Signal Processing: Image Communication 87 (2020) 115925

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Image super-resolution reconstruction based on sparse representation and deep learning

Jing Zhang *, Minhao Shao, Lulu Yu, Yunsong Li

State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Keywords:
Sparse representation
Deep learning
Super-resolution
Feature fusion

ABSTRACT

Super-resolution reconstruction technology has important scientific significance and application value in the field of image processing by performing image restoration processing on one or more low-resolution images to improve image spatial resolution. Based on the SCSR algorithm and VDSR network, in order to further improve the image reconstruction quality, an image super-resolution reconstruction algorithm combined with multi-residual network and multi-feature SCSR(MRMFSCSR) is proposed. Firstly, at the sparse reconstruction stage, according to the characteristics of image blocks, our algorithm extracts the contour features of non-flat blocks by NSCT transform, extracts the texture features of flat blocks by Gabor transform, then obtains the reconstructed high-resolution (HR) images by using sparse models. Secondly, according to improve the VDSR deep network and introduce the feature fusion idea, the multi-residual network structure (MR) is designed. The reconstructed HR image obtained by the sparse reconstruction stage is used as the input of the MR network structure to optimize the high-frequency detail residual information. Finally, we can obtain a higher quality super-resolution image compared with the SCSR algorithm and the VDSR algorithm.

1. Introduction

The vision is the primary path for human to obtain information, and the effect of vision information mostly dependents on the quality of the image. The most important index to measure the quality of the image is spatial resolution, because high-resolution(HR) images contain richer detail information. In the actual scene, due to the limitation of imaging equipment's resolution, shooting conditions and artificial factors, it is difficult for the imaging system to obtain the information without distortion, which makes it hard to obtain high quality images in practice. Improving image quality, especially improving the spatial resolution of images is a hot and urgent topic in image science research and engineering application. It is costly to improve the spatial resolution of the image by promoting the hardware device, so it is more feasible to research the software approach to increase the resolution of the image, that is, the image super-resolution technology to be investigated in this paper.

Image super-resolution technology is an image processing approach, and high-resolution clear images can be reconstructed from the image of single frame low-resolution (LR) images based on the imaging model and descending prior information, which not only have detail information but also obtain the resolution beyond the limitation of imaging system. So image super-resolution can improve the visual effect of the image, and provide some benefits for subsequent processes, such as feature extraction, information recognition, traffic and security monitoring, etc.

The concept and approach of image super-resolution reconstruction was proposed by Harris and Goodman in 1964, here the approach of extrapolation with limited signal is applied to the super-resolution reconstruction of optical image in a dual way, and good results are achieved. So far, a variety of image super-resolution reconstruction approaches have been proposed. the image super-resolution reconstruction is generally classified as three categories: interpolation-based approach, reconstruction-based approach and learning based approach. Among these, the learning-based approach has better performance than others and it is divided into shallow learning and deep learning-based approaches. The approach based on interpolation, reconstruction and shallow learning is generally referred to as the traditional image super-resolution reconstruction technology.

The simplest approach of SR is based on interpolation [1–3], which estimate the HR pixels using their neighborhoods. The computational cost of this approach is very low and the real-time performance is high, while simple interpolation methods such as Bilinear or bicubic interpolation tend to generate overly smooth images with ringing and jagged artifacts, so the image reconstruction quality is poor. The approach based on reconstruction is widely used to reconstruct high-resolution images from low-resolution images, and good results are obtained under certain scenes, such as [4–6]. But there are still many

* Corresponding author.
E-mail address: jingzhang@xidian.edu.cn (J. Zhang).

https://doi.org/10.1016/j.image.2020.115925
Received 15 May 2019; Received in revised form 9 February 2020; Accepted 16 June 2020
Available online 23 June 2020

problems to be solved, the computational cost is high, the convergence time is slow, the real-time efficiency is not high, and the quality of the reconstructed image decreases rapidly when the desired magnification factor is large. In these cases, the result may be overly smooth, lacking important high-frequency details [7].

