



# Image super-resolution reconstruction based on deep dictionary learning and A+

Yi Huang<sup>1,2</sup> · Weixin Bian<sup>1,2</sup> · Biao Jie<sup>1,2</sup> · Zhiqiang Zhu<sup>1,2</sup> · Wenhui Li<sup>1,2</sup>

Received: 18 October 2023 / Revised: 23 November 2023 / Accepted: 1 December 2023 / Published online: 3 January 2024  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

## Abstract

The method of image super-resolution reconstruction through the dictionary usually only uses a single-layer dictionary, which not only cannot extract the deep features of the image but also requires a large trained dictionary if the reconstruction effect is to be better. This paper proposes a new deep dictionary learning model. Firstly, after preprocessing the images of the training set, the dictionary is trained by the deep dictionary learning method, and the adjusted anchored neighborhood regression method is used for image super-resolution reconstruction. The proposed algorithm is compared with several classical algorithms on Set5 dataset and Set14 dataset. The visualization and quantification results show that the proposed method improves PSNR and SSIM, effectively reduces the dictionary size and saves reconstruction time compared with traditional super-resolution algorithms.

**Keywords** Deep dictionary learning · Image super-resolution · Anchored neighborhood regression · Sparse representation

## 1 Introduction

Image super-resolution (SR) reconstruction technology refers to the process of reconstructing a high-resolution image of a corresponding scene from a low-resolution image of a given scene. The single-image super-resolution problem focuses on how to generate believable, visually pleasing high-resolution (HR) output images from low-resolution (LR) input images. Unlike similar image enhancement algorithms, in SR, it is usually assumed that the input image, although of lower resolution in the number of pixels, is still of high resolution at the original scale. SR method mainly focuses on increasing the scale of the image while ensuring that the image is as clear as possible. The most direct way to improve image resolution is to improve the optical hardware in the acquisition system, however, since the manufacturing process is difficult to greatly improve, and the manufacturing cost is very high, it is often too expensive to increase the image resolution physically. Therefore, the

technology of image super-resolution reconstruction from the perspective of software and algorithms has become a research hotspot in many fields such as image processing and computer vision. Traditional super-resolution reconstruction algorithms mainly rely on basic digital image processing techniques for reconstruction, such as interpolation-based super-resolution reconstruction, degradation model-based super-resolution reconstruction, and learning-based super-resolution reconstruction.

Interpolation-based methods [1, 2] treat each pixel on the image as a point on the image plane, so the estimation of super-resolution images can be seen as fitting unknown pixel information to the plane using known pixel information on process. This type of algorithm is usually completed by a predefined transformation function or interpolation kernel. The method based on interpolation is simple to calculate and easy to understand, but there are some obvious defects. First, it assumes that the change of pixel gray value is a continuous and smooth process, but in fact this assumption is not fully established; second, in the reconstruction process, super-resolution is calculated only according to a pre-defined transfer function. The image, regardless of the degradation model of the image, often leads to blurring, jaggedness and other phenomena in the restored image, so the effect is not good. Nearest neighbor interpolation, bilinear interpolation,

✉ Weixin Bian  
bwx2353@ahnu.edu.cn

<sup>1</sup> School of Computer and Information, Anhui Normal University, Wuhu 241002, Anhui, China

<sup>2</sup> Anhui Provincial Key Laboratory of Network and Information Security, Wuhu 241002, Anhui, China

and bicubic interpolation are common interpolation-based methods.

Degradation model-based super-resolution reconstruction [3] considers that HR images are properly transformed by motion, blur and noise, resulting in reduced image resolution. This method constrains super-resolution image reconstruction by extracting key information in LR images and incorporating prior knowledge of images that require super-resolution reconstruction. Common methods include iterative back projection method [4], projection to convex set [5] and maximum a posteriori method [6], etc.

The learning-based method believes that there are certain correspondences between the LR images and the HR images, mainly to learn these correspondences, so a large amount of training data is needed, and these data are used to learn mapping relationships, which can be used to predict the HR images corresponding to the LR images to complete the reconstruction. Common learning-based methods include manifold learning [7], sparse coding methods [8], etc. Gradient priors [3, 9, 10], non-local self-similarity priors [11–13] and sparsity priors [14–16] are some typical image priors. These methods are able to recover sharp edges and suppress aliasing artifacts. However, the prior imposed on the HR image will affect the performance of these methods to a relatively large extent, and the reconstruction effect is poor when the zoom scale of the image is large.

Yang et al. [17] used the sparseness of natural images, introduce the theory of compressed sensing, build a dictionary that can sparsely represent image blocks through dictionary learning, and then, the reconstructed image can be generated through the linear combination of the dictionary and the sparse representation coefficients obtained by linear programming, so as to restore the high-frequency details of the image and achieve better results. On the basis of Yang, Zeyde et al. [18] used K-SVD to train the dictionary, which not only greatly accelerated the dictionary training time but also improved the efficiency of learning. Compared with the method of Yang et al. [17], the reconstruction effect Significant improvement. After that, Timofte et al. [19] performed manifold learning on dictionaries obtained by dictionary sparse training, which accelerated the image reconstruction time while ensuring the image reconstruction effect. Subsequently, he proposed an improved A + algorithm (Timofte et al. [20]) which resulted in faster training times and improved reconstruction quality. Perez-Pellitero [21] proposed an improved SR linear regressor training strategy and a reverse search method for accelerating regression-based SR methods. Zhao et al. [22] proposed a multi-resolution dictionary learning (MRDL) model that not only restores details well, reduces aliasing and noise, but also have a large improvement in computational efficiency. Zhang et al. [23] proposed a joint super-resolution framework with structurally modulated sparse representations, which

also improves the performance of image super-resolution. But all the work are based on a single-layer training dictionary for reconstruction, which leads to the fact that if the reconstruction effect is expected to be good, the size of the trained dictionary is relatively large, and the deep features of the image in these methods do not have much mention and attention.

Tariyal et al. [24] proposed the method of deep dictionary learning (DDL). In single-layer dictionary learning, one utilizes the learned dictionary to synthesize data from the coefficients. DDL extends this concept from a single layer to multiple layers, learning a multi-layer dictionary so that the deepest coefficients can be used to synthesize data. Mahdizadehghadam et al. [25] proposed a new model that attempts to learn a deep dictionary for classification tasks. Compared with traditional deep neural networks, deep dictionary learning methods usually first divide images into patches and extract features at the patch level, which is the most obvious difference between it and the deep neural network. The learned patch-based feature dictionary is then used to transform the input data into a global sparse feature representation. Song et al. [26] proposed multi-layer discriminative dictionary learning (MDDL) with local constraints for image classification. Through multi-layer dictionary learning, a robust dictionary is learned in the last layer, and the separability of encoded vectors belonging to different categories is improved compared to other methods. Classification accuracy is also guaranteed. The classification accuracy is also guaranteed. Tang et al. [27] proposed a new deep-to-point encoding network for image multi-classification problems. Thereafter, Montazeri et al. [28] proposed a multi-layer K-SVD method, which was also used to solve classification problems. However, image super-resolution reconstruction using deep dictionary learning is relatively rare. Recently, Huang et al. [29] proposed an image super-resolution model for deep dictionary learning. Different from other multi-layer dictionaries, their architecture contains L-1 analysis dictionary and synthesis dictionary for extracting high-level features. However, the results are slightly deficient with the method proposed in this paper. Singhal et al. [30] proposed deep coupled dictionary learning for solving the image inverse problem. Inspired by such research results, this paper will propose an improved image super-resolution reconstruction method for the limitations of image super-resolution reconstruction methods through dictionary learning. In the training part, this paper uses the deep dictionary learning method to learn the deep dictionary, and then uses the A + method to complete the super-resolution reconstruction of the image. Experimental results show that in the case of the same dictionary size, the proposed algorithm in this paper has good improvements in PSNR and SSIM, and effectively reduces the dictionary size.

The main contributions to our work are summarized as follows:

- (1). A new method of super-resolution based on deep dictionary learning is proposed.
- (2). While ensuring the reconstruction effect, the size of the dictionary is reduced and the reconstruction time is shortened.
- (3). Compared with some classical methods, our method performs better with the same dictionary size. Compared with some methods in recent years, our method is also quite competitive.

