Wang et al. EURASIP Journal on Image and Video Processing (2021) 2021:16
https://doi.org/10.1186/s13640-021-00552-8

EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Single image super resolution based on multi-scale structure and non-local smoothing

Wenyi Wang[1†], Jun Hu[2†], Xiaohong Liu[3*], Jiying Zhao[2] and Jianwen Chen[1]

*Correspondence:
liux173@mcmaster.ca
†Wenyi Wang and Jun Hu contributed equally to this work.
[3]Department of Electrical and Computer Engineering, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada
Full list of author information is available at the end of the article

Abstract
In this paper, we propose a hybrid super-resolution method by combining global and local dictionary training in the sparse domain. In order to present and differentiate the feature mapping in different scales, a global dictionary set is trained in multiple structure scales, and a non-linear function is used to choose the appropriate dictionary to initially reconstruct the HR image. In addition, we introduce the Gaussian blur to the LR images to eliminate a widely used but inappropriate assumption that the low resolution (LR) images are generated by bicubic interpolation from high-resolution (HR) images. In order to deal with Gaussian blur, a local dictionary is generated and iteratively updated by *K*-means principal component analysis (K-PCA) and gradient decent (GD) to model the blur effect during the down-sampling. Compared with the state-of-the-art SR algorithms, the experimental results reveal that the proposed method can produce sharper boundaries and suppress undesired artifacts with the present of Gaussian blur. It implies that our method could be more effect in real applications and that the HR-LR mapping relation is more complicated than bicubic interpolation.

Keywords: Super-resolution, Sparse representation, Dictionary learning, K-PCA, Non-local mean

1 Introduction

The problem of enlarging images to the ones with bigger spatial size is regarded as image super-resolution (SR), which builds the mathematical relation between the low-resolution (LR) image and the high-resolution (HR) image. Although image SR is a frequent manipulation in image processing, this problem still remains challenging because it is under constrained and there is no closed form without extra constraints. The concept of image SR was first proposed and studied by Tsai and Huang in the 1980s [1]. Through the past three decades, varieties of SR algorithms and models have been developed to deal with this problem. Among them, single-image super-resolution (SISR) is an important branch, which enlarges an image based on the image itself as the observation. In general, the existing SISR algorithms can be classified into three categories: interpolation-based, reconstruction-based and learning-based SR.

The interpolation-based SR usually utilizes fixed function [2] or adaptive structure kernels [3, 4] to predict the missing pixels in HR grid. It assumes that the LR observations are degraded by down-sampling, and the unknown HR pixels can be estimated from their observed neighbors. Now, the interpolation-based methods are often used as the comparison baseline. However, considerable blurring and aliasing artifacts are often inevitable in the up-scaled images.

Different from the interpolation-based SR, the reconstruction-based methods [5, 6] refine the observation model. In addition to sampling, reconstruction-based SR assumes that the LR image is obtained by a series of degradations: down-sampling, blurring, and additive noise. Using different prior assumptions, this family of approaches are capable of enhancing the features of low-resolution images through a regularized cost function. As a result, they are able to produce sharper edges and clearer textures while removing the undesired artifacts. However, these methods are usually inadequate in producing novel details and perform unsatisfactory under high scaling factor.

Compared with the aforementioned methods, the learning-based SR is generally superior since it is capable of generating convincing novel details that are almost lost in the low-resolution image. Basically, these algorithms and models exploit the prior texture knowledge from extensive sample images to learn the underlying mapping relations between LR and HR images. During the past decades, numerous mapping formulations have been designed, the most representative methods include neighbor-based SR, regression-based SR, sparsity-based SR, and the ones using deep neural networks. Compared with neighbor-based [7] and regression-based SR [8] that always heavily rely on the quality and the size of sample images, the sparsity-based SR is capable of learning more compact dictionaries based on signal sparse representation. In this case, it has been widely studied for its superior performance in producing clear images with low computational complexity. Based on this idea, Yang et al. proposed a classic sparse model by training a joint dictionary pair in the sparse domain for image reconstruction [9]. Subsequently, they introduced a coupled dictionary training approach with two acceleration schemes which overcame the sparse coding bottleneck [10]. With the similar framework of [8], Zeyde et al. [11] provided the refinements in dictionary training and image optimization, so that it improves the efficiency and performance. Timofte et al. proposed an anchored neighborhood regression (ANR) model to reconstruct the LR image via neighbor-based dictionaries in a fast way and also introduced the global regression (GR) model for some extreme cases [12]. They subsequently introduced the A+ [13] method that improves the SR performance by combining ANR with SF. Shi et al. further used the anchored neighborhood as the image prior to the deep network [14]. However, [15] implies that a large dictionary may cause unstable HR restoration. Hence, the concept of sub-dictionary has been widely used. Dong et al. raised an adaptive sparse domain selection (ASDS) model with PCA-based sub-dictionary training [16]. Afterwards, they further applied this method on their non-locally centralized sparse representation (NCSR) model [17], which has been turned out to be one of the state-of-the-art SR algorithms which can deal with the blur effect during the down-sampling. In addition, based on Yang's framework in [9], Zhang et al. also utilized the PCA-based clustering to the sample image patches, so that multiple mapping relations can be trained [18]. In recent years, the deep neural networks have shown powerful capability in various computer vision tasks and brought significant benefit to image SR. To our best knowledge, Dong et al. [19] first proposed the deep learning-based SR method which is based on the CNN architecture. Afterwards, Kim et al. proposed to use a much deeper residual network to achieve superior performance [20]. In addition, Kim's method was capable of dealing with different enlarge ratios by expanding the training dataset to include the LR-HR patch pairs with different scaling factors. Although the aforementioned learning-based methods performed well on their test dataset, there was often an underlying assumption that the LR image was directly down-sampled from the HR image by bicubic interpolation method, which might not be the actual case in real scenarios that the blur effect could happen along with the down-sampling. And there is no sufficient evidence that can prove that the learning model is still valid if it is trained with the bicubic interpolation assumption.