Considering the disadvantage of interpolation approach and reconstruction approach, and in order to make better use of prior information to obtain more high-resolution image detail information, some approaches of super-resolution reconstruction based on learning are put forward. the basic idea is to select low-resolution images and their corresponding of the same type as training sets, and to obtain unknown information by machine learning approach rather than artificially defining. So the rich high frequency information can be obtained by the characteristics of different components during the training process to improve the reconstruction image quality. The typical approach based on learning is divided into two categories based on shallow learning approach and deep learning approach. Among the shallow learning approaches, sparse representation can achieve the better performance, Typical representative is Yang et al. [8] (SCSR), which proposes the sparse representation model to reconstruct the HR image from its LR counterpart by establishing the mapping relationship between the LR and HR patches. This approach can be divided into three stages of the feature extraction, the sparse dictionary training and the image reconstruction. it can improve the image detail information at the same time maintain the geometric structure information, resulting in a better reconstruction image. Later, many researchers make improvements of SCSR [9–11]. Dong et al. [12] also adopt the sparse representation model, clustering the patches into several classes and obtaining the different overcomplete dictionaries, then it adds a new constraint condition for obtaining better quality images.

Because of the learning ability limitation of the shallow learning model, the high frequency details of the reconstructed image are not rich enough. In order to make up for the disadvantages of the shallow learning model, some deep learning models [13–15] are proposed and have a wide application in SR.

In [16], Dong et al. propose a Super-Resolution Convolutional Neural Network (SRCNN) to learn a nonlinear LR-to-HR mapping function. This network architecture has been extended to embed a sparse coding model [17] and it also adds some recursive layers [18,19]. The VDSR network [20] shows significant improvement over the SRCNN method by increasing the network depth from 3 to 20 convolutional layers. To facilitate training a deeper model with a fast convergence speed, the VDSR method adopts the global residual learning paradigm to predict the differences between the ground truth HR image and the bicubic up sampled LR image instead of the actual pixel values. The DRRN approach [19] further trains a 52-layer network by extending the local residual learning approach of the ResNet [21] with deep recursion. We note that the above methods use bicubic interpolation to up sample input LR images before feeding into the deep networks, which increases the computational cost and requires a large amount of memory.

The emergence of those approaches makes up for the deficiency of traditional shallow learning approaches and enhances the high frequency detail information of the reconstructed image comparing with the sparse method. In addition, the deep learning approach can easily combine with other traditional super-resolution models, efficiently improving the performance of reconstruction.

The major contributions of the proposed method are summarized as follows:

(1) In order to improve the reconstructed image quality, according to the characteristics of high- and low-resolution image blocks, the MFSCSR algorithm is proposed. In this algorithm, LR image blocks are divided into flat blocks and non-flat blocks through their variance. The contour features of non-flat blocks are extracted by Nonsubsampled Contourlet (NSCT) transform and the texture features of flat blocks are extracted by Gabor transform, then the contour dictionary and texture dictionary are jointly trained by using contour features and texture features, finally the HR image blocks are reconstructed by sparse model and all these blocks synthesize the reconstructed HR image .

(2) To optimize high frequency detail information for super-resolution images. Based on the VDSR network, we design a lightweight residual network structure, which has only four layers of convolution and introduces feature fusion. Finally, the MFSCSR algorithm is combined with proposed Multi-residual network to form our algorithm (MRMFSCSR) in this paper.

This paper is organized as follows. In Section 2, we study the technical theory of SCSR algorithm and residual network structure model. In Section 3, we improve the feature extraction method of SCSR algorithm and propose MFSCSR. At the same time, we design MR network according to VDSR network, and finally combine these two models into MRMFSCSR algorithm to enrich the high-frequency detail information and improve the quality of reconstructed super-resolution image. The experimental results are shown in Section 4, and finally Section 5 provides a summary of the work.