The rest of the paper are organized as follows. The second part outlines the relevant technologies and concepts used in our approach. The third part introduces the detailed process of the method. In Sect. 4, we conducted an experimental evaluation of the proposed scheme, and in Sect. 5, we summarized our work.

## 2 Related work

This section details the main methods involved in the proposed scheme.

### 2.1 Deep dictionary learning

The shallow dictionary learning model is:

$$X = DZ \tag{1}$$

where  $X$  is the training data,  $D$  represents the dictionary, and  $Z$  is the sparse coefficient.

DDL draws on the idea of deep learning, and further extracts more abstract deep features by extending shallow dictionary learning to multiple layers. Figure 1 shows a schematic diagram of two-layer dictionary learning. Mathematically, it can be modeled as:

$$X = D_1\varphi(D_2Z_2) \tag{2}$$

here  $\varphi$  is a nonlinear activation function. Extending this idea, the problem of multi-level dictionary learning can be expressed as:

$$X = D_1\varphi(D_2\varphi(\dots\varphi(D_NZ_N))) \tag{3}$$

### 2.2 Anchored neighborhood regression

In the image super-resolution problem, most of the neighborhood embedding (NE) and sparse coding (SC) methods

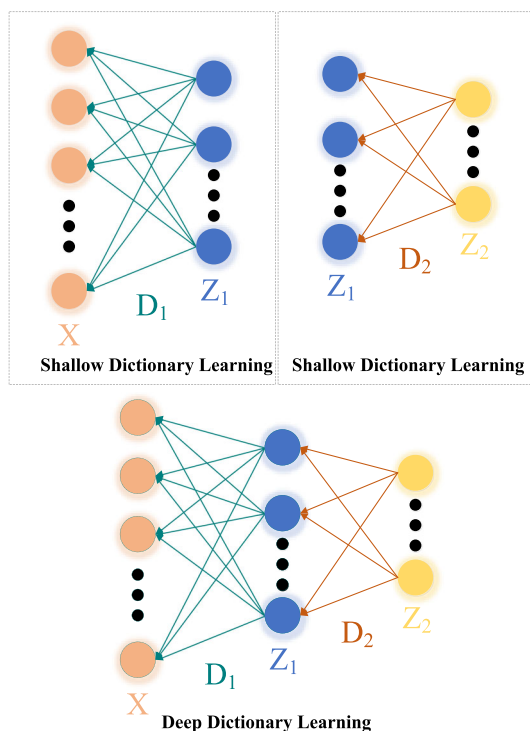


Fig. 1 Schematic diagram of two-layer deep dictionary learning

use the  $l_1$  norm of coefficients to constrain or regularize the least squares (LS) problem, resulting in a high computational capacity requirement for the algorithm. In order to solve this problem, we can reformulate the problem as the least square regression of  $l_2$  norm regularization of coefficients. Therefore, this paper uses Anchored Neighborhood Regression (ANR) to transform the problem into solving the following optimization problems:

$$\min_{\beta} \|y - N_l\beta\|_2^2 + \lambda\|\beta\|_2 \tag{4}$$

where,  $y$  represents the input LR patches feature, and  $N_l$  represents the neighborhood of LR dictionary space,  $\beta$  Represents the sparse coefficient,  $\lambda$  Is the regularization coefficient. The algebraic solution of the optimization problem can be expressed as:

$$\beta = (N_l^T N_l + \lambda I)^{-1} N_l^T y \tag{5}$$

Since the sparse coefficients of the LR dictionary and the HR dictionary are shared, that is, their sparse coefficients are the same, then the reconstructed HR patches can be written as:

$$x = N_h\beta \tag{6}$$

where  $x$  is the reconstructed HR patches, and  $N_h$  is the neighborhood of the HR dictionary space corresponding to  $N_l$ .

For the atoms in the learned dictionary, each atom selects its  $K$ -nearest neighbor atoms, and define these neighbor atoms as the neighborhood of the atom. Once the neighborhood is defined, a separate projection matrix  $P_j$  can be calculated for each dictionary atom  $d_j$  according to its neighborhood. Then, the super-resolution problem can be solved by calculating the adjacent atom  $d_j$  of each input patches feature  $y$  in the dictionary, and then using the stored projection matrix  $P_j$  to map it to the HR space:

$$x = P_j y \quad (7)$$

This method is an approximation of NE method with low complexity, so it can greatly reduce the execution time and time complexity of the algorithm.

### 2.3 Adjust anchored neighborhood regression(A +)

In ANR and other sparse coding methods, training samples are only used in the dictionary construction (learning) stage, while the adjusted anchored neighborhood regression (A +) algorithm is different. When looking for the neighborhood of each atom (calculated in the training phase and used in the reconstruction phase), the neighborhood used for regression is obtained from the training sample set. Therefore, training samples in the reconstruction phase is also very important. The ANR method is used again to conduct regression during training. The difference is that the neighborhood is defined based on training samples instead of sparse dictionary atoms. Therefore, the optimization problem is transformed into:

$$\min_{\delta} \|y - S_l \delta\|_2^2 + \lambda \|\delta\|_2 \quad (8)$$

Here, the matrix  $S_l$  is used instead of  $N_l$  as the neighborhood of the atom, which represents the  $K$  closest training samples to the dictionary atoms that matching the input patches  $y$ .

## 3 Proposed scheme

In this section, we introduce the details of the proposed scheme. The method learns the required high- and low-resolution dictionary from the training set through deep dictionary learning, and selects dictionary atomic pairs to calculate the projection matrix. Finally, the test image is read and reconstructed through the dictionary and projection matrix. The symbols involved are shown in Table 1.

**Table 1** The notations used in this paper

Notations	Description
$X$	LR training data
$Y$	HR training data
$D$	Learned dictionary
$Z$	Sparse coefficient
$D_1$	The first layer dictionary
$Z_1$	The first layer coefficients
$\lambda$	Regularization factor
$I$	Regularization term <sub><math>i</math></sub>
$\varphi$	Activation function
$D_2$	The second layer dictionary
$Z_2$	The second layer coefficients
$Z_{N-1}$	The (N-1)th layer coefficients
$D_N$	The Nth layer dictionary
$Z_N$	The Nth layer coefficients
$D_l$	LR dictionary
$D_h$	HR dictionary
$S_l$	Neighborhood of LR dictionary atoms
$S_h$	Neighborhood of HR dictionary atoms
$\beta$	Sparse coefficients when reconstructing
$P_j$	Projection matrix

### 3.1 Pretreatment

The image super-resolution algorithm usually divides the image into blocks in the spatial domain, carries out super-resolution reconstruction on the image patches, fuses the reconstructed images of every patch, and then, the super-resolution reconstruction is completed. The spatial characteristics of the original image patches cannot represent the image well. The common practice is to subtract the average value and divide it by the standard deviation to normalize the image contrast. The commonly used image features are the first and second derivatives of image patches. The performance of these two feature types seems similar, while Bevilacqua et al. [31] showed that the performance of using only the average subtraction is slightly better than that of using only the first derivative and second order derivative. In this paper, the same features as Zeyde et al. [18] are used to extract the first and second order features of an image patch, and PCA is applied to reduce the dimensions.

We subtract the LR image bicubically interpolated image from the HR image to create a normalized HR block. The patches generated from the SR process are added to the bicubic interpolated image of the LR input image (overlapping pixel values are averaged) to reconstruct the output.

### 3.2 Dictionary construction

Dictionary construction is critical to the performance of any SR reconstruction method with sparse representations. Generally speaking, the performance of the algorithm is positively correlated with the size of the dictionary, that is, the larger the dictionary, the better the effect, but this also brings higher computational costs. The LR input image itself can be used to build a dictionary, in which case there will be an "inner" dictionary [32]. However, many methods prefer to use different images to build an "external" dictionary beyond the input query.

In the dictionary construction process of this paper, the same set of training samples are used as those used by Zeyde et al. [23] and Yang et al. [17]. For the dictionary learning method, this paper draws on the deep dictionary learning method proposed by TARIYAL et al. [29].