In this paper, a novel learning-based SR method is proposed based on hybrid dictionary training in the sparse domain [21]. In our proposed algorithm, multiple global dictionary pairs are trained from a large natural image dataset. Each global dictionary pair is trained to reveal the general mapping relation between the LR and the HR images under different scaling factors. In addition to the global dictionaries trained from general dataset, we also predict a local dictionary along with a self-similarity metric based on the input LR image. Since the local dictionary is more consistent with the input image, the reconstructed HR result can be more robust and the blur effect during the down-sampling could be suppressed. Given the dictionaries and an LR image, we iteratively enhance the details of LR image. In each iteration, the appropriate global dictionary is chosen according to the current LR image quality. Afterwards, the undesired artifacts are suppressed by self-similarity prior and non-locally centralized constraints based on local dictionary. The main contributions of this paper are as follows: (1) we combine the global and local dictionary to estimate the HR image so that the result could benefit from the large training dataset and the blur effect during the down-sampling could be also suppressed; (2) we apply multiple global dictionaries from multi-scale structures to iteratively improve the HR image quality.

The remained part of this paper is organized as follows. In Section 2, we will provide a brief introduction to the related works of sparsity-based SR and dictionary training methodologies. In Section 3, our proposed SR method will be described in detail, including the multi-scale global dictionary training, K-PCA-based local dictionary training, and the single image super-resolution based on the hybrid dictionaries. In Section 4, the results and comparisons will be represented to prove that our proposed method can generate state-of-the-art HR images with the present of Gaussian blur effect. Finally, the conclusion will be drawn in Section 5.

2 Related work

Based on the observed LR image, single image super-resolution can be modeled by Eq. (1):

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}, \tag{1}$$

where $\boldsymbol{H}$ is the degradation operator that combines down-sampling and blurring, $\boldsymbol{x}$ is the HR image, $\boldsymbol{v}$ is the additive noise, and $\boldsymbol{y}$ is the observed LR image.

2.1 Image SR based on sparse representation

Since the problem of solving $\boldsymbol{x}$ from Eq. (1) is ill-posed, the HR recovery from LR image is uncertain and the artifacts will be introduced. In this case, researchers proposed and studied different regularization terms such as total variation (TV) [22, 23], the non-local similarity [24], and the sparsity representation [25, 26]. The total variation suppresses the artifacts at the expense of over-smoothing the texture. In order to maintain the discontinuities or spatially inhomogeneous, the sparsity-based regularization is adopted in many recent single image super resolution methods [9, 16, 17, 27]. According to the sparse representation, the observation model in Eq. (1) can be re-written as:

$$y_i = \boldsymbol{H}D\alpha_i + v_i, \tag{2}$$

where $x_i = D\alpha_i$ is a patch of the high resolution image, D is called as the dictionary, α_i is the sparse code of patch x_i, y_i is the observed low-resolution patch of x_i, and v_i represents the additive noise.

Based on the observation model, the sparsity-based image super-resolution can be formulated by Eq. (3):

$$\boldsymbol{\alpha}_{\boldsymbol{y}} = \arg\min_{\alpha} \left\{ \|\boldsymbol{y} - \boldsymbol{HD\alpha}\|_2^2 + \lambda R(\boldsymbol{\alpha}) \right\}, \tag{3}$$

where $\|\boldsymbol{y} - \boldsymbol{HD\alpha}\|_2^2$ is called the fidelity term, $R(\boldsymbol{\alpha})$ is the sparsity based regularization term, and λ is the Lagrange multiplier which controls the balance between the fidelity and the regularization.

In order to reliably represent the image by using sparse code α, the dictionary choice becomes a critical issue. Generally, the dictionaries can be classified into two categories: analytical dictionary and learning-based dictionary. Analytical dictionaries such as the ones from DCT or Haar wavelet are easily generated, but they are not adaptive to diverse images. Learning-based dictionary is trained according to information from real natural images. Therefore, it contains more comprehensive characteristics which make the reconstruction precise. The learning-based dictionaries can be further divided into global dictionary and local dictionary. The global dictionary extracts texture from a large set of natural images which have abundant details. Therefore, it may reconstruct the details, which are almost lost in LR image, based on the experience drawn from other images. However, there are still risks to completely rely on global dictionary. The LR-HR mapping relation in global dictionary is learned from large scale database, which has its own LR-HR scaling method. Given an image, the HR estimation may fail if its scale and blur model does not fit with the image pairs in the database. This problem could be a worthy consideration especially when most of the learning based SR methods built their training dataset with the assumption that LR images is generated from HR ones by bicubic interpolation. On the other hand, the local dictionary is trained based on the observed LR image itself by using self-similarity and the feature statistics. Therefore, it is possible to use the local dictionary as supplementary information to enhance the SR image quality by handling the problem that LR image suffers from different degradation to the ones in training dataset.

2.2 Initial SR image for local dictionary training

As aforementioned, the quality of SR image estimated by global dictionary could be improved by using local dictionary to suppress the blur effect not learned from the global training dataset, and vice versa—the SR image from the global learning, which can be regarded as an initial guess or the side information to a local dictionary based method, can affect the final estimation of the high resolution image. In order to verify this, a simple example is given as follows. First, an initial HR estimation is made for an LR image

Table 1 PSNR(dB)/SSIM values of the SR results by using K-PCA and non-locally centralized sparse representation [17] with different initial HR values

Algorithm	Bicubic	Bilinear	Nearest neighbor	GR [12]	Yang's [9]
PSNR/SSIM	28.49/0.8894	28.42/0.8889	28.17/0.8869	28.79/0.8910	28.86/0.8916

by using 5 different SR algorithms: bicubic, bilinear, nearest neighbor, global regression (GR) [12], and SC-SR [9]. Based on these initial guess, the HR estimation is refined by applying the same local procedure—K-PCA and non-locally centralized sparse representation (NSCR) [17]. Table 1 presents the final HR estimation by using the local dictionaries trained from different initial HRs. It can be observed from Table 1 that the final HR estimation would be better if the initial HR value was better generated. Similar phenomenon was also found and analyzed in another recent paper [28]. Based on this observation, [28] proposed to use ridge regression based method to produce the initial HR image. In this paper, we propose to update the initial HR values in each iteration of the SR processing, so that the final result can be further improved. The details of our proposed SR method will be introduced in the following section.

