

# Single Image Super Resolution with Neighbor Embedding and In-place Patch Matching

Zhong-Qiu Zhao<sup>1</sup>, Zhen-Wei Hao<sup>1</sup>(✉), Run Su<sup>2</sup>, and Xindong Wu<sup>1</sup>

<sup>1</sup> College of Computer and Information, Hefei University of Technology, Hefei, China  
1216459032@qq.com

<sup>2</sup> The No. 38 Research Institute of CETC, Hefei, China

**Abstract.** In this paper, we present a novel image super-resolution framework based on neighbor embedding, which belongs to the family of learning-based super-resolution methods. Instead of relying on extrinsic set of training images, image pairs are generated by learning self-similarities from the low-resolution input image itself. Furthermore, to improve the efficiency of image reconstruction, the in-place matching is introduced to the process of similar patches searching. The gradual magnification scheme is adopted to upscale the low-resolution image, and iterative back projection is used to reduce the reconstruction error at each step. Experimental results show that our method achieves satisfactory performance not only on reconstruction quality but also on time efficiency, as compared with other super-resolution methods.

**Keywords:** Super-resolution · Neighbor embedding · Self-similarity · In-place matching

## 1 Introduction

Image super resolution (SR) reconstruction is the process of generating a high-resolution (HR) image by using one or more low-resolution (LR) inputs. As a post-processing technique, it breaks through the limitations of image equipment and environment, and has a wide range of applications in real life such as high-definition television, medical imaging, video surveillance etc. The problem of image SR was first studied by Tsai and Huang in 1980s [1], and now existing SR methods can be broadly categorized into three categories: interpolated-based, reconstruction-based, and learning-based methods. In this paper, we focus on the learning-based methods which exploit the information from the training data to learn the mapping relationship between the LR and HR image patches.

As a kind of very promising SR technique, learning-based methods were first studied by Freeman et al. [2]. Nowadays, based on different learning models, lots of learning-based methods have been proposed. According to the ways how the mapping relationship is formulated, learning-based SR methods can be divided into: manifold learning [3–6], sparse coding [7–9] and regression-based methods [10–12]. Although learning-based methods can effectively restore the missing details, these methods primarily rely on example data. As the contents of natural images are ever-changing, it is difficult to

choose a universal training set to fit the reconstruction of all images. When the example data cannot provide accurate information which is similar to the LR input image, lots of artifacts can be easily introduced to the result image. More recently, Glasner et al. [13] proposed a method that exploited patch redundancy in natural image, and realized image SR reconstruction without external database.

In [3], Chang et al. proposed an SR method using neighbor embedding (NE), in which the HR image can be estimated with a set of weighted image patches. Until now, many modified algorithm over the neighbor embedding have been proposed. Bevilacqua et al. [4] proposed to use non-negative least squares to compute the weights of the neighbor patches. Gao et al. [5] proposed a sparse neighbor embedding method which chooses the neighbor patches and computes the corresponding weights simultaneously. Based on the multiscale similarity learning, Zhang et al. [6] accomplished the image SR reconstruction with only the LR input image. Although the image reconstruction methods above can obtain promising reconstruction performance, they need to search a vast example dataset or the whole image for similar patterns, leading to low efficiency in practical applications. Recently, several efforts have been made to improve time efficiency. With a precomputed projection matrix, Timofte et al. [14] accomplished the image reconstruction and acquired high time efficiency. Based on the in-place examples, Yang et al. [15] proposed to learn a set of regression mapping for fast image SR. Dong et al. [16, 17] used convolution neural network to learn an end-to-end mapping model, and obtained satisfactory reconstruction performance with high time efficiency.

In this paper, we propose a NE-based single image SR scheme which considers both the selection of example data and time efficiency. With the similarity patch redundancy in natural images, our method adopt the multiscale similarity learning to generate the example image pairs using only the input image. And contrary to [6], the proposed method conducted the k-nearest neighbor search with the in-place patch matching. We adopt gradual magnification scheme to upscale the LR image and implement iterative back projection algorithm [18] to correct the reconstruction image at each step.

The remainder of this paper is organized as follows. Section 2 introduces the NE-based approach for SR. Section 3 presents the details of our SR framework. The experimental results and comparisons with several other SR methods are demonstrated in Sect. 4. Finally, several conclusions are drawn in Sect. 5.