2. Related work

2.1. SCSR

The SCSR algorithm applies the sparse representation theory to image super-resolution reconstruction. It mainly includes three stages: training sample extraction, joint dictionary training and image reconstruction. Firstly, the edge information of LR image block is extracted by the gradient operator, then it groups with the HR image blocks to form the sample training set. Secondly, The LR image patches and their corresponding HR image patches are used to train a joint dictionary pair $\{D_h, D_l\}$. Finally, each LR image block and D_l are used to solve the sparse coefficient α, D_h and α can reconstruct the corresponding HR image blocks.

(1) Training sample extraction stage: The literature suggests extracting different features for the low-resolution image patches in order to boost the prediction accuracy. The SCSR algorithm uses the first- and second-order derivatives as the features for the low-resolution patches due to their simplicity and effectiveness. The four 1-D filters used to extract the derivatives are:

$$f_1 = [-1, 0, 1],\ f_2 = f_1^T,\ f_3 = [1, 0, -2, 0, 1], f_4 = f_3^T \tag{1}$$

where the superscript T means transpose. Applying these four filters yields four feature vectors for each patch, which are concatenated into one vector as the final representation of the low-resolution patch.

(2) Joint dictionary training stage: Given the sampled training image patch pairs $P = \{X_h, Y_l\}$, where $X_h = \{x_1, x_2 \dots x_n\}$ are the set of sampled high resolution image patches and $Y_l = \{y_1, y_2 \dots y_n\}$ are the corresponding low resolution image patches (or features), our goal is to train the two dictionaries D_h and D_l for high resolution and low resolution image patches, so that the sparse representation of the high resolution patch is the same as the sparse representation of the corresponding low resolution patch.

In order to satisfy the isomorphism of the two dictionaries under the sparse representation, the sparse coding problems are solved by:

$$D_h, D_l = \underset{\{D_h, D_l, A\}}{\arg\min} \|Y_c - D_c A\|_2^2 + \lambda(\frac{1}{N} + \frac{1}{M})\|A\|_1 \tag{2}$$

$$Y_c = \left[\frac{1}{\sqrt{N}} Y \ \ \frac{1}{\sqrt{M}} X\right], D_c = \left[\frac{1}{\sqrt{N}} D_h \ \ \frac{1}{\sqrt{M}} D_l\right] \tag{3}$$

where N and M are the feature vector dimensions of the high-resolution and low-resolution image patches respectively. The trained dictionaries are the key priori information in the reconstruction phase.

(3) Image reconstruction stage: the feature is extracted from the input LR image block to obtain the low-resolution feature block y, and the low-resolution dictionary D_l is used to calculate the sparse coefficient a corresponding to y. The calculation process is as follows:

$$a = \arg\min_a \|x - D_l a\|_2^2 + \lambda\|a\|_1 \tag{4}$$

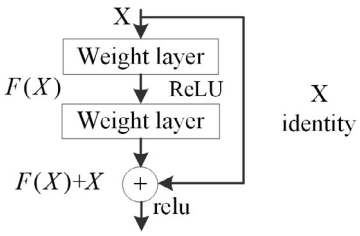

Fig. 1. Residual learning: a building block.

The sparse coefficient a is obtained by Eq. (4), and the high-resolution image is calculated by combining D_h as shown in Eq. (5).

$$x = D_h a \tag{5}$$

2.2. Deep learning

The typical residual network (ResNet) proposed by Microsoft Research He Kai Ming et al. in 2015 is a deep neural network, which successfully develops into a 152-layer ultra-deep network with remarkable results. The network is easily integrated with other networks and used for image processing.

The ResNet is the first proposed deep residual network and its primary building block is shown in Fig. 1. For practical application, the depth of this network can be designed arbitrarily, but with the increase of network depth, the computational complexity will naturally increase.