3 Methods: image super-resolution based on multi-scale structure and non-local mean (MSNM-SR)

According to recent researches of single image super-resolution (SR), the sparsity representation has shown the advantages in recovering discontinuous and inhomogeneous image regions [9, 16, 17, 27, 29]. Therefore, our proposed SR method still uses sparse coding to represent the image features. As we mentioned in Section 2, the global dictionary is beneficial to provide comprehensive image structure, and the local dictionary is more relevant to the image to be enhanced. In this case, we propose our SR method that utilizes the global and local dictionaries together to generate high resolution images with clear texture and suppressed blurring artifact.

3.1 Overview of our proposed SR method

The flowchart of our proposed single image super-resolution is presented in Fig. 1. First, a set of global dictionary pairs $\left\{D_i^H, D_i^L | i = 1.1, 1.2, \ldots, 4\right\}$ is trained from a large amount of natural images. Here, i represents the upscaling ratio from LR to HR images, and D^H and D^L represent the LR and HR image dictionary, respectively. These global dictionary pairs are generated based on the assumption that LR and HR images share the same sparse representation. By training multiple dictionary pairs, multi-scale mapping relation between HR and LR images can be established.

Given a low resolution image $I_{\mathrm{LR}} \in \mathbb{R}^{m\times n}$, and the scale factor s, a high resolution image $I_{\mathrm{HR}} \in \mathbb{R}^{sm\times sn}$ will be gradually generated by our proposed SR method. First, the magnification factor s_i is initialized as $s_1 = s$. According to the value of s_i, the corresponding global dictionary pair $\left\{D_{s_i}^H, D_{s_i}^L\right\}$ is used to magnify the low resolution image. In order to suppress the artifacts and the noises introduced by sparse representation, a local dictionary $\left\{D_{s_i}^0\right\}$ is generated. Since $\left\{D_{s_i}^0\right\}$ is constructed based on the self-information of the image, this dictionary would be more consistent with the image content. Based on $\left\{D_{s_i}^0\right\}$, a sparse fidelity term and a non-local smoothing term are used as the constraints, so that the structure of the reconstructed HR image is similar to the original input image.

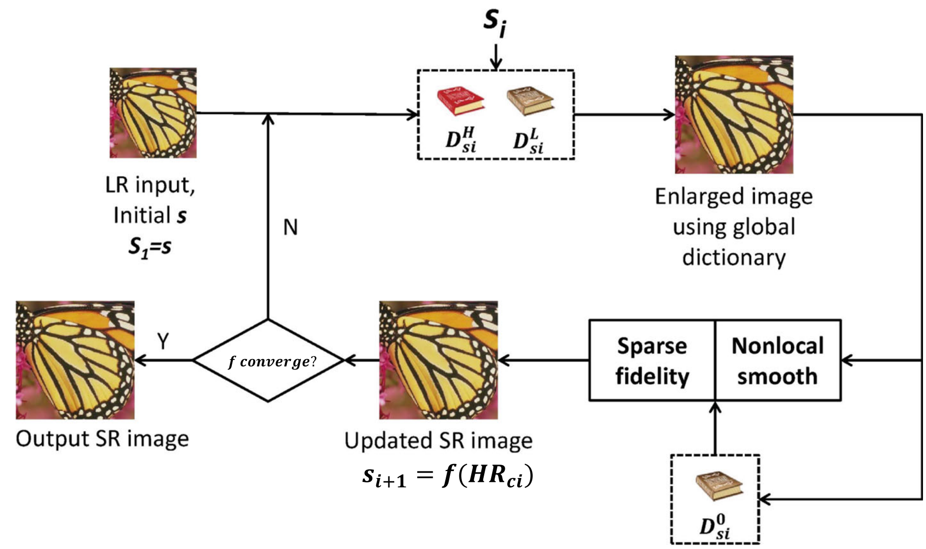

Fig. 1 Flowchart of the proposed SR method. s is the size ratio between the output HR and the input LR; s_i is the image magnification factor in the ith iteration, and it determines the choice of the global dictionaries D_{si}^H and D_{si}^L; D_{si}^0 is the local dictionary trained from the image itself

Afterwards, the magnifying factor s_i is updated according to a blind image quality estimation function $f(\mathrm{HR}_c)$, where HR_c is the current estimated HR image. The HR image is iteratively updated until function $f(\mathrm{HR}_c)$ converge.

The detailed descriptions of our proposed SR method will be introduced as follows.

3.2 Global dictionary training based on multi-scale image structures

According to Section 2.2, the initial value significantly affects the quality of the final HR image in local dictionary-based SR method such as NCSR. Compared with using the NCSR's default initial value, which is generated by bicubic interpolation, the quality of the estimated HR image can be significantly improved if the initial values are better generated by other SR methods. Therefore, we propose to train global dictionaries from large dataset within multiple scaling factors to better generate the initial guess of the HR image.

In global dictionary-based sparse image representation, it is often assumed that the same image patch should have the same sparse code in different resolutions. Given an LR image I_{lr} and a dictionary trained D_{lr} from LR image dataset, the sparse codes of image patches in I_{lr} can be estimated. According to the assumption that the corresponding HR image I_{hr} shares the same sparse code with I_{lr}, we can reconstruct the high resolution image from the low-resolution one if an appropriate high resolution dictionary D_{hr} is also available. It is obvious that the most important step is to find out the dictionary pair D_{hr} and D_{lr} that can reliably represent the HR image and its LR version with the same sparse code.

Given a large high-resolution training dataset $S_{hr} = \{I_{hr1}, I_{hr2}, \ldots, I_{hrn}\}$ with clear natural images I_{hri}, the low-resolution training dataset $S_{lr} = \{I_{lr1}, I_{lr2}, \ldots, I_{lrn}\}$ is generated by applying Gaussian blurring, down-sampling, and bicubic scaling to the same size as the images I_{hri} in S_{hr}. Afterwards, the images in S_{lr} and S_{hr} are decomposed into patch sets $P_{lr} = \{p_{lr1}, p_{lr2}, \ldots, p_{lrm}\}$ and $P_{hr} = \{p_{hr1}, p_{hr2}, \ldots, p_{hrm}\}$, where m is the number of patches extracted from the dataset.