## 2 Neighbor Embedding Approach for Super Resolution

The NE-based SR method was first proposed by Chang et al. [3], which uses a manifold learning called locally linear embedding (LLE) to estimate the HR patches corresponding to the LR ones. Its main idea is that the image patches in different resolutions have similar geometric manifold spaces and each image patch can be represented as the linear combination of its K neighbors.

Let  $Y_s = \{y_s^j\}_{j=1}^m$  be the training patches of LR images,  $X_s = \{x_s^j\}_{j=1}^m$  be the corresponding HR training patches. We represent the LR test image as a set of overlapping image patches  $Y_t = \{y_t^i\}_{i=1}^n$ , and  $X_t = \{x_t^i\}_{i=1}^n$  are their corresponding HR patches to be estimated. The NE-based SR reconstruction can be summarized as follows:

- (1) For each patch  $y_t^i \in Y_t$ 
  - (a) Find the  $K$  nearest neighbors in  $Y_s$  in terms of Euclidean distance:

$$N_i = \arg \min_{1 \leq j \leq m}^k \|y_t^i - y_s^j\|^2 \tag{1}$$

- (b) Compute the reconstruction weights of the neighbors that minimize the reconstruction error:

$$\varepsilon^i = \left\| y_t^i - \sum_{y_s^j \in N_i} w_{ij} y_s^j \right\|_2^2 \quad s.t. \quad \sum_{y_s^j \in N_i} w_{ij} = 1 \tag{2}$$

- (c) Compute the HR image patch using the corresponding HR patches and the reconstruction weights:

$$x_t^i = \sum_{x_s^j \in N_i} w_{ij} x_s^j \tag{3}$$

- (2) Construct the HR image by combining the obtained HR image patches.

### 3 Proposed Algorithm

In this section, we provide the details of our modified algorithm on the NE-based method. The procedure is divided into three steps: (1) Generating example data, (2) NE-based relationship learning with in-place patch matching, (3) SR enhancement using iterative back projection. The proposed SR method is summarized in Algorithm 1. The details in each step of the algorithm are as follows:

#### 3.1 Generating Example Data

It is the most fundamental and important step for learning-based image SR to select the appropriate HR-LR training image pairs. In this paper, utilizing the self-similarity redundancy of the natural image, the HR-LR image pairs can be extracted from the LR input image.

For an observed LR image  $Y$ , the problem of image super-resolution can be modeled as:

$$Y = DHX + \sigma \tag{4}$$

where  $H$  and  $D$  is the operator of blurring and down-sampling, respectively.  $X$  is the HR image, and  $\sigma$  is the noise term.

In the process of image reconstruction, we magnify the LR image to the desired size with a gradual scheme. The scale parameter is set to be  $k = 0, \dots, \lceil \log(S)/\log(s) \rceil$ , where  $S$  and  $s$  are the overall and gradual up-scaling factors, respectively. To obtain

example image pairs, we firstly regard the original LR input image  $Y$  as the HR image  $X^{(0)}$ . Then the corresponding LR image  $Y^{(0)}$  can be obtained by down-sampling  $X^{(0)}$  and up-scaling the down-sampled image to the size of  $X^{(0)}$  using the bicubic interpolation:

$$Y^{(0)} = ((X^{(0)} * B_s) \downarrow_s) \uparrow_s \quad (5)$$

where  $B_s$  is the blurring kernel,  $\uparrow_s$  and  $\downarrow_s$  are the up-scaling and down-scaling operators with the factor of  $s$ . Then  $X^{(0)}$  and  $Y^{(0)}$  are treated as the initial example image pairs for learning-based algorithm. Iteratively, to obtain the HR image  $X^{(1)}$ , we produce  $Y^{(1)}$  through magnifying  $X^{(0)}$  using bicubic interpolation with a factor of  $s$ . With the similarity assumption, for each patch in  $Y^{(1)}$ , we can find several similar patches in  $Y^{(0)}$ . The HR patch in  $X^{(1)}$  can be linearly represented by the image patches in  $X^{(0)}$  which correspond to the patches found in  $Y^{(0)}$ .  $X^{(1)}$  and  $Y^{(1)}$  are then taken as the HR-LR pairs for succeeding learning. The synthesis processes are repeated until the HR image at scale  $\lceil \log(S)/\log(s) \rceil$  is obtained.