For single-layer dictionaries,  $\mathbf{D}$  and  $\mathbf{Z}$  can be obtained by solving the following optimization problems:

$$\min_{\mathbf{D}, \mathbf{Z}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 \quad (9)$$

For the problem of sparse representation, the purpose is to learn the basis that can represent samples in a sparse way, that is,  $\mathbf{Z}$  is required to be sparse. The most commonly used algorithm to solve this problem is K-SVD algorithm. Fundamentally, it solves the following optimization problems:

$$\min_{\mathbf{D}, \mathbf{Z}} \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 \text{ s.t. } \|\mathbf{Z}\|_0 \leq \varepsilon \quad (10)$$

For deep dictionary learning, as shown in Eq. (3), the proposed method uses the greedy learning method to learn one layer at a time, that is, first learn the dictionaries and coefficients of the first layer:

$$\min_{\mathbf{D}_1, \mathbf{Z}_1} \|\mathbf{X} - \mathbf{D}_1\mathbf{Z}_1\|_F^2 \quad (11)$$

$\mathbf{Z}_1$  is not required to be sparse, except for the coefficient of the last layer.

The above optimization problems can be solved by alternately solving  $\mathbf{D}_1$  and  $\mathbf{Z}_1$ , namely:

$$\mathbf{Z}_1 = (\mathbf{D}_1^T \mathbf{D}_1 + \lambda \mathbf{I})^{-1} \mathbf{D}_1^T \mathbf{X} \quad (12)$$

$$\mathbf{D}_1 = \mathbf{Z}_1 \mathbf{X}^T (\mathbf{Z}_1 \mathbf{Z}_1^T + \lambda \mathbf{I})^{-1} \quad (13)$$

alternate the above two processes to solve.

For the second layer, we need to solve the following problems:

$$\min_{\mathbf{D}_2, \mathbf{Z}_2} \|\varphi^{-1}(\mathbf{Z}_1) - \mathbf{D}_2 \mathbf{Z}_2\|_F^2 \quad (14)$$

it can also be solved by alternately solution, that is:

$$\mathbf{Z}_2 = (\mathbf{D}_2^T \mathbf{D}_2 + \lambda \mathbf{I})^{-1} \mathbf{D}_2^T \varphi^{-1}(\mathbf{Z}_1) \quad (15)$$

$$\mathbf{D}_2 = \mathbf{Z}_2 \varphi^{-1}(\mathbf{Z}_1)^T (\mathbf{Z}_2 \mathbf{Z}_2^T + \lambda \mathbf{I})^{-1} \quad (16)$$

solve alternately until the last layer:

$$\min_{\mathbf{D}_N, \mathbf{Z}} \|\varphi^{-1}(\mathbf{Z}_{N-1}) - \mathbf{D}_N \mathbf{Z}_N\|_F^2 + \lambda \|\mathbf{Z}_N\|_0 \quad (17)$$

Among them, adding a regularization term requires the last layer coefficient  $\mathbf{Z}_N$  to be sparse. The last layer can be solved using K-SVD method.

First, use deep dictionary learning to learn the final LR dictionary through LR image patches, and the resulting final LR dictionary can be expressed as:

$$\mathbf{D}_l = \mathbf{D}_1 \varphi(\mathbf{D}_2 \varphi(\dots \varphi(\mathbf{D}_N))) \quad (18)$$

where  $\varphi$  is the nonlinear activation function. Then, by forcing  $\mathbf{D}_h$  to share the same sparse coefficients as the last layer in  $\mathbf{D}_l$ ,  $\mathbf{D}_h$  can be built from the sparse coefficients and HR patches, which can be calculated by the following expression:

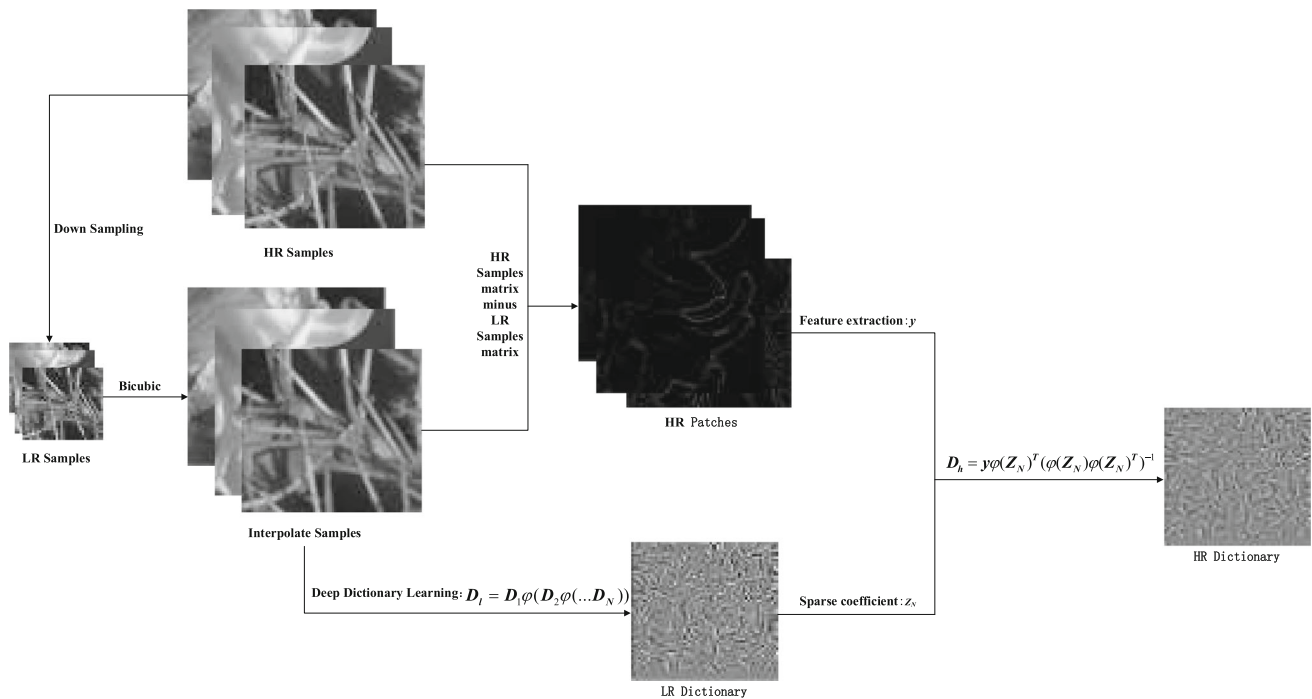
$$\mathbf{D}_h = \mathbf{Y} \varphi(\mathbf{Z}_N)^T (\varphi(\mathbf{Z}_N) \varphi(\mathbf{Z}_N)^T)^{-1} \quad (19)$$

where  $\mathbf{y}$  represents the features of the HR patches, and  $\mathbf{Z}_N$  is the sparse coefficient learned by the last layer of dictionary learning. The process of dictionary construction is shown in Fig. 2.

### 3.3 Sampling anchor neighborhoods

In order for robust regressors to be anchored to dictionary atoms, we need to have a sample neighborhood centered on the atom, in the sense of Euclidean distance, or up to the unit l2 norm, on the surface of the unit hypersphere. The local manifold on the unit hypersphere can be approximated by a set of adjacent samples. At the same time, since we have a consistent scale factor between the LR feature and its HR patch at the time of feature extraction, when we bring the LR patch feature into the hypersphere by l2 normalization, we must also pass the same factor by the corresponding HR patches are scaled to linearly transform which aims to preserve the relationship between LR and HR spaces.

In the LR dictionary, each atom finds several atoms closest to it to form the neighborhood  $\mathbf{S}_l$ , these atoms are extracted from the LR patches in the training set, and the distance metric used for the nearest neighbor atom search is the anchor Euclidean distance between atoms and training samples; in the HR dictionary, the corresponding atoms are also extracted to form the neighborhood  $\mathbf{S}_h$ , which is also extracted from the



**Fig. 2** dictionary building

HR patches in the training set. Compute the projection matrix between these LR, HR atoms. After traversing all dictionary atoms, in the end, each atom has a corresponding group of neighbor atoms and a projection matrix. The calculation of the projection matrix will be described in detail in the next section.