In order to guarantee the dictionary a good representation of viewing-sensitive textures, we represent the image by its high-frequency component other than the original image. Similar with [11], the features in HR patch (p_{hri}) are extracted by subtracting the corresponding LR from the original HR, while the features in LR patch (p_{lri}) are extracted by using first- and second-order gradient filters. Afterwards, two training matrix can be generated: HR training matrix ($X_{hr} = [x_{hr1}, x_{hr2}, \ldots, x_{hrm}]$) and LR training matrix ($X_{lr} = [x_{lr1}, x_{lr2}, \ldots, x_{lrm}]$), where each x is a column vector reshaped from one training patch. Given X_{hr} and X_{lr}, the high- and low-resolution dictionaries can be estimated by Eqs. (4) and (5), respectively:

$$D_h = \underset{D_h,\alpha}{\arg\min} \left\{ \|X_{hr} - D_h\alpha\|_2^2 + \lambda\|\alpha\|_1 \right\}, \tag{4}$$

$$D_l = \underset{D_l,\alpha}{\arg\min} \left\{ \|X_{lr} - D_l\alpha\|_2^2 + \lambda\|\alpha\|_1 \right\}. \tag{5}$$

Because the sparse code α is shared between LR and HR patches, Eqs. (4) and (5) can be combined in Eq. (6):

$$[D_l, D_h] = \underset{D_l,D_h,\alpha}{\arg\min} \left\{ \|X_{lr} - D_l\alpha\|_2^2 + \|X_{hr} - D_h\alpha\|_2^2 + \lambda\|\alpha\|_1 \right\}. \tag{6}$$

The global dictionary pair is trained to reveal the underlying relation between the LR and HR images based on the knowledge from the training dataset. If the LR training images are generated by down-sampling the HR training images with a fixed scaling factor s, the LR images can only provide the structure knowledge in a fixed level. Therefore, one global dictionary pair trained from such datasets may not be suitable in different situations. For example, we can generate the LR training images by down-sampling the HR training images with a scale factor 4 and then estimate the global dictionary pair to reveal the LR-HR mapping relation. Although this dictionary pair may be effective if we scale an LR image to 4 times of its original size, it may performs unsatisfactorily if we scale the LR image to some other sizes. With this concern, we down-sample the HR training images by different scaling factors to generate multiple LR training set $\{S_{lr1}, S_{lr2}, \ldots, S_{lrm}\}$. The LR images in different training sets S_{lri} represent the low-resolution structures in multi-scale. As shown in Fig. 2, we generate the LR-HR dictionary pair for each LR-HR training set pair $\{S_{lri}, S_{hr}\}$. Given the global dictionary pairs under different structure level, we can always choose the appropriate dictionary to enhance the LR image according to different situations.

In order to illustrate the necessity of multiple global dictionary pairs under different structure level, we generate 30 LR training datasets $\{S_{lr1}, S_{lr2}, \ldots, S_{lr30}\}$ from one HR training dataset S_{hr}. Every LR set is generated by down-sampling the HR images at the ratio of s_i. In this paper, $s_i \in \{1.1, 1.2, 1.3, \ldots, 4.0\}$. In Fig. 3, we reconstruct the HR image from the LR image by using different global dictionary pairs. The horizontal axis represent the down-sampling ratio s_i at which the LR training set is generated. The vertical axis represent the PSNR value of the reconstructed HR image. In Fig. 3a, the LR image is scaled by 2.4 to generate the HR image, and the best reconstruction result is obtained when the LR training set is down-sampled by 2.5. Similarly, the LR image is scaled by 3.3 to generate the HR image in Fig. 3b, and the best reconstruction result is obtained when the LR training set is down-sampled by 3.5. It is clear that the choice of global dictionary pair affects the HR reconstruction result. In Fig. 4, the visual quality of the reconstructed

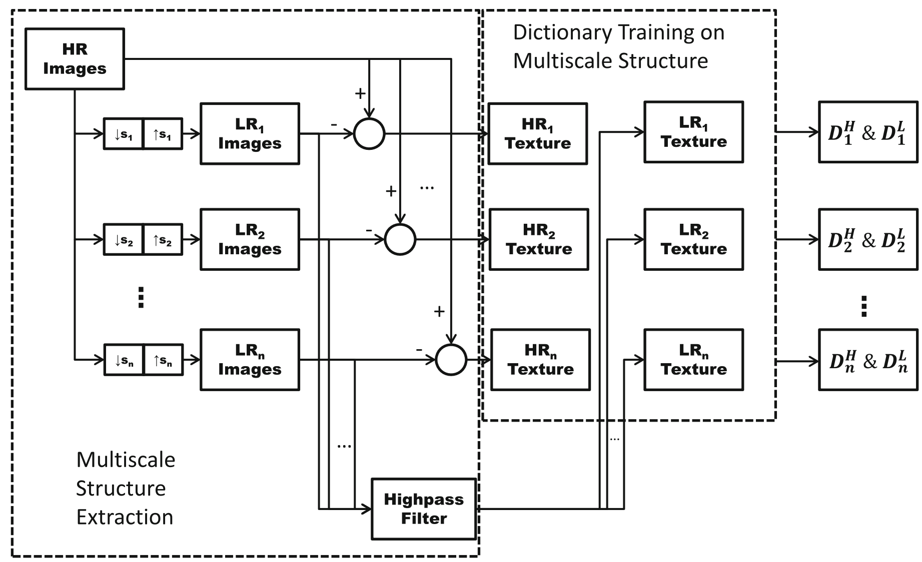

Fig. 2 Global dictionary training

HR images are presented along with the ground truth HR. It is visually noticeable that the quality of reconstructed HR would be better if appropriated global dictionary pair is used.

3.3 Local dictionary training using K-PCA

There is a need for great diversity in global dictionary, so that it can be used to recover general images. Despite the comprehensive information provided by global dictionary, it is proved to be unstable for sparse representation because of the highly diversity [15]. In order to represent the image by using a robust and compact dictionary, we use K-PCA and non-locally centralized sparse representation (NSCR) [17] to generate the local dictionary that is consistent with the input image.