### 3.2 NE-Based Relationship Learning with in-Place Patch Matching

Once the training example pairs are formed, the relationship between the HR and the corresponding LR images can be estimated from the pair mappings. In this paper, we take advantage of the NE-based method for this purpose.

For learning the relationship between HR-LR pairs, K-nearest neighbors (K-NN) search was used to find several of the most similar patches in the training set for each LR image patch. Contrary to previous methods which need to search in a vast external example set or the whole image, we conduct the similar patches search in a restricted local region which is built on the patch matching with small scaling factors is in-place. Thereby, the time efficiency of image reconstruction is improved by narrowing the searching scope. For an image patch at the location  $(x, y)$  of  $Y^{(k+1)}$ , we can find the  $K$  most similar patches around its original coordinates  $(x_r, y_r)$  in image  $Y^{(k)}$ , where  $x_r = \lfloor x/s + 0.5 \rfloor$  and  $y_r = \lfloor y/s + 0.5 \rfloor$ . Then we use the LLE algorithm to compute the reconstruction weights between the image patch in  $Y^{(k+1)}$  and the matched similar image patches in  $Y^{(k)}$  by minimizing the reconstruction error. Then the HR patch can be recovered by using the reconstruction weights and the corresponding HR neighbor patches. Once all the HR patches are generated, the HR image can be obtained through merging HR patches.

### 3.3 SR Enhancement Using Iterative Back Projection

In each step, due to the unavoidable reconstruction error, the reconstruction result which is fed to next step directly usually leads to the accumulation of errors. Thus, through the iterative back projection (IBP) algorithm we reduce the reconstruction error at each step, and ensure that image reconstruction is conducted along the right direction. The reconstruction error can be formulated as:

$$E(X) = \|(X \otimes H) \downarrow - Y\|^2 \quad (6)$$

where  $X$  is the estimated HR image,  $Y$  is the LR input image, and  $H$  represents blurring operator.

This problem can be efficiently solved by the gradient descent method, and the updating equation for this iterative method is:

$$X^{t+1} = X^t + ((Y - DHX^t) \uparrow) * p \quad (7)$$

where  $t$  is iteration index,  $p$  is back-projection filter, and  $X^t$  is the HR image at the  $t^{\text{th}}$  iteration.

---

**Algorithm 1.** Proposed Image Super-Resolution Algorithm

---

**Input:** Low-resolution input image  $Y$ .

**1. Initialize:**

- 1) Set the scale parameter  $n = \lceil \log(S) / \log(s) \rceil$ ;
- 2) Set HR image  $X^{(0)} = Y$ , and the corresponding LR image  $Y^{(0)} = (X^{(0)}) \downarrow \uparrow$ .

**2. NE-based reconstruction:**

for  $k = 1$  to  $n$  do

- 1) Calculate the LR image  $Y^{(k)}$  with the bicubic interpolation of  $X^{(k-1)}$ ;
- 2) For each patch in  $Y^{(k)}$ , find its K-NN around its original coordinates in  $Y^{(k-1)}$ , then use LLE algorithm to compute the reconstruction weights;
- 3) Compute HR image patches using the reconstruction weights and the corresponding HR patches in  $X^{(k-1)}$ ;
- 4) Produce the HR image  $X^{(k)}$  by merging the HR patches;
- 5) Use the IBP algorithm to reduce the reconstruction error, and obtain the new  $X^{(k)}$ .

end for

**3. Post-Processing:**

- 1) Use the bicubic interpolation to resize the HR image  $X^{(n)}$  to the  $S$  times of the LR image size. The non-local means is finally used to suppress the artifacts in the HR image.

**Output:** High-resolution output image  $X$ .

---

## 4 Experiment Results and Analysis

In this section, the experiments on several benchmark images are shown to validate our proposed method. Several state-of-the-art learning-based SR methods are used as comparison baselines, including NE-based [3], SC-based [7], Glasner's method [13] and Zhang's method [6]. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are adopted to evaluate the objective quality of the SR results, and reconstruction time is used as the indicator of reconstruction efficiency.