### 3.4 Super-resolution reconstruction

Input the LR image to be processed and segment it with the same block size as the dictionary training stage, and then perform feature extraction on all the LR image segmentation blocks. For each LR feature block  $X$ , finding which class it is closest to in the neighborhood of the LR dictionary atoms can be expressed as the following optimization problem:

$$\min_{\beta} \|x - S_l\beta\|_2^2 + \lambda\|\beta\|_2 \quad (20)$$

where  $x$  represents the input LR image patches, and  $S_l$  represents the neighborhood of the LR dictionary atoms. Solve the above optimization problem to get the coefficient  $\beta$ . The algebraic form solution can be expressed as:

$$\beta = (S_l^T S_l + \lambda I)^{-1} S_l^T x \quad (21)$$

Directly use the projection matrix to which this class belongs to obtain the corresponding HR block  $Y$ . which is:

$$y = S_h\beta \quad (22)$$

where,  $y$  represents the output HR image patch, and  $S_h$  represents the neighborhood of the HR dictionary atom corresponding to  $S_l$ . which is:

$$y = S_h(S_l^T S_l + \lambda I)^{-1} S_l^T x \quad (23)$$

According to Eq. (7), the projection matrix can be expressed as:

$$P_j = S_h(S_l^T S_l + \lambda I)^{-1} S_l^T \quad (24)$$

after the sampling of the LR neighborhood and the HR neighborhood is completed, the projection matrix can be calculated.

The calculation is performed on all input image patches until all input low-resolution image patches have corresponding reconstructed high-resolution image patches. All the reconstructed high-resolution image blocks are pasted in reverse according to the previous segmentation coordinates (the average value of the overlapping area between blocks is calculated) to obtain the final high-resolution image. The super-resolution reconstruction process is shown in Fig. 3.

## 4 Experimental results and analysis

This section compares the proposed method with several typical image super-resolution reconstruction methods NE + LLE, Yang [17], Zeyde [18], ANR [19], GR [19], MCSR [33], A + (same dictionary size) [20] and CDDL[30]. At

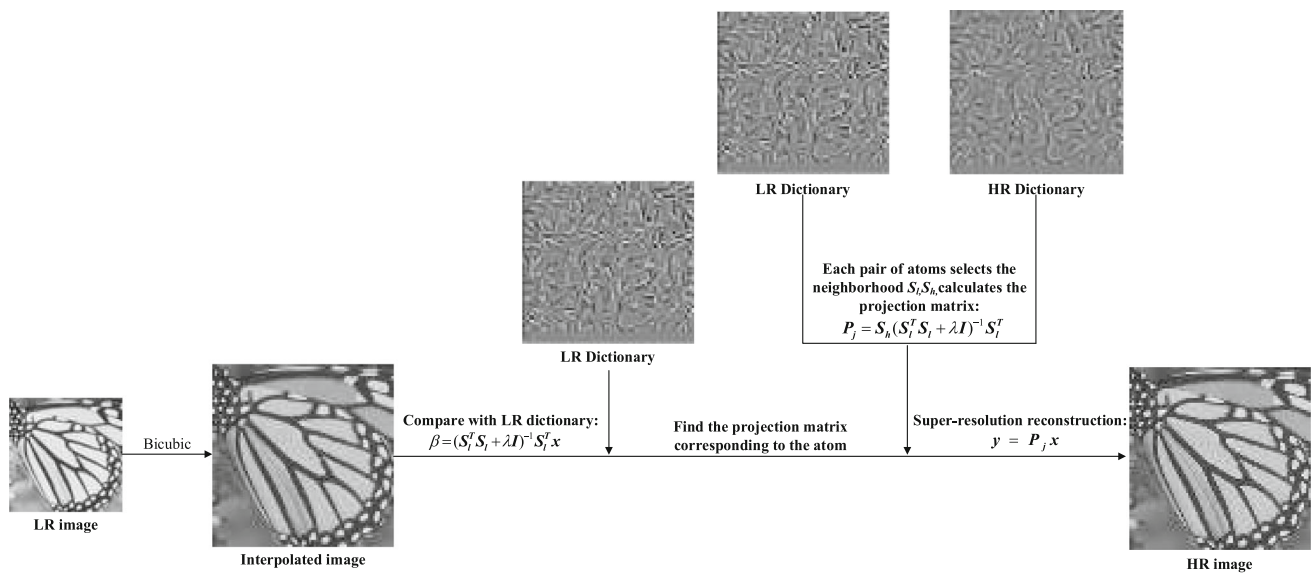


Fig. 3 super-resolution reconstruction

Table 2 PSNR/dB of each algorithm with different zoom scales on the Set5 dataset

Image	Scale	Yang	Zeyde	GR	ANR	NE + LLE	MCSR	A+	Our method
Baby		–	38.20	38.30	38.40	38.30	–	38.13	<b>38.44</b>
Bird		–	39.90	39.00	40.00	40.00	–	38.76	<b>41.08</b>
Butterfly	× 2	–	30.60	29.10	30.40	30.40	–	28.83	<b>31.98</b>
Head		–	35.60	35.60	35.60	35.60	–	35.51	<b>35.72</b>
Woman		–	34.50	33.70	34.50	34.50	–	33.57	<b>35.24</b>
Average		–	35.76	35.14	35.82	35.76	–	34.96	<b>36.49</b>
Baby		34.30	35.10	34.90	35.10	35.10	35.00	34.59	<b>35.10</b>
Bird		34.10	34.60	33.90	34.60	34.60	<b>35.40</b>	33.49	35.38
Butterfly	× 3	25.60	25.90	25.00	25.90	25.70	<b>28.80</b>	24.77	27.14
Head		33.20	33.60	33.50	33.60	33.60	33.60	33.3	<b>33.73</b>
Woman		29.90	30.40	29.70	30.30	30.20	<b>31.40</b>	29.36	31.12
Average		31.42	31.92	31.4	31.9	31.84	<b>32.80</b>	31.1	32.49
Baby		–	33.10	32.80	33.00	33.00	<b>32.90</b>	32.55	33.24
Bird		–	31.70	31.30	31.80	31.70	32.40	31.03	<b>32.58</b>
Butterfly	× 4	–	23.60	23.10	23.50	23.40	<b>25.90</b>	22.82	24.34
Head		–	32.20	32.10	32.30	32.20	32.30	31.98	<b>32.51</b>
Woman		–	27.90	27.40	27.80	27.70	<b>28.80</b>	27.17	28.64
Average		–	29.70	29.34	29.68	29.60	<b>30.50</b>	29.11	30.26

The best result in each row are shown in bold

the same time, our proposed method is compared with some methods in recent years, these methods are Huang’s method [29], recurrent residual regressor (RRR) [34], single image super-resolution (SISR) [35], based on deep learning-dilated convolution (DC) [36] and multi-scale encoder decoder (MSED) [37]. By using the most commonly used PSNR and SSIM in image evaluation indicators for objective evaluation. Where PSNR stands for peak signal-to-noise ratio and SSIM

for structural similarity. Its expressions are:

$$PSNR = 10 \log_{10} \left( \frac{255^2 \cdot MN}{\|\hat{x} - x\|^2} \right) \tag{25}$$