The input LR image is scaled to a set of images $S_I = \{I_{s_k} | k = 1, 2, 3, \ldots, N\}$ with different sizes by using bicubic interpolation. If the input $LR \in \mathbb{R}^{m \times n}$, the desired output $HR \in \mathbb{R}^{sm \times sn}$, the height and width of the scaled image I_{s_i} is $0.8^{s_k} sm$ and $0.8^{s_k} sn$. By extracting 7×7 image patches from S_I, we generate the patch set $\boldsymbol{P}$, which is further clustered into K groups $\boldsymbol{P} = \{\boldsymbol{P_i} | i = 1, 2, \ldots, K\}$ using K-means clustering. We assume the patches within one group are similar, so these patches can be robustly represented

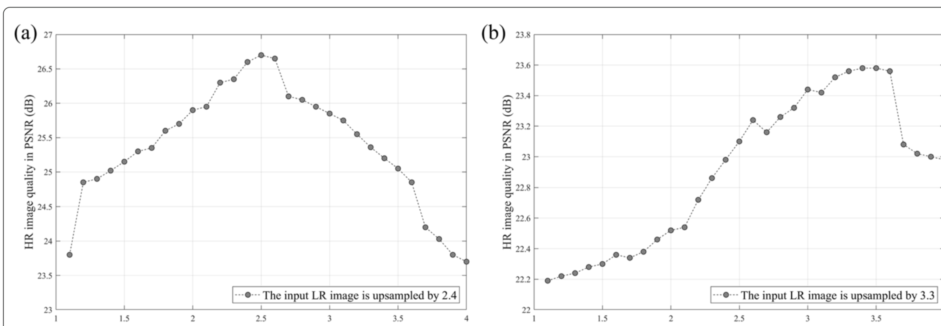

Fig. 3 Quality of HR image reconstructed by using different global dictionary pairs

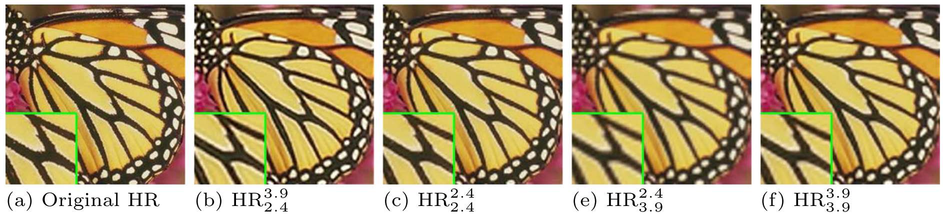

Fig. 4 Enhance LR image to HR with different global dictionaries. HR_a^b represents the estimated HR image with different scale factors and global dictionaries, where **a** is the scale factor of the test image and **b** is the scale factor used in the training phase

by using a compact dictionary D_i. Principal component analysis (PCA) is applied, and the PCA bases is regarded as D_i for group $\boldsymbol{P_i}$. After we combine all D_i $(i = 1, 2, \ldots, K)$ together, a complete local dictionary $D^0 = [D_1, D_2, \ldots, D_K]$ can be generated based on the input LR image itself.

3.4 Image super-resolution based on local and global training

In this section, we introduce the high resolution image reconstruction based on the global and local dictionaries. As shown in Eq. (3), a standard solution for image sparse representation can be formulated by the minimum optimization of an energy function with the fidelity term and the regularization term. The fidelity term ensures that the observed low-resolution image is a blurred and down-sampled version of the high-resolution image that is constructed by sparse representation. In this case, a reliable sparse representation is critical for high-resolution image reconstruction. In this paper, we adopt the global and the local dictionaries at the same time to ensure that the sparse representation can provide rich texture details and can be consistent with the observed low resolution image. With these concerns, we reformulate Eqs. (3) to (7).

$$\boldsymbol{\alpha_y} = \arg\min_{\alpha_l} \left\{ \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{D^0}\boldsymbol{\alpha_l}\|_2^2 + \lambda \|\boldsymbol{\alpha_l} - \boldsymbol{\beta_l}\|_1 \right\}, \tag{7}$$

s.t.

$$\|U_{IP}(\boldsymbol{y}) - \boldsymbol{D^L}\boldsymbol{\alpha_g}\|_2^2 < \epsilon, \tag{8}$$

where ϵ is a small factor, $\boldsymbol{y}$ is the observed low-resolution image, $\boldsymbol{H}$ is a matrix for blurring and down-sampling, $\boldsymbol{D^0}$ is the local dictionary, $\boldsymbol{\alpha_l}$ is the sparse code of the high-resolution image according to local dictionary, $\boldsymbol{D^L}$ is the global LR dictionaries, $\boldsymbol{\alpha_g}$ is the sparse code of the image according to global dictionary, $U(\cdot)$ is the upscaling operator, $U_{\mathrm{IP}}(\boldsymbol{y})$ is the initial prediction of the upscaled $\boldsymbol{y}$ in the gradient decent based optimization for Eq. (7), and $\boldsymbol{\beta_l}$ is the non-local mean of $\boldsymbol{\alpha_l}$, which is formulated as follows:

$$\boldsymbol{\beta_i} = \sum_{n \in N_i} \omega_{i,n} \boldsymbol{\alpha_n}, \tag{9}$$

where $\boldsymbol{\alpha_n}$ denotes the sparse code of an image patch x_n, N_i denotes the N most similar patches to patch x_i, $\boldsymbol{\beta_i}$ is the sparse code of patch x_i after nonlocal smoothing, and $\omega_{i,n}$ is the weight factor defined as follows:

$$\omega_{i,n} = \frac{\exp\left(-\|x_i - x_n\|_2^2\right)}{\sum\limits_{n \in N_i} \exp\left(-\|x_i - x_n\|_2^2\right)}. \tag{10}$$

Given an LR image, its HR version is estimated by iteratively solving Eq. (7). In the ith iteration, the global sparse code of the current LR image α_g^i is estimated by using Eq. (11):

$$\alpha_{gi} = \arg\min_{\alpha} \left\{ \|X_i^L - D_i^L \alpha\|_2^2 + \lambda \|\alpha\|_1 \right\}, \tag{11}$$

where X_i^L is the initial LR image in the ith iteration and also the final HR image estimation in the $(i-1)$th iteration, note that $U(\boldsymbol{y})$ in Eq. (8) is the general representation for X_i^L; D_i^L is the LR global dictionary used in the ith iteration; α is the sparse code of X_i^L; and α_{gi} is the optimal α that can minimize Eq. (11).