In the residual network structure, the input signal is transmitted directly to the output by means of a shortcut connection. The output results are:

$$H(x) = F(x) + x \tag{6}$$

Instead of having each few stacked layers directly fit a desired underlying mapping, we explicitly let these layers fit a residual mapping. It is assumed that the residual mapping is easier to be optimized than the original unreferenced mapping. To the extreme, if an identity mapping is optimal, it would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers. With adding a simple addition operation, the training speed of the network is getting faster, and the training effect is much better.

3. MRMFSCSR algorithm

3.1. Improved SCSR

Image super-resolution reconstruction is to reconstruct the SR image that is close to the original HR image and minimize the loss of edge and texture information. Therefore, extracting LR image feature information in the sparse representation reconstruction process is a key step, the more prior knowledge to joint training dictionary is extracted, the more accurate prediction of the missing HR information is able to be obtained. The traditional SCSR algorithm uses the first- and second-order derivatives to extract image feature information. In this way, only the feature information in the horizontal and vertical direction can be extracted, so the acquired texture and edge information are not rich enough resulting in poor reconstruction image quality. In order to solve the above problems, we propose a multi-feature extraction sparse representation super-resolution reconstruction algorithm (MFSCSR).

In the sample extraction stage of the MFSCSR's training model, according to variance, the LR image is segmented into LR image blocks, which contain flat blocks (small variance) rich in texture features and non-flat blocks (large Variance) with rich contour features, Fig. 2 shows the differences between flat blocks and non-flat blocks of Lena.

For different features of LR image blocks, MFSCSR uses NSCT to extract feature information of non-flat image blocks and uses Gabor transform to extract feature information of flat image blocks. Then these two different types of feature blocks and their corresponding HR image blocks constitute a training sample set.

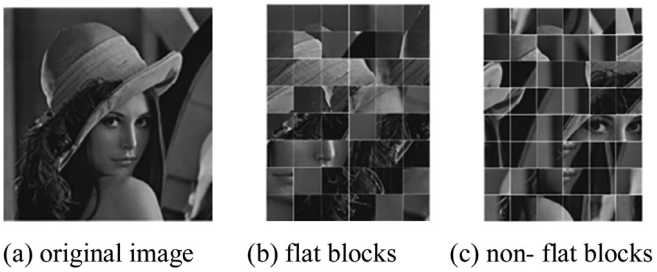

Fig. 2. The differences between flat blocks and non-flat blocks of Lena.

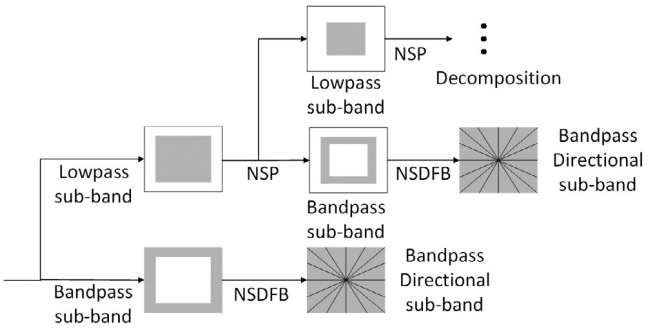

Fig. 3. The flow chart of the NSCT transform.

For flat image blocks with few details, the filter in the SCSR algorithm cannot obtain enough feature information. In order to obtain sufficient image feature information, it is effective to use the Gabor transform to process multiple scales of the image. The 2-D Gabor filter function used to extract the features is:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp(-\frac{(x'^2 + \gamma^2 y'^2)}{2\sigma^2} + i(2\pi \frac{x'}{\lambda} + \psi)) \tag{7}$$

$$x' = x\cos\theta + y\sin\theta \tag{8}$$

$$y' = -x\sin\theta + y\cos\theta \tag{9}$$

where λ defines the wavelength of the Gabor kernel, θ is the direction of the Gabor kernel function to be determined, ψ defines the Phase shift, σ defines the Gaussian standard deviation, and γ represents the spatial aspect ratio, which determines the shape of the Gabor function.