## 4.1 Experimental Settings

In our experiments, the LR images are generated from the original HR images through down-sampling by bicubic interpolation with a factor of 3. Since human vision is sensitive to brightness changes, the SR reconstruction is only conducted in the gray-channel for color images, and the images in the remaining color channels are directly magnified to the desired size using bicubic interpolation. In the reconstruction phase, the LR patch size is set as  $5 \times 5$  pixels, and 3 pixels are overlapped between adjacent patches. The neighborhood size  $K$  is set to be 3, and when conducting the similar patches search across different scales, the search scope is set as  $11 \times 11$  pixels around the corresponding coordinates. For all the SR experiments, the up-scaling factor  $S$  is set to be 3, and we magnify the LR image step-by-step by the factor  $s$  which is set to be 1.25. At each step, we use the IBP algorithm to enhance the quality of the reconstruction image, and the number of iterations is set to be 20.

## 4.2 Experimental Results

Table 1 lists the PSNR and SSIM values of various SR approaches. From Table 1, we can see that our proposed method can obtain the best performance of PSNR and SSIM on most of benchmark images. This fact indicates that with the LR image itself, we can

**Table 1.** The PSNR and SSIM performance on eight test images of super resolution of  $3\times$  magnification

Methods		NE-based	SC-based	Glasner	Zhang [6]	Proposed
Images		[3]	[7]	[13]		
bike	PSNR	23.42	23.87	23.96	24.45	<b>24.56</b>
	SSIM	0.741	0.770	0.783	0.792	<b>0.803</b>
butterfly	PSNR	24.92	25.51	26.74	<b>28.18</b>	27.18
	SSIM	0.830	0.861	0.890	<b>0.919</b>	0.909
girl	PSNR	32.79	33.35	33.17	33.52	<b>33.53</b>
	SSIM	0.813	0.823	0.825	0.826	<b>0.829</b>
plants	PSNR	31.58	32.39	32.85	33.59	<b>33.72</b>
	SSIM	0.879	0.902	0.910	0.917	<b>0.922</b>
raccoon	PSNR	28.67	28.89	28.79	29.06	<b>29.15</b>
	SSIM	0.754	0.759	0.762	0.761	<b>0.768</b>
lenna	PSNR	32.23	32.85	32.59	33.51	<b>33.58</b>
	SSIM	0.865	0.877	0.879	0.884	<b>0.887</b>
baby	PSNR	33.77	34.51	34.29	<b>35.20</b>	35.09
	SSIM	0.901	0.917	0.922	0.923	<b>0.924</b>
pepper	PSNR	30.67	31.45	32.82	33.36	<b>34.05</b>
	SSIM	0.863	0.879	0.885	0.889	<b>0.893</b>

recover the HR image and obtain promising results. The reason is that similar patches across different scales are beneficial to recover faithful details.

We also assess the visual quality of the SR methods. Figures 1 and 2 show the SR images obtained by different SR methods on the test images of butterfly and lenna, respectively.



**Fig. 1.** The SR images(3 $\times$ ) of the butterfly image by various methods. Top row: the ground truth HR image, NE-based [3], SC-based [7]. Bottom row: Glasner [13], Zhang [6], and ours.



**Fig. 2.** The SR images(3 $\times$ ) of the lenna image by various methods. Top row: the ground truth HR image, NE-based [3], SC-based [7]. Bottom row: Glasner [13], Zhang [6], and ours.

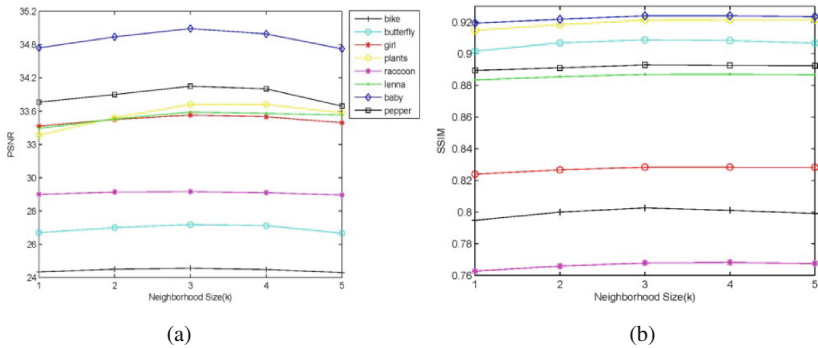
In addition, we use computation time to evaluate the reconstruction efficiency. Table 2 lists the reconstruction time of different SR methods. From the table, we can see that our method takes the least computation time, and possesses the highest reconstruction efficiency. This improvement is due to the adoption of in-place patch matching, which reduces the searching scope for similar patches. Thus, our method is more feasible in real applications, comparing with other state-of-the-art SR methods.