With the global sparse code α_{gi} from Eq. (11), the HR estimation from global dictionary in the ith iteration is given in Eq. (12):

$$X_{i+1/3}^L = D_i^H \alpha_{gi}, \tag{12}$$

where D_i^H is the HR global dictionary used in the ith iteration, and $X_{i+1/3}^L$ is an intermediate HR estimation, and it is also the initial HR guess feeding to the following local dictionary based HR estimation.

Given $X_{i+1/3}^L$, the local dictionary $\boldsymbol{D^0}$ and the corresponding local sparse code $\boldsymbol{\alpha_l}$ in Eq. (7) can be generated by K-PCA as mentioned in Section 3.3. According to Eqs. (9) and (10), the sparse code $\boldsymbol{\beta_i}$ of each patch's non-local mean can be estimated. With $\boldsymbol{\beta_i}$, the regularization term $X_i^{\beta_i}$ can be calculated as follows:

$$X_i^{\beta_i} = D_i^0 \beta_i. \tag{13}$$

After we have H, $\boldsymbol{D^0}$, the initial sparse code $\boldsymbol{\alpha_l}$, and the sparse code $\boldsymbol{\beta_l}$ of non-local mean, the optimal $\boldsymbol{\alpha_y}$ in Eq. (7) could be iteratively approached. First, we fix the second regularization term, and the optimization problem becomes a least square problem which can be efficiently solved, the gradient decent-based updating processing is given in matrix form by Eq. (14):

$$X_{i+2/3}^L = \boldsymbol{D^0}\boldsymbol{\alpha_l} + \theta H^T \left(y - H\boldsymbol{D^0}\boldsymbol{\alpha_l}\right), \tag{14}$$

where θ is the learning step size, which is set to 2.4 in this paper. Afterwards, we fix the first fidelity term in Eq. (7), and now, this function can be solved by iterative shrinkage algorithm:

$$X_{i+1}^L = S_\tau \left(X_{i+2/3}^L - X_i^{\beta_i}\right) + X_i^{\beta_i} \tag{15}$$

where S_τ is a soft-thresholding operator; $X_i^{\beta_i}$ is calculated in Eq. (13). According to the work presented in [30], the aforementioned algorithm is empirically converge. We also present the PSNR convergence of our proposed method in Fig. 5. There are three cases being compared: (1) the global SR $D^H\alpha_g$ is not used, and the SR image is generated based on local dictionary only; (2) the global SR is used only for once as the initial estimation for local SR; and (3) the global SR is used for two times to update the initial estimation during the gradient decent optimization of local SR. It can be observed that the use of global SR significantly improves the SR quality. It is worth notice that the global SR at the 201th iteration firstly reduces the PSNR but the final PSNR converges to a higher value compared with the cases without using global SR. It is very likely that the global SR updates the estimation during the gradient decent (GD) processing in local SR, and it avoids the

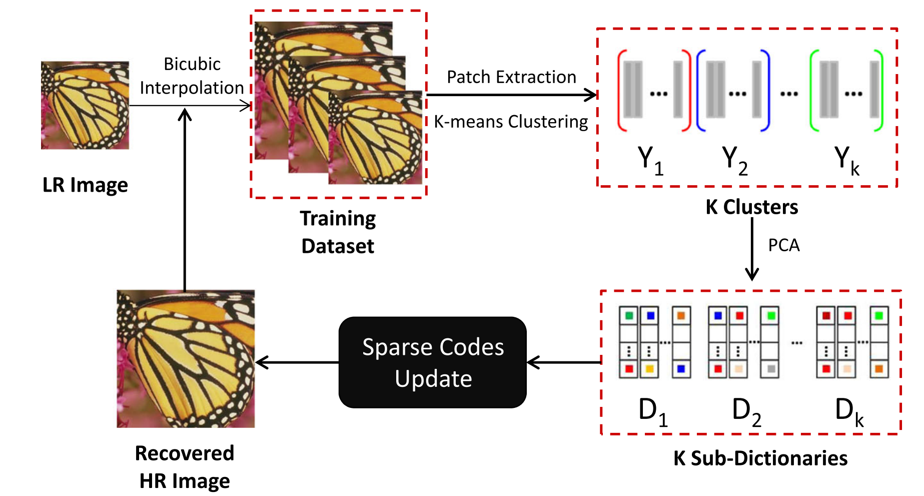

Fig. 5 The convergence comparison of our proposed method

GD processing being trapped in the local minima. According to our experiments, the proposed SR method can converge within 300 iterations for calculating Eq. (14).

The last problem is to find out the proper global dictionary in each iteration. Although we already generated global dictionary pairs in multi-scale structure, it is still difficult to select the most appropriate one in each iteration. According to our extensive experiments, we use the iteration number to stand for the degradation level and introduce a non-linear function in Eq. (16) to imply the selection of global dictionary pairs:

$$f(i) = \frac{a}{i} + b \tag{16}$$

where a and b are numeric parameters, i is the iteration number, and $f(i)$ is the index number for global dictionary selection. In this paper, we set a and b to be $s - 1$ and 1 for all test images, and here ,s is the scaling factor of the HR image.

4 Results and discussion

We evaluate and compare our proposed method with the present of Gaussian blurring during the down-sample procedures. All the methods are compared on Set5 [31], Set9 [17], and Set14 [11]. The hyperparameters in our method are set as follows: the patch size is 6×6 and the dictionary size is 1024 in global training, the patch size is 7×7, the K-PCA cluster size is 70, the learning rate in gradient decent is 2.4, and the sparsity lambda is 0.35 in local training.