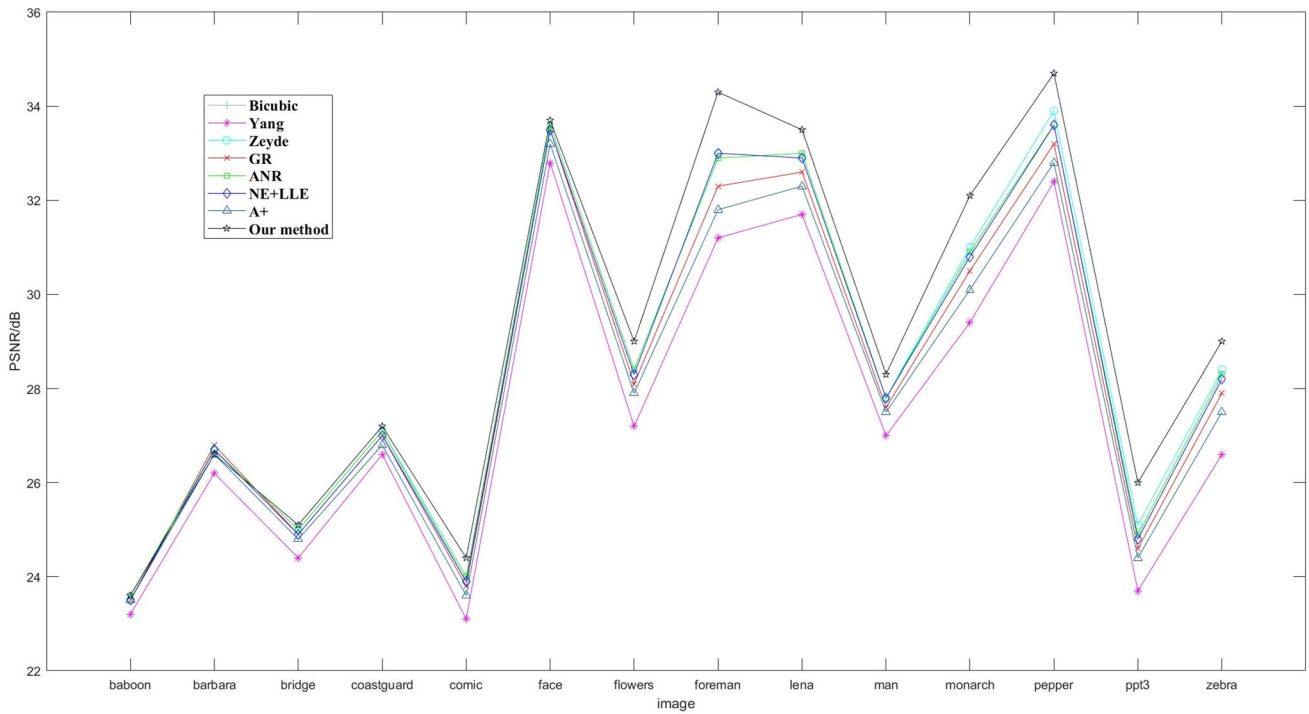
$$SSIM = \frac{(2\mu_{\hat{x}}\mu_x + C_1)(2\sigma_{\hat{x}x} + C_2)}{(\mu_{\hat{x}}^2 + \mu_x^2 + C_1)(\sigma_{\hat{x}}^2 + \sigma_x^2 + C_2)} \tag{26}$$

where  $\hat{x}$  represents the reconstructed image,  $x$  represents the original image, M and N represent the row and number of the

**Table 3** SSIM of each algorithm at different zoom scales on the Set5 dataset

Image	Scale	Yang	Zeyde	GR	ANR	NE + LLE	MCSR	A +	Our method
Baby		–	0.9362	<b>0.9648</b>	0.9645	0.9631	–	0.9628	0.9643
Bird		–	0.9837	0.9812	0.9844	0.9834	–	0.9804	<b>0.9862</b>
Butterfly	× 2	–	0.9532	0.9297	0.9514	0.9509	–	0.9290	<b>0.9643</b>
Head		–	0.8818	<b>0.8848</b>	0.8838	0.8829	–	0.8813	0.8847
Woman		–	0.9647	0.9604	0.9656	0.9648	–	0.9595	<b>0.9690</b>
Average		–	0.9493	0.9442	0.9499	0.9490	–	0.9426	<b>0.9537</b>
Baby		0.9043	0.9211	0.9216	<b>0.9225</b>	0.9210	0.9191	0.9155	0.9218
Bird		0.9391	0.9477	0.9404	0.9488	0.9477	0.9518	0.933	<b>0.9543</b>
Butterfly	× 3	0.8611	0.877	0.8316	0.8717	0.8705	<b>0.9213</b>	0.8250	0.9076
Head		0.8024	0.8204	0.8223	0.8239	0.8226	0.8230	0.8166	<b>0.8262</b>
Woman		0.9037	0.9176	0.9048	0.9168	0.9161	0.9280	0.8969	<b>0.9282</b>
Average		0.8821	0.8968	0.8841	0.8968	0.8956	<b>0.9080</b>	0.8774	0.9076
Baby		–	0.8797	0.8782	0.8811	0.8797	0.8716	0.8722	<b>0.8835</b>
Bird		–	0.8999	0.8908	0.9018	0.9004	0.9051	0.8842	<b>0.9136</b>
Butterfly	× 4	–	0.7965	0.7483	0.7887	0.7850	<b>0.8641</b>	0.7395	0.8365
Head		–	0.7736	0.7734	0.7762	0.7751	0.7774	0.7685	<b>0.7825</b>
Woman		–	0.8643	0.8459	0.8619	0.8608	<b>0.8913</b>	0.8390	0.8836
Average		–	0.8428	0.8237	0.8419	0.8402	<b>0.8628</b>	0.8207	0.8599

The best result in each row are shown in bold



**Fig. 4** PSNR at scale 3 on set14 dataset



**Table 4** Comparison with recent methods(zoom scale 2)

Images	PSNR							SSIM					
	RRR	SISR	Huang	DC	MSED	CDDL	Our method	RRR	SISR	DC	MSED	CDDL	Our method
<i>set5</i>													
Baby	36.47	37.16	38.21	38.38	38.34	38.38	<b>38.50</b>	0.84	0.86	<b>0.96</b>	<b>0.96</b>	<b>0.96</b>	<b>0.96</b>
Bird	36.86	37.46	39.12	38.76	38.76	40.60	<b>41.22</b>	0.91	0.98	0.98	0.96	0.98	<b>0.99</b>
Butterfly	26.85	30.14	29.17	32.50	31.99	<b>32.88</b>	32.03	0.84	0.85	0.96	0.93	<b>0.97</b>	0.96
Head	33.55	34.67	35.53	35.74	35.61	<b>35.78</b>	35.75	0.81	0.80	0.87	0.87	<b>0.88</b>	<b>0.88</b>
Woman	32.16	34.01	33.85	35.26	35.17	<b>35.60</b>	35.30	0.86	0.97	0.97	0.96	<b>0.98</b>	0.97
Average	33.18	34.69	35.18	36.13	35.97	<b>36.65</b>	36.56	0.85	0.89	<b>0.95</b>	0.94	<b>0.95</b>	<b>0.95</b>
<i>set14</i>													
Baboon	23.28	24.07	–	25.16	25.29	25.24	<b>25.62</b>	0.67	0.68	0.74	0.75	<b>0.76</b>	<b>0.76</b>
Barbara	26.19	28.02	–	28.23	28.57	28.53	<b>28.73</b>	0.79	0.78	0.86	0.87	<b>0.88</b>	<b>0.88</b>
Brigde	26.14	27.11	–	27.63	27.40	27.66	<b>27.75</b>	0.75	0.74	0.84	0.84	<b>0.85</b>	<b>0.85</b>
Coastguard	28.30	29.19	–	30.51	30.43	<b>30.69</b>	30.56	0.75	0.75	0.84	0.85	<b>0.86</b>	0.84
Comic	26.33	27.05	–	28.30	27.53	<b>28.74</b>	28.24	0.81	0.81	0.91	0.89	<b>0.92</b>	0.91
Face	34.04	34.65	–	35.77	35.69	<b>35.79</b>	35.72	0.80	0.80	<b>0.88</b>	<b>0.88</b>	<b>0.88</b>	<b>0.88</b>
Flower	30.41	31.10	–	33.38	33.75	<b>33.78</b>	33.02	0.85	0.84	0.93	0.92	<b>0.95</b>	0.93
Foreman	31.71	31.88	–	34.59	34.50	34.58	<b>37.03</b>	0.81	0.85	0.96	0.95	0.96	<b>0.97</b>
Lena	33.47	34.55	–	36.59	36.56	<b>36.99</b>	36.58	0.82	0.83	0.93	0.93	<b>0.94</b>	0.93
Man	29.50	30.23	–	30.99	30.28	<b>31.25</b>	30.85	0.79	0.78	0.87	0.87	<b>0.89</b>	<b>0.89</b>
Monarch	33.40	35.22	–	37.64	36.51	<b>37.90</b>	37.00	0.87	0.87	0.97	0.97	<b>0.98</b>	<b>0.98</b>
Pepper	33.37	34.10	–	35.52	34.98	35.74	<b>37.00</b>	0.81	0.82	<b>0.92</b>	0.91	<b>0.93</b>	0.92
2ppt3	26.38	28.04	–	30.45	30.30	<b>30.26</b>	30.08	0.80	0.80	0.97	0.95	0.97	<b>0.98</b>
Zebra	30.74	31.32	–	33.44	33.45	33.44	<b>33.66</b>	0.83	<b>0.94</b>	<b>0.94</b>	0.92	0.91	<b>0.94</b>
Average	29.52	30.47	–	32.01	31.80	32.19	<b>32.27</b>	0.80	0.81	0.90	0.89	<b>0.91</b>	0.90

The best result in each row are shown in bold

image,  $\mu$  represents the mean,  $\sigma^2$  represents the variance,  $\sigma$  represents the covariance, and  $C_1$  and  $C_2$  are constants. The larger the PSNR value, the better the reconstructed image quality. The closer the SSIM value is to 1, the more similar the reconstructed image is to the original image.