Compared with flat image blocks, non-flat image blocks contain more feature information, and feature extraction using filters in the SCSR algorithm misses much feature information. The NSCT transform can extract the feature information multiple times for the non-flat image block as needed, which greatly contributes to the improvement of the image reconstruction effect. The NSCT process consists of two parts, including multiscale decomposition and multidirectional decomposition. The nonsubsampled pyramid filter banks (NSPFB) structure is first used to achieve multi-scale decomposition, and then the nonsubsampled directional filter banks (NSDFB) structure is used to achieve multi-directional decomposition.

When an image is decomposed by NSCT, NSFFB is used for multi-scale decomposition, which is shown in Fig. 3. At each decomposition scale, a low pass sub-band image and a bandpass sub-band image are obtained, which are the same size as the original image. Then NSPFB is used to decompose the low pass sub-band images on each scale, and iterates to all the decomposition scales to obtain multi-scale and multi-directional decomposition of the image. Finally, NSDFB is used to decompose the bandpass sub-band images on each scale.

In this paper, flat blocks are trained to generate a texture dictionary pair $\{D_h^t, D_l^t\}$, which use to reconstruct a flat HR image block, and non-flat blocks are trained to obtain a contour dictionary pair $\{D_h^c, D_l^c\}$,

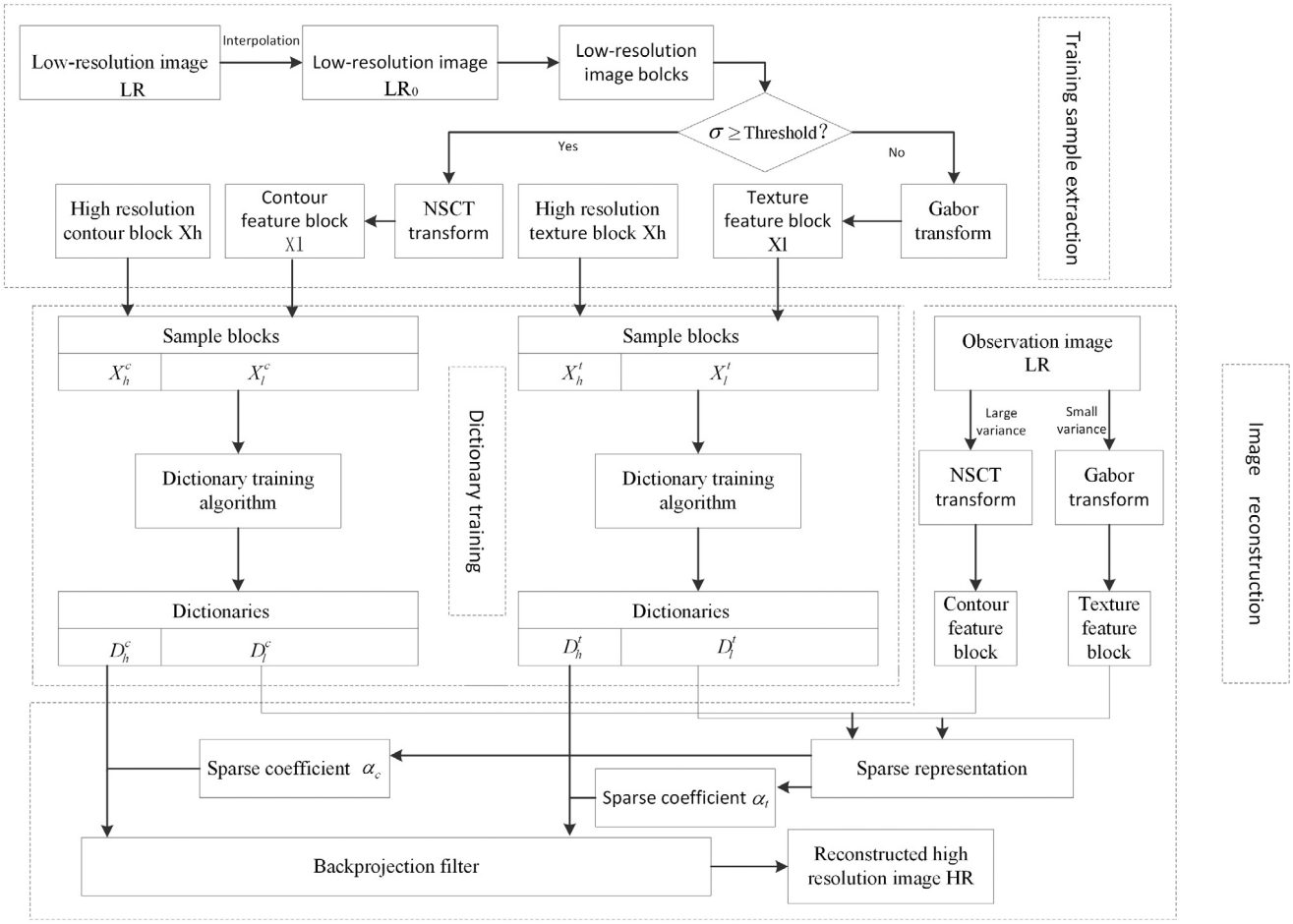

Fig. 4. The flow chart of MFSCSR algorithm.

which use to reconstruct a non-flat HR image block, finally two types of HR image blocks constitute a HR image. The flow chart of the MFSCSR algorithm is shown in Fig. 4.

The detailed reconstruction process of the MFSCSR algorithm is shown as follows:

Algorithm1 : **MFSCSR**

Task: Reconstruct high-resolution image.

- Input: low-resolution image x, contour dictionary $\{D_h^c, D_l^c\}$,texture dictionary $\{D_h^t, D_l^t\}$,variance threshold 0.27;
- Output: high-resolution image;
