performance of the above four reconstruction algorithms is evaluated by PSNR index. We compared the average reconstruction time of SRCNN, FSRCNN-s and convolutional neural network proposed in this paper as shown in Tab. 2.

The computer used in this experiment is 64 bits, with the Intel Core i5-4590 3.30GHz processor and the 4GB memory.

The size of the array images is 1200×1600, so the number of pixels is close to 2 million in each image. Then the convolution neural network reconstruction method is used to reconstruct the array images, which generates approximately 18 million pixels of each HR image, and verifies that the resolution of HR images has been improved to a certain extent.

From Tab. 3, it can be seen that the sharpness and the resolution of LR images acquired by microarray camera are obviously improved after reconstruction by convolution neural network.

Table 2: The average PSNR values and time of four reconstruction methods

Test Set	Magnification	Bi-cubic Interpolation (dB)	SRCNN (dB/s)	FSRCNN-s (dB/s)	Improved CNN (dB/s)
Set5	2	33.66	36.34/1.88	36.58/2.31	36.69/1.95
Set14	2	30.23	32.18/5.78	32.28/4.60	32.44/3.89
Array images	2	29.83	32.11/62.32	31.77/40.90	32.52/35.77
Set5	3	30.39	32.39/1.99	32.61/1.14	32.65/1.01
Set14	3	27.54	29.00/5.52	29.12/2.18	29.17/1.84
Array images	3	27.43	28.37/62.11	28.48/18.89	28.59/16.31

Table 3: Wedge comparison of LR and HR images

	LR image	HR image
Sharpness (LW/PH)	960	1500
Valid pixel	1229556	3002149
Pattern pixel	1920000	17280000

After reconstructed by the improved convolution neural network method, the edge details of the image are clearer and the blurring phenomenon in the image is obviously reduced as shown in Fig. 9.

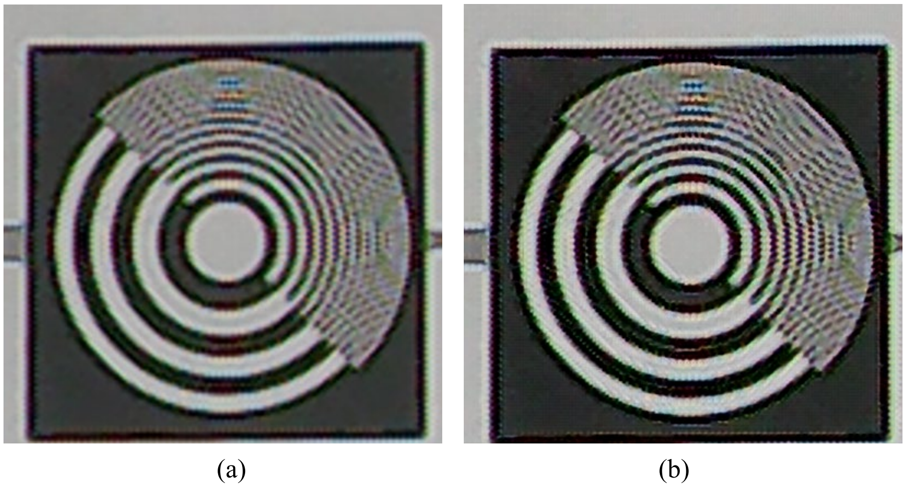


Figure 9: Comparison of reconstruction results by two methods. (a) Reconstructed by bi-cubic interpolation method. (b) Reconstructed by improved convolutional neural network

4.3 SR reconstruction for interpolation results

In this section, the array images have been sharpened, registered and interpolated. Then, the interpolated image is reconstructed by using the three-fold magnification model parameters of the improved convolution neural network. The original image captured by Lens 5 is compared with the reconstructed image in detail as shown in Fig. 10.

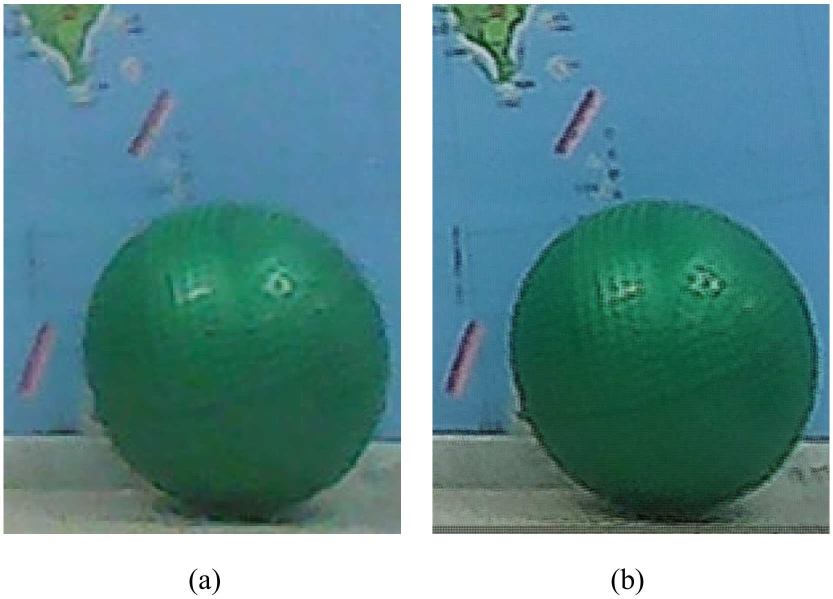


Figure 10: Comparison of reconstruction results with original in detail. (a) The original image of Lens 5. (b) The final HR image processed by our method

From Fig. 10, we can see that the image edge details after processing by convolution neural network are clearer and richer. The texture details of the sphere are clearer, and the noise is significantly reduced.