The datasets used in the experiment set5, set14, BSDS100, Manga109 and Urban100. Set5 and Set14 datasets are low-complexity single image super-resolution data sets based on non-negative neighborhood embedding. These two training sets are used for single image super-resolution reconstruction, that is, to reconstruct HR images based on LR images to obtain more details. BSDS100 is a data set containing 100 natural images, Urban100 is a data set containing 100 Urban landscape images, and Manga109 is a data set containing 109 comic images. These three data sets are very challenging data. sets because they contain different scenes and genres.

#### 4.1 Experimental parameters

The experimental platform is Intel Core i7-11800H@2.30 GHz, the operating system is 64bit

Windows10 Professional Edition, Matlab R2019a. The dictionaries used in this experiment are two-layer dictionaries. When the magnification is 2, 3 and 4, the size of the deep dictionaries are  $2048 \times 1024 \times 512$ ,  $4096 \times 2048 \times 1024 \times 512$ , and  $8192 \times 4096 \times 2048 \times 1024$ , respectively. The neighborhood size  $K = 2048$ .

#### 4.2 Experimental results and analysis

Tables 2 and 3 give the quantitative experimental results on the set5 dataset. Figure 4 shows the experimental results of the method proposed in this paper when the scale is 3 on the set14 dataset.

Through the experimental results, it is not difficult to find that compared with some classic methods, the proposed method in this paper has obvious improvements in PSNR and SSIM. Table 4 shows the comparison results of the proposed method with recent methods when the zoom scale is 2. It can be seen that compared with some methods in recent years, our method is also very competitive in PSNR, and in terms

**Table 5** PSNR and SSIM on the BSDS100 dataset, Manga109 dataset, and Urban100 dataset

Dataset	Scale		Zeyde	GR	ANR	NE + LLE	A +	MSED	CDDL	Our method
BSDS100	× 2	PSNR	29.57	29.57	29.58	29.57	31.18	31.42	31.66	<b>31.66</b>
		SSIM	0.8436	0.8436	0.8436	0.8436	0.8873	0.8938	<b>0.9003</b>	0.9001
	× 3	PSNR	27.22	27.23	27.22	27.22	28.30	28.56	28.70	<b>28.72</b>
		SSIM	0.7390	0.7394	0.7391	0.7390	0.7834	0.7846	0.7860	<b>0.7862</b>
	× 4	PSNR	25.98	25.99	25.98	25.98	26.84	27.03	<b>27.21</b>	27.20
		SSIM	0.6682	0.6688	0.6684	0.6682	0.7090	0.7103	<b>0.7130</b>	0.7123
Manga109	× 2	PSNR	30.85	30.83	30.86	30.85	35.22	35.59	<b>35.71</b>	35.70
		SSIM	0.9345	0.9344	0.9346	0.9345	0.9664	0.9691	0.9699	<b>0.9702</b>
	× 3	PSNR	26.96	26.99	26.99	26.96	29.28	29.57	29.65	<b>29.66</b>
		SSIM	0.8546	0.8554	0.8554	0.8547	0.9101	0.9218	0.9336	<b>0.9342</b>
	× 4	PSNR	24.94	24.95	24.94	24.94	27.02	27.09	27.31	<b>27.34</b>
		SSIM	0.7858	0.7859	0.786	0.7858	0.8471	0.8499	0.8510	<b>0.8522</b>
Urban100	× 2	PSNR	26.90	26.89	26.91	26.90	29.18	29.44	29.52	<b>29.55</b>
		SSIM	0.8405	0.8405	0.8407	0.8405	0.8927	0.8986	0.9004	<b>0.9011</b>
	× 3	PSNR	24.48	24.48	24.48	24.48	26.04	26.22	26.36	<b>26.38</b>
		SSIM	0.7353	0.7354	0.7354	0.7353	0.7979	0.8088	<b>0.8120</b>	0.8112
	× 4	PSNR	23.16	23.17	23.17	23.16	24.35	24.50	24.58	<b>24.62</b>
		SSIM	0.6583	0.6585	0.6584	0.6582	0.7192	0.7211	<b>0.7234</b>	0.7225

The best result in each row are shown in bold

of SSIM, our method has achieved good results on most of the test pictures. Table 5 gives the quantitative experimental results on the BSDS100, Manga109 and Urban100 datasets.

Figures 5 and 6 show the visualization experimental results of the super-resolution reconstructed part of the image. It can be seen from the figure that the reconstruction result of bicubic interpolation is the worst, not only quite blurry, but also missing a lot of high-frequency detail information, and the visual effect is not good compared with the original image; the algorithm of Yang et al. has improved performance and reconstructed the original image, and the image retains most of the information of the original image, but the edge of the reconstructed image is not clear enough, and the overall effect is not particularly good; the edge information of the reconstructed image of Zeyde's algorithm is better than that of Yang's, but the visual effect of the reconstructed image is slightly poor; the reconstructed image of the NE + LLE algorithm has good detailed information, but the graininess is strong, which affects the vision; the A + method has the same dictionary size as the method in this paper, the reconstruction effect is not particularly ideal, and some places are more Blurred; the proposed method in this paper combines A + and deep dictionary learning algorithm to reconstruct the image with super-resolution, so that the edge of the reconstructed image is clear, and the visual effect is better than the previous methods. We have enlarged an area in the results, and the enlarged area is shown as a box in the figure. The enlarged image of Fig. 5 shows that the proposed method not only has

fewer artifacts but also has a clearer reconstruction effect than other methods. The enlarged image of Fig. 6 shows that the proposed method better removes jagged edges that appear in other methods, and the visual effect is also better.

### 4.3 Ablation experiment

We conducted some parameter-related experiments on the proposed method. First, we conducted experiments with different number of layers and different dictionary sizes to observe the effect. Figure 7a shows the PSNR of different  $\lambda$  on the set5 data set at a magnification factor of 2 and a dictionary size of  $1024 \times 512 \times 256$ . The spacing of  $\lambda$  is set to 0.05. It can be seen that as  $\lambda$  increases, the PSNR start to increases and then gradually decreases. It can be inferred that the PSNR reaches the highest peak when  $\lambda$  is around 0.15. Figure 7b shows the PSNR on the set5 data set when the amplification factor is 2, the first and last dictionary sizes are 1024 and 256 respectively, and  $\lambda$  is 0.1, and the number of dictionary layers is different. As the number of intermediate layers increases, PSNR shows a downward trend after being flat at the beginning. It is speculated that too many layers will cause over-fitting. Figure 7c shows the PSNR on the set5 dataset when the magnification is 2,  $\lambda$  is 0.1, the first layer dictionary size is 2048, the number of dictionary layers is different, and the last layer dictionary size is different. It can be seen that the PSNR gradually decreases because the