- Initialization: divided x into several $5 * 5$ blocks x_i (4 pixels overlapped between blocks);
- Workflow:

(1) For each block x_i, calculate the variance and classify it into flat block or non-flat block;

(2) If the variance is smaller than 0.27, the Gabor transform extracts the flat block texture feature x_i^t; If the variance is greater, the NSCT transform extracts the non-flat block contour feature x_i^c;

(3) If block x_i is a flat block, the texture sparse coefficient α_t is solved by using D_l^t and the flat texture feature x_i^t as Eq. (4);

(4) Calculate the reconstructed flat image block y_i^t by using D_h^t and α_t as Eq. (5).

(5) The non-flat image block reconstruction solution process is the same as step 3–4, and use the $\{D_h^t, D_l^t\}$ to calculate the reconstructed non-flat image block y_i^c.

(6) Integrate all the y_i^t and y_i^c to generate the reconstructed high-resolution image.

3.2. Multi-residual network structure

In this paper, a multi-residual network (MR) is designed based on Very Deep Convolutional Networks (VDSR), which only has one residual structure, and the multi-feature fusion idea is used to integrate the extracted features to complete the high-frequency detail optimization in our method. The multi-residual network framework is shown in Fig. 5.

The multi-residual network is used to optimize the high-frequency detail residual image between the interpolation LR image and the original HR image, the feature maps of the input interpolation LR image (ILR) is extracted by convolution, then we will fuse these obtained feature maps. Under the effective iteration number, training is carried out by optimizing the regression objective using mini-batch gradient descent based on backpropagation. Finally, the high-frequency detail residual is obtained, which can constitute with the ILR image to generate reconstructed high-resolution image. For the ILR image not only can be obtained by the interpolation method but also use the other models. The closer the ILR image is to the original HR image, the easier the high-frequency detail residual image is to be trained.