Given the test HR image, it is first blurred by Gaussian kernel (windows size=7, standard deviation=1.6), and afterwards, it is directly down-sampled to generated the test LR image. It is worth to notice that the LR image generation in this paper is different from many existing SR methods that applied bicubic interpolation. In practice, it is not convincing to make the assumption that the relation between LR and HR images obeys the bicubic interpolation models. Therefore, introducing the blurring effect during the down-sampling makes the SR method more robust in real practice. In Table 2, the PSNR/SSIM comparison is made in the Y channel of YC_bC_r color space. Specifically, we compare our method with 5 other SR methods, which include the bicubic interpolation, the sparse representation based on global dictionary (SC) [11], A+ [13], anti-blur

Table 2 The PSNR(dB)/SSIM values of reconstructed SR images in 3 datasets

Upscale ratio	Dataset	Bicubic (PSNR/SSIM)	SC [11] (PSNR/SSIM)	A+ [13] (PSNR/SSIM)	$SRCNN_b$ [19] (PSNR/SSIM)	$SRCNN_g$ [19] (PSNR/SSIM)	NCSR [17] (PSNR/SSIM)	Proposed (PSNR/SSIM)
x3	Set5	26.48/0.79	26.72/0.80	26.72/0.81	26.73/0.81	29.67/0.85	33.19/0.91	**33.40/0.92**
	Set9	25.20/0.70	25.49/0.72	25.51/0.73	25.51/0.73	26.92/0.76	29.86/0.85	**30.06/0.85**
	Set14	24.72/0.68	24.92/0.70	24.91/0.70	24.93/0.70	27.20/0.76	29.38/0.82	**29.51/0.83**
x4	Set5	24.75/0.73	24.67/0.74	24.52/0.74	24.43/0.74	28.54/0.81	30.65/0.87	**31.27/0.88**
	Set9	23.91/0.65	23.90/0.66	23.78/0.66	23.72/0.66	26.13/0.72	27.71/0.79	**28.24/0.80**
	Set14	23.39/0.62	23.31/0.63	23.18/0.63	23.06/0.63	26.22/0.72	27.50/0.76	**27.91/0.76**

The PSNR/SSIM of the reconstructed images from our proposed method (the numbers in boldface) outperforms other SOTA conventional SR methods and the DL based SR methods

SR based on local dictionary (NCSR) [17], and the deep learning-based SR (SRCNN) [19], since the performance of deep learning-based methods highly depends on the training dataset. Therefore, we provide two deep learning models: SRCNN_b and SRCNN_g. SRCNN_b is the original model in [19]; the training HR-LR pairs are generated by bicubic interpolation. SRCNN_g is the model trained by ourselves; the training HR-LR pairs are generated by Gaussian blur (windows size=7, standard deviation=1.6) and direct down-sample.

Because the LR-HR mapping relation in the test phase is not consistent with the assumption made (i.e., bicubic interpolation) in the training phase of SC, A+, and SRCNN_g, it is not surprising that our proposed method outperforms SC, A+, and SRCNN_g in Table 2. The anti-blur SR based on local dictionary (NCSR) performs well with the present of Gaussian blur. As we mentioned in Section 2.2, the initial SR estimation in NCSR benefits the final result. Therefore, our proposed method outperforms NCSR by iteratively feeding the SR prediction from global dictionary to the local estimation. Although the training dataset for SRCNN_g includes Gaussian blur effect, our proposed method can still generate better SR images with respect to PSNR/SSIM.

In Fig. 6, we provide the visual comparison in the RGB color space. Due to the combination of global and local dictionary, our proposed method provide clearer details. Although it is potentially possible to produce artifacts and even errors when using global dictionaries, the self-similarity constraints and non-local centralized sparsity will efficiently suppress this problem. Compared with the other methods, our proposed SR method can produce convincing novel details, sharper boundaries, and clearer textures while it also effectively suppresses undesired artifacts.

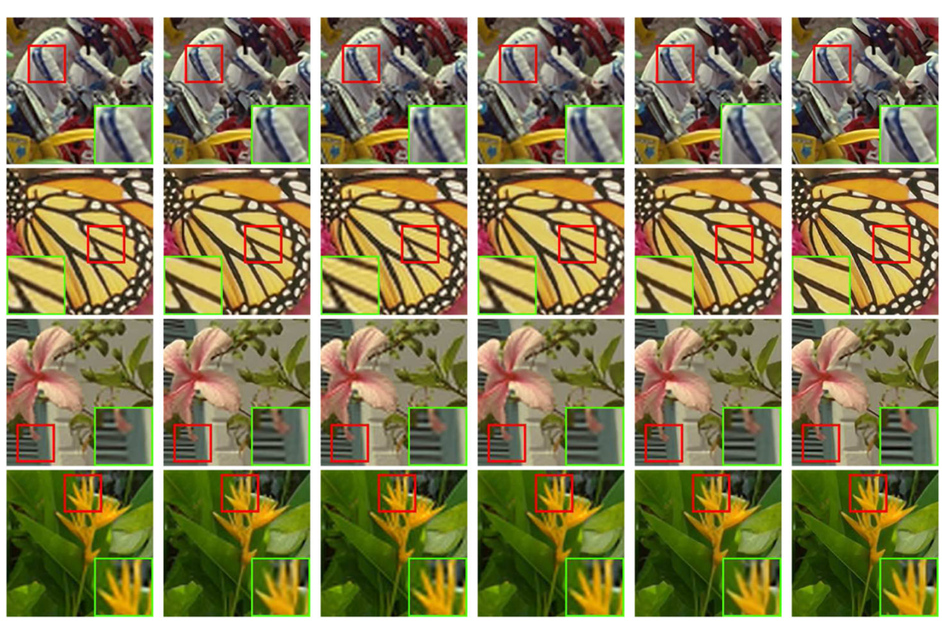

Fig. 6 Visual comparison between different SR methods (scale ratio=4). Each column presents the HR estimations from one SR method. From the left to the right, the compared methods are SC [11], A+ [13], SRCNN_b [19], SRCNN_g [19], NCSR [17], and our proposed method

5 Conclusion

In this paper, we propose a novel and effective SR method that utilize multi-scale image structures and non-local similarities. Specifically, a set of global dictionary pairs are trained under different image resolutions, so that the LR-HR mapping relations can be comprehensively established. When an LR image is enhanced to an HR image, the appropriate global dictionary pair can be chosen from the dictionary set by using a non-linear function. In addition, a K-PCA-based local dictionary is also trained according to the input LR image content. This local dictionary is more consistent with the input, and it helps to reduce the artifacts introduced by the feature inconsistency between the test image and the global training dataset. Furthermore, the sparsity-based non-local mean, which is proved to be effective in many SR methods, is used to smooth the estimated HR image in every iteration. In this case, less artifact will be propagated to the next iteration. The experimental results show that our proposed SR model is capable to recover HR images with clear textures, sharp edges, and convincing novel details.