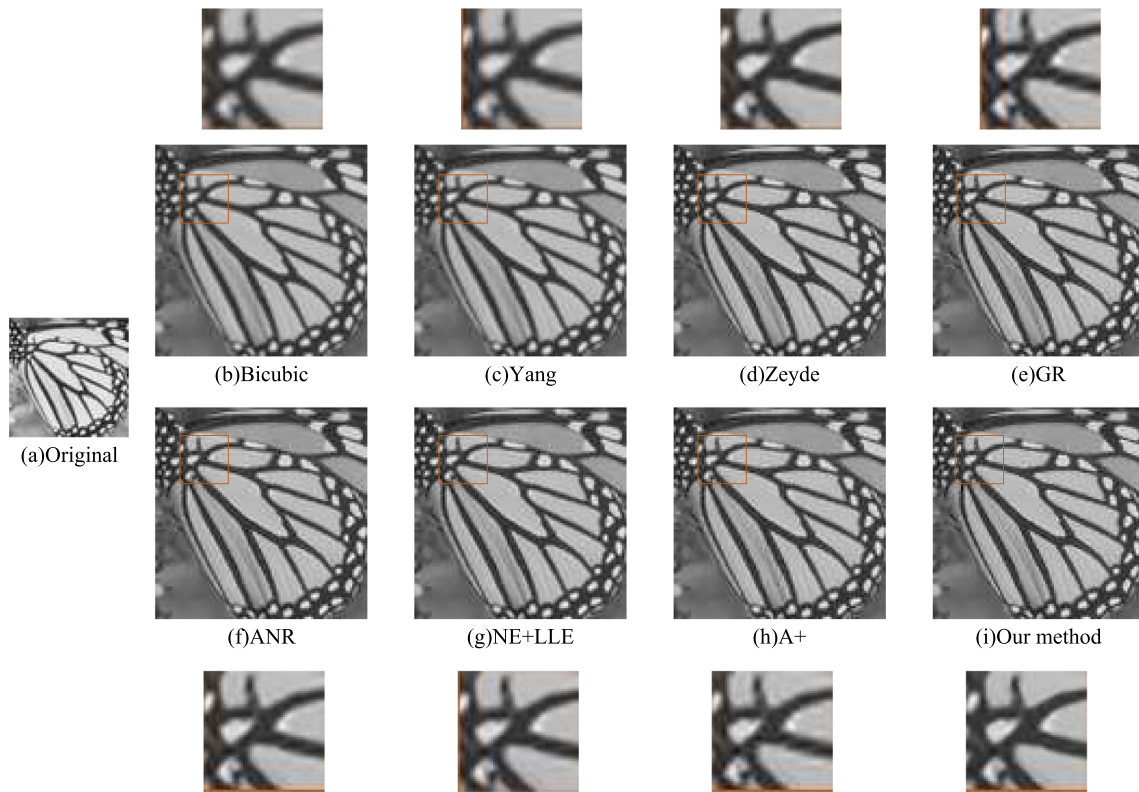


Fig. 5 The reconstruction effect of each algorithm of butterfly when the zoom scale is 3

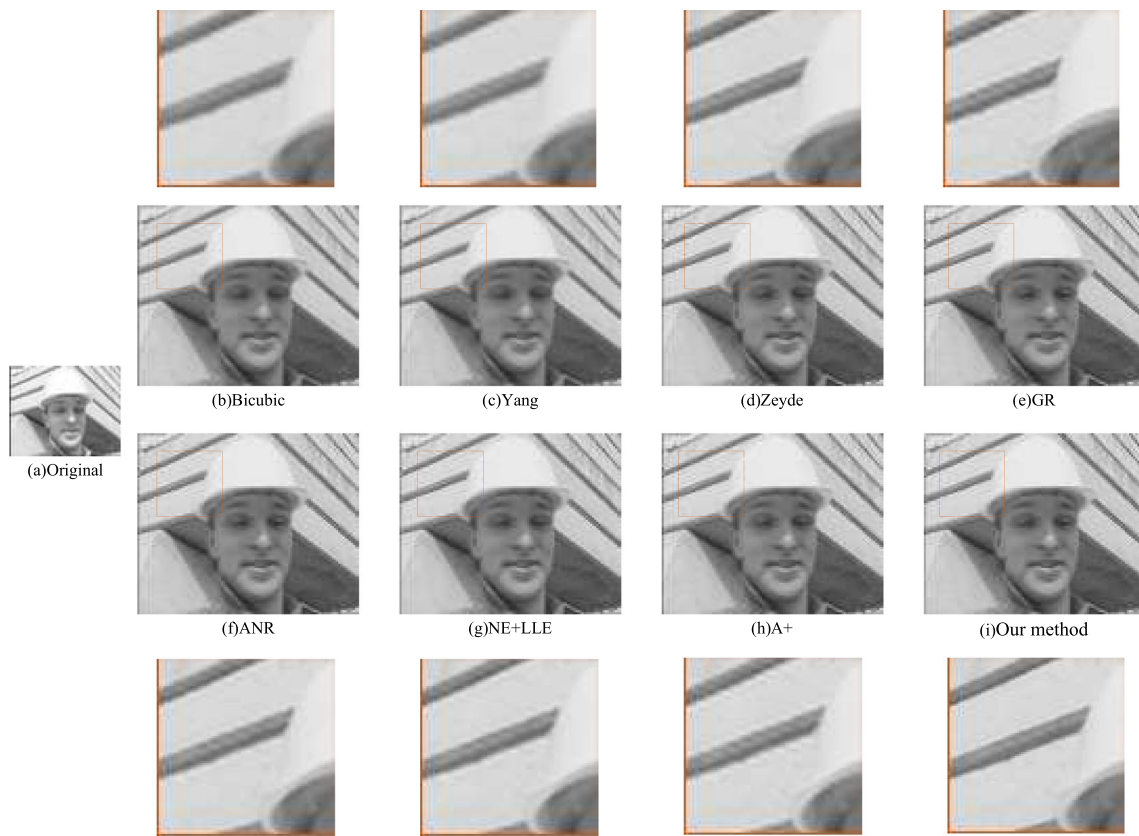
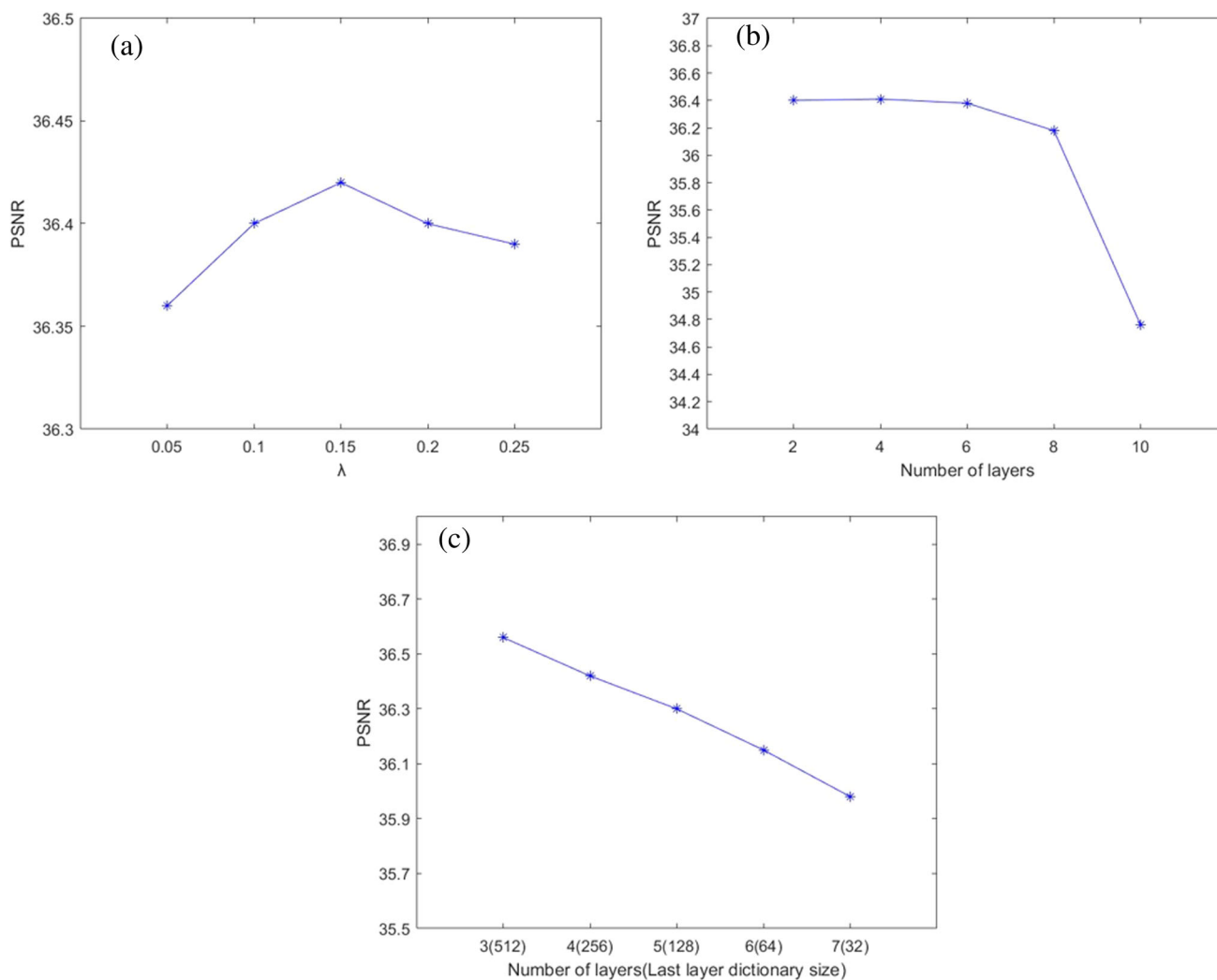


Fig. 6 The reconstruction effect of each algorithm of foreman when the zoom scale is 3



**Fig. 7** **a** PSNR with different  $\lambda$  (top left), **b** PSNR with different number of layers (top right), **c** PSNR with different number of layers and different tail layer dictionary sizes (bottom)

overall dictionary size gradually decreases, which reduces the reconstruction effect.