SSIM and PSNR image evaluation standards need to take the original image as a prerequisite to measure the effect of improving the original image. The array images used in this paper cannot be evaluated using SSIM and PSNR because there is no original image. Gray evaluation gradient (GMG) is an effective method to measure the effect in the condition of no original image. The formula is as follows:

$$GMG = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{[X(i+1,j)-X(i,j)]^2+[X(i,j+1)-X(i,j)]^2}{2}}}{MN} \tag{4}$$

where $X(i+1,j)$, $X(i,j)$, $X(i,j+1)$ are the gray values of the corresponding points of the images, and MN is the size of the images.

Table 4: The GMG comparison between the original image and the reconstruction result

	The original of Lens 5	After reconstruction
GMG	7.7315	17.8630

5 Conclusion

In this paper, firstly, we use a microarray camera to capture the scene images, then the array images are sharpened. Afterwards the array images are registered by multi-scale feature registration method, and the initial estimated HR image is obtained by interpolating the additional information in the array images. Finally, an improved convolutional neural network reconstruction algorithm is proposed and used to further enhance the quality of the initial estimated HR image. Experiments show that the proposed method is effective.

Acknowledgement: The work of this paper is supported by the National Natural Science Foundation of China under grant numbers 61572038, and the Innovation Project Foundation NCUT.

References

Cui, Z.; Chang, H.; Shan, S. (2014): Deep network cascade for image super-resolution. *European Conference on Computer Vision*, pp. 49-64.

Dong, C.; Chen, C. L.; He, K. (2014): Learning a deep convolutional network for image super-resolution. *European Conference on Computer Vision*, pp. 184-199.

Dong, C.; Chen, C. L.; Tang, X. (2016): Accelerating the super-resolution convolutional neural network. *European Conference on Computer Vision*, pp. 391-407.

Freeman, W. T.; Jones, T. R.; Pasztor, E. C. (2002): Example-based super-resolution. *Computer Graphics & Applications*, vol. 22, no. 2, pp. 56-65.

Gerchberg, R. W. (1974): Super-resolution through error energy reduction. *Optica Acta: International Journal of Optics*, vol. 21, no. 9, pp. 709-720.

Glasner, D.; Bagon, S.; Irani, M. (2009): Super-resolution from a single image. *International Conference on Computer Vision*, pp. 349-356.

Harris, J. L. (1964): Diffraction and resolving power. *Journal of the Optical Society of America*, vol. 54, pp. 931-933.

He, K.; Zhang, X.; Ren, S. (2015): Deep residual learning for image recognition. *IEEE Computer Society*, pp. 770-778.

He, K.; Zhang, X.; Ren, S. (2015): Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026-1034.

Huang, T. S.; Tsai, R. Y. (1981): Image sequence analysis: motion estimation. *Image Sequence Analysis*, pp. 1-18.

Irani, M.; Peleg, S. (1991): Improving resolution by image registration. *Cvgip Graphical Models & Image Processing*, vol. 53, no. 3, pp. 231-239.

Jia, Y.; Shelhamer, E. (2014): Caffe: convolutional architecture for fast feature embedding. *ACM International Conference on Multimedia*, pp. 675-678.

Kartik, V.; Dan, L.; Andrew, M.; Gabriel, M.; Priyam, C. (2013): PiCam: an ultra-thin high performance monolithic camera array. *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 1-13.

Kim, S. P.; Su, W. Y. (1993): Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 534-539.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012): Image-Net classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*, pp. 1097-1105.

Lecun, Y.; Bottou, L.; Bengio, Y. (1998): Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324.

Lin, M.; Chen, Q.; Yan, S. (2015): Network in network. Computer Science.

Meng, R.; Steven G. R.; Wang, J.; Sun, X. (2018): A fusion steganographic algorithm based on Faster R-CNN. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 1-16.

Osendorfer, C.; Soyer, H.; Smagt, P. V. D. (2014): Image super-resolution with fast approximate convolutional sparse coding. *International Conference on Neural Information Processing*, pp. 250-257.

Park, S. C.; Min, K. P.; Kang, M. G. (2003): Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21-36.

Sorgi, L.: Multiscale stereo features matching. *Features*.

Wadaka, S.; Sato, T. (1975): Super-resolution in incoherent imaging system. *Journal of the Optical Society of America*, vol. 65, no. 3, pp. 354-355.

Wilburn, B.; Joshi, N.; Vaish, V. (2004): High-speed videography using a dense camera array. *IEEE Computer Vision and Pattern Recognition*, vol. 2, pp. 294-301.

Zhu, X.; Milanfar, P. (2011): Restoration for weakly blurred and strongly noisy images. *IEEE Workshop on Applications of Computer Vision*, pp. 103-109.