The detailed description of MR is shown as follows:

In the training stage, Using the ILR image and the original high-resolution image as the input image of the network. MR network extracts the feature information of the image sequentially through only the 4-layer convolution network and the ReLU activation function, then introduces multi-residual network structure, which not only transmit the input image directly to the output but also transmit the adjusted feature maps of each layer obtained by convolution by means of

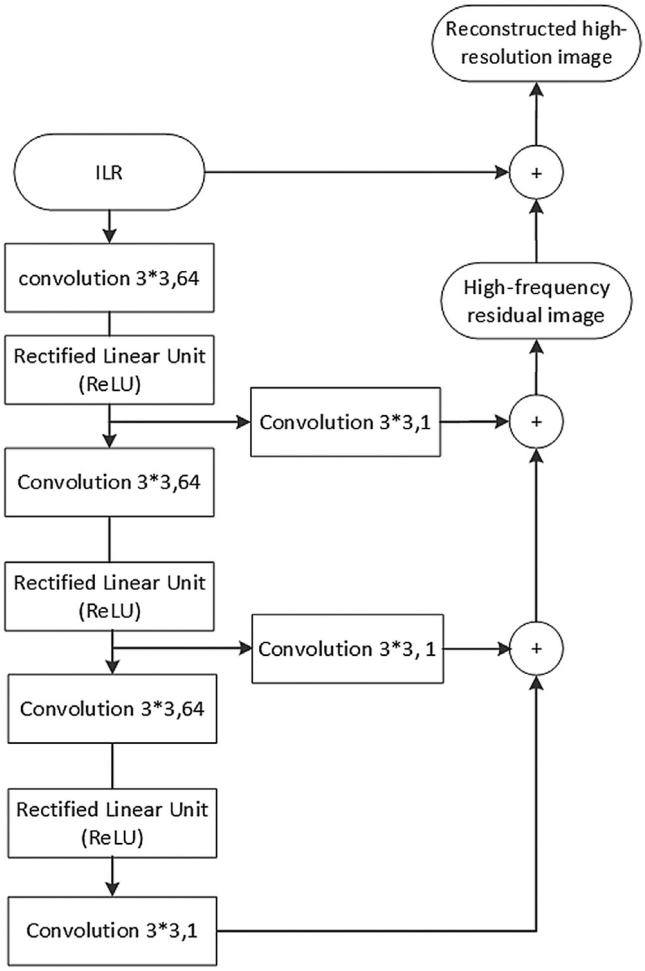

Fig. 5. The multi-residual network framework.

Multiple shortcut connection. The aim of adjusting feature map is to make channel consistent in the multi-residual structure. Due to the use of multiple residuals structure, a shallower network can still achieve the expected performance to reconstruct HR image.

3.3. MRMFSCSR

In order to make full use of the sparse representation and the neural network, this paper combines the MFSCSR algorithm with the MR network and proposes the MRMFSCSR algorithm, which is shown in Fig. 6. First, the reconstructed high-resolution image is obtained by the MFSCSR algorithm described in 3.1, which is imported to the MR network for reconstructing high-frequency detail image, and the reconstructed high-resolution image and high-frequency detail image are combined to obtain the final SR image.

Under the premise of balancing computational efficiency and image quality, this paper proposes an algorithm that combines sparse representation and residual network to deal with image super-resolution reconstruction, enriching high-frequency detail information while having less computational complexity comparing with VDSR. Therefore, MRMFSCSR has great value for the subsequent application of image processing such as remote sensing image target detection and feature extraction.

4. Performance testing and results analysis

In this section, we first demonstrate the super-resolution results on natural, remote sensing and infrared images. Then we continue to discuss various influential factors of the proposed algorithm including dictionary size and residual structure of network.

In the MRMFSCSR algorithm we choose a 2-scale, 4-direction Gabor filter bank and a 2-layer NSCT filter to extract features, the dictionary size is 512 and the variance threshold is 0.27, the patch size is $5 * 5$ pixels for both low- and high-resolution images. Simultaneously we use the network with a depth of 4, the batch size in training is 64, the convolution kernel size is $3 * 3$, stride is 1, padding is 1, and learning rate is 0.01. We evaluate the results both visually and qualitatively in Peak Signal to Noise Ratio (PSNR) and Structure Similarity Index (SSIM). The training of the network of the experiment is completed under the Caffe environment