Abbreviations

LR: Low-resolution; HR: High-resolution; SISR: Single image super-resolution; CNN: Convolutional neural network; K-PCA: *K*-means principal component analysis

Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Authors' contributions

All authors are involved in deriving the algorithm and making the validation experiments. All authors read and approved the final manuscript.

Funding

Work presented in this paper has been partially supported by the Provincial Key Research and Development Program of Sichuan, China (2020YFG0149).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Electronics Engineering, University of Electronic Science and Technology of China, No. 2006, Xiyuan Ave., West Hi-Tech Zone, 611731 Chengdu, People's Republic of China. [2]School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Ave., K1N 6N5 Ottawa, Canada. [3]Department of Electrical and Computer Engineering, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada.

Received: 8 January 2018 Accepted: 17 March 2021
Published online: 17 May 2021

References

1. R. Y. Tsai, T. S. Huang, *Advances in computer vision and image processing*. (JAI Press, Greenwich, CT, USA, 1984)
2. T. Blu, P. Thevenaz, M. Unser, Linear interpolation revitalized. IEEE Trans. Image Process. **13**(6), 710–719 (2004)
3. X. Li, M. Orchard, New edge-directed interpolation. IEEE Trans. Image Process. **10**(10), 1521–1527 (2001)
4. L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans. Image Process. **15**(8), 2226–2238 (2006)
5. J. Sun, Z. Xu, H. Shum, Gradient profile prior and its applications in image super-resolution and enhancement. IEEE Trans. Image Process. **20**(6), 1529–1542 (2011)
6. K. Zhang, X. Gao, D. Tao, X. Li, Single image super-resolution with nonlocal means and steering kernel regression. IEEE Trans. Image Process. **21**(11), 4544–4556 (2012)
7. W. Freeman, T. Jones, E. Pasztor, Example-based super-resolution. IEEE Comput. Graph. Appl. Mag. **22**(2), 56–65 (2002)
8. K. Ni, T. Nguyen, Image super-resolution using support vector regression. IEEE Trans. Image Process. **16**(6), 1596–1610 (2007)
9. J. Yang, J. Wright, T. S. Huang, Y. Ma, Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)

10. J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution. IEEE Trans. Image Process. **21**(8), 3467–3478 (2012)
11. R. Zeyde, M. Elad, M. Protter, in *Proceedings of 8th International Conference: Curves and Surfaces*, On single image scale-up using sparse-representations, pp. 711–730 (2010)
12. R. Timofte, V. D. Smet, L. V. Gool, in *Proceedings of IEEE International Conference on Computer Vision*, Anchored neighborhood regression for fast example-based super-resolution, pp. 1920–1927 (2013)
13. R. Timofte, V. D. Smet, L. V. Gool, in *Proceedings of Asian Conference on Computer Vision*, Adjusted anchored neighborhood regression for fast super-resolution, pp. 111–126 (2014)
14. W. Shi, S. Liu, F. Jiang, D. Zhao, Z. Tian, Anchored neighborhood deep network for single-image super-resolution. EURASIP J. Image Video Process. **2018**(34), 1–12 (2018)
15. M. Elad, I. Yavneh, A plurality of sparse representations is better than the sparsest one alone. IEEE Trans. Inf. Theory. **55**(10), 4701–4714 (2009)
16. W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans. Image Process. **20**(7), 1838–1857 (2011)
17. W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration. IEEE Trans. Image Process. **22**(4), 1620–1630 (2013)
18. K. Zhang, D. Tao, X. Gao, X. Li, Z. Xiong, Learning multiple linear mappings for efficient single image super-resolution. IEEE Trans. Image Process. **24**(3), 846–861 (2015)
19. C. Dong, C. C. Loy, K. He, X. Tang, in *Proceedings of European Conference on Computer Vision*, Learning a deep convolutional network for image super-resolution, pp. 184–199 (2014)
20. J. Kim, J. K. Lee, K. M. Lee, in *IEEE Conference on Computer Vision and Pattern Recognition*, Accurate image superresolution using very deep convolutional networks, (2016)
21. J. Hu, J. Zhao, in *Proceedings of IEEE International Conference on Instrumentation and Measurement Technology*, A joint dictionary based method for single image super-resolution, pp. 1440–1444 (2016)
22. A. Marquina, S. J. Osher, Image super-resolution by tv-regularization and bregman iteration. J. Sci. Comput. **37**(3), 367–382 (2008)
23. H. A. Aly, E. Dubois, Image up-sampling using total-variation regularization with a new observation model. IEEE Trans. Image Process. **14**(10), 1647–1659 (2005)
24. G. Peyre, S. Bougleux, L. Cohen, Non-local regularization of inverse problems. Inverse Probl. Imaging. **5**(2), 511–530 (2011)
25. I. Daubechies, M. Defrise, C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pur. Appl. Math. **57**(11), 1413–1457 (2004)
26. J. A. Troopp, S. J. Wright, Computational methods for sparse solution of linear inverse problems. Proc. IEEE. **98**(6), 948–958 (2010)
27. K. I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior. IEEE Trans. Pattern Anal. Mach. Intell. **32**(6), 1127–1133 (2010)
28. Y. Zhang, W. Yang, Z. Guo, Image super-resolution based on structure-modulated sparse representation. IEEE Trans. Image Process. **24**(9), 2797–2810 (2015)
29. Y. Zhang, J. Liu, W. Yang, Z. Guo, Image super-resolution based on structure-modulated sparse representation. IEEE Trans. Image Process. **24**(9), 2797–2810 (2015)
30. E. Candés, Enhancing sparsity by reweighted l_1 minimization. J. Fourier Anal. Appl. **14**(5-6), 877–905 (2008)
31. M. Bevilacqua, A. Roumy, C. Guillemot, M.-L. Alberi-Morel, in *Proceedings of BMVC*, Low complexity single-image super-resolution based on nonnegative neighbor embedding, pp. 135–113510 (2012)

Publisher's Note