**Table 2.** The reconstruction time(seconds) of different SR methods.

Methods Images	NE-based [3]	SC-based [7]	Glasner [13]	Zhang[6]	Proposed
bike	98	63	142	341	<b>44</b>
butterfly	99	55	135	331	<b>44</b>
girl	98	69	141	345	<b>45</b>
plants	97	62	149	336	<b>44</b>
raccoon	146	111	241	580	<b>66</b>
lenna	391	266	650	2815	<b>178</b>
baby	328	285	703	2850	<b>179</b>
pepper	331	272	682	2865	<b>178</b>

### 4.3 Experimental Analysis

As one type of learning-based SR methods, the NE-based algorithm is employed in our method to build the relationship between the HR and LR image patches. So, the neighborhood size  $K$  is a variable parameter in our method. To study the effect of neighborhood size  $K$  on reconstruction performance, we take different  $K$  values to test our method on the benchmark images. The PSNR and SSIM results are presented in Fig. 3, from



**Fig. 3.** Performance comparison of testing images with different neighborhood sizes. (a) PSNR value with different neighborhood sizes (b) SSIM value with different neighborhood sizes.



which we can see that our method yields better results when the neighborhood size  $K$  is set to be three.

## 5 Conclusion

This paper presents a novel single image SR framework by incorporating both similarity redundancy and in-place patch matching. The similarity redundancy across different scales is used to generate example data for the learning-based reconstruction, and in-place patch matching is utilized to fulfill the similar patches search in order to improve the reconstruction efficiency. To exploit the similarity redundancy across different scales, we magnify the LR image to the desired size with a gradual scheme, and use the IBP algorithm to reduce the reconstruction error at each step. The experimental results indicate that our method can obtain highly satisfactory performance in objective evaluation, while maintaining relatively high reconstruction efficiency.

## References

1. Huang, T.S., Tsai, R.Y.: Multi-frame image restoration and registration. *Adv. Comput. Vis. Image Process.* **1**(2), 317–339 (1984)
2. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super resolution. *IEEE Comput. Graph. Appl.* **22**(2), 56–65 (2002)
3. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. *CVPR* **1**, 275–282 (2004)
4. Bevilacqua, M., Roumy, A., Guillemot, C., Morel, M.L.A.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: *BMVC*, pp. 1–10 (2012)
5. Gao, X., Zhang, K., Tao, D., Li, X.: Image super-resolution with sparse neighbor embedding. *IEEE Trans. Image Process.* **21**(7), 3194–3205 (2012)
6. Zhang, K., Gao, X., Tao, D., Li, X.: Single image super-resolution with multiscale similarity learning. *IEEE Trans. Neural Networks Learn. Syst.* **24**(10), 1648–1659 (2013)
7. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: *CVPR*, pp. 1–8 (2008)
8. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., Chenin, P., Cohen, A., Gout, C., Lyche, T., Mazure, M.-L., Schumaker, L. (eds.) *Curves and Surfaces 2011. LNCS*, vol. 6920, pp. 711–730. Springer, Heidelberg (2012)
9. Dong, W., Zhang, L., Shi, G., Wu, X.: Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. Image Process.* **20**(7), 1838–1857 (2011)
10. Li, D., Simske, S., Mersereau, R.M.: Single image super-resolution based on support vector regression. In: *IJCNN*, pp. 2898–2901 (2007)
11. Yang, M.C., Chu, C.T., Wang, Y.C.F.: Learning sparse image representation with support vector regression for single-image super-resolution. In: *ICIP*, pp. 1973–1976 (2010)
12. Yang, M.C., Wang, Y.C.F.: A self-learning approach to single image super-resolution. *IEEE Trans. Multimedia* **15**(3), 498–508 (2013)
13. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: *ICCV*, pp. 349–356 (2009)

14. Timofte, R., Smet, V.D., Gool, L.V.: Anchored neighborhood regression for fast example-based super-resolution. In: ICCV, pp. 1920–1927 (2013)
15. Yang, J., Lin, Z., Cohen, S.: Fast image super-resolution based on in-place example regression. In: CVPR, pp. 1059–1066 (2013)
16. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part IV. LNCS, vol. 8692, pp. 184–199. Springer, Heidelberg (2014)
17. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2015)
18. Irani, M., Peleg, S.: Improving resolution by image registration. *CVGIP Graphical Models Image Process.* **53**(3), 231–239 (1991)