## 5 Conclusion

In view of the reason that the image super-resolution reconstruction algorithm ignores the image deep features, this paper proposed a deep dictionary model for image super-resolution, which can extract the image features to train the deep dictionary, and then carries out super-resolution reconstruction by adjust anchored neighborhood regression. The algorithm in this paper effectively reduces the size of the dictionary. The experimental results show that the reconstruction results of the algorithm in this paper are more ideal than those of multiple image super-resolution methods in the case of the same dictionary size, and have achieved good improvement in the objective evaluation criteria PSNR and SSIM.

**Author contributions** W.B. and Y.H. wrote the main manuscript text, B.J. and Z.Z. and W.L prepared all the figures and tables. All authors reviewed the manuscript.

**Funding** The work is partially supported by Natural Science Foundation of Anhui Provincial (Grant No. 2108085MF206) and National Natural Science Foundation of China (Grant No. 61976006).

**Availability of data and materials** The data come from the common datasets.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

- Li, X., Orchard, M.T.: New edge-directed interpolation. *IEEE Trans. Image Process.* **10**(10), 1521–1527 (2001)
- Wei, Z., Ma, K.K.: Contrast-guided image interpolation. *IEEE Trans. Image Process.* **22**(11), 4271–4285 (2013)

3. Sun, J., Xu, Z., Shum, H.Y.: Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE Trans. Image Process.* **20**(6), 1529–1542 (2010)
4. Xu, J., Sun, C., Huang, Y., Huang, X.: Residual neural network for filter kernel design in filtered back-projection for ct image reconstruction. In: *Bildverarbeitung für die Medizin 2021: Proceedings, German Workshop on Medical Image Computing, Regensburg, March 7–9, 2021*, pp. 164–169. Springer Fachmedien Wiesbaden. (2021)
5. Stark, H., Oskoui, P.: High-resolution image recovery from image-plane arrays, using convex projections. *JOSA A* **6**(11), 1715–1726 (1989)
6. Denisova, N.: Bayesian maximum-a-posteriori approach with global and local regularization to image reconstruction problem in medical emission tomography. *Entropy* **21**(11), 1108 (2019)
7. Ding, X., Li, Q., Lin, L., He, Q., Shao, Y.: Fast time-frequency manifold learning and its reconstruction for transient feature extraction in rotating machinery fault diagnosis. *Measurement* **141**, 380–395 (2019)
8. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008). IEEE
9. Chen, L., Fang, F., Wang, T., Zhang, G.: Blind image deblurring with local maximum gradient prior. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1742–1750 (2019)
10. Ma, C., Rao, Y., Cheng, Y., Chen, C., Lu, J., Zhou, J.: Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7769–7778, (2020)
11. Zha, Z., Yuan, X., Zhou, J., Zhu, C., Wen, B.: Image restoration via simultaneous nonlocal self-similarity priors. *IEEE Trans. Image Process.* **29**, 8561–8576 (2020)
12. Li, W., Liu, Y., Ou, X., Wu, J., Guo, L.: Enhancing image clarity: a non-local self-similarity prior approach for a robust Dehazing Algorithm. *Electronics* **12**(17), 3693 (2023)
13. Dong, W., Zhang, L., Shi, G., Li, X.: Nonlocally centralized sparse representation for image restoration. *IEEE Trans. Image Process.* **22**(4), 1620–1630 (2012)
14. Wen, F., Ying, R., Liu, Y., Liu, P., Truong, T.K.: A simple local minimal intensity prior and an improved algorithm for blind image deblurring. *IEEE Trans. Circuits Syst. Video Technol.* **31**(8), 2923–2937 (2020)
15. Chen, Y., He, W., Yokoya, N., Huang, T.Z.: Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition. *IEEE Trans. Cybern.* **50**(8), 3556–3570 (2019)
16. Zhang, Y.Q., Ding, Y., Liu, J., Guo, Z.: Guided image filtering using signal subspace projection. *IET Image Proc.* **7**(3), 270–279 (2013)
17. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
18. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: *International Conference on Curves and Surfaces*, pp. 711–730. Springer, Berlin, Heidelberg (2010)
19. Timofte, R., De Smet, V., & Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1920–1927 (2013)
20. Timofte, R., De Smet, V., Van Gool, L.: A+: adjusted anchored neighborhood regression for fast super-resolution. In: *Asian Conference on Computer Vision*, pp. 111–126. Springer, Cham (2014)
21. Pérez-Pellitero, E., Salvador, J., Torres-Xirau, I., Ruiz-Hidalgo, J., Rosenhahn, B.: Fast super-resolution via dense local training and inverse regressor search. In: *Asian Conference on Computer Vision*, pp. 346–359. Springer, Cham (2014)
22. Wei, Z., Xiaofeng, B., Fang, H., Jun, W., Abidi, M.: Fast image super-resolution algorithm based on multi-resolution dictionary learning and sparse representation. *J. Syst. Eng. Electron.* **29**(3), 471–482 (2018)
23. Zhang, Y., Liu, J., Yang, W., Guo, Z.: Image super-resolution based on structure-modulated sparse representation. *IEEE Trans. Image Process.* **24**(9), 2797–2810 (2015)
24. Tariyal, S., Majumdar, A., Singh, R., Vatsa, M.: Deep dictionary learning. *IEEE Access* **4**, 10096–10109 (2016)
25. Mahdizadehaghdam, S., Panahi, A., Krim, H., Dai, L.: Deep dictionary learning: A parametric network approach. *IEEE Trans. Image Process.* **28**(10), 4790–4802 (2019)
26. Song, J., Xie, X., Shi, G., Dong, W.: Multi-layer discriminative dictionary learning with locality constraint for image classification. *Pattern Recogn.* **91**, 135–146 (2019)
27. Tang, H., Liu, H., Xiao, W., Sebe, N.: When dictionary learning meets deep learning: deep dictionary learning and coding network for image recognition with limited data. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(5), 2129–2141 (2020)
28. Montazeri, A., Shamsi, M., Dianat, R.: MLK-SVD, the new approach in deep dictionary learning. *Vis. Comput.* **37**(4), 707–715 (2021)
29. Huang, J.J., Dragotti, P.L.: A deep dictionary model for image super-resolution. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6777–6781. (2018). IEEE
30. Singhal, V., Majumdar, A.: A adaptation approach to solve inverse problems in imaging via coupled deep dictionary learning. *Pattern Recogn.* **100**, 107163 (2020)
31. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)
32. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: *2009 IEEE 12th International Conference on Computer Vision* (pp. 349–356). IEEE.
33. Li, Y.F., Fu, R.D., Jin, W., Ji, N.: Image super-resolution method based on multi-channel convolution. *Chin. J. Image Graph.* **22**(12), 1690–1700 (2017)
34. Zhang, K., Wang, Z., Li, J., Gao, X., Xiong, Z.: Learning recurrent residual regressors for single image super-resolution. *Signal Process.* **154**, 324–337 (2019)
35. Chen, Y., Wang, J., Chen, X., Zhu, M., Yang, K., Wang, Z., Xia, R.: Single-image super-resolution algorithm based on structural self-similarity and deformation block features. *IEEE Access* **7**, 58791–58801 (2019)
36. Zhang, Z., Wang, X., Jung, C.: DCSR: dilated convolutions for single image super-resolution. *IEEE Trans. Image Process.* **28**(4), 1625–1635 (2018)
37. Liu, H., Fu, Z., Han, J., Shao, L., Hou, S., Chu, Y.: Single image super-resolution using multi-scale deep encoder–decoder with phase congruency edge map guidance. *Inf. Sci.* **473**, 44–58 (2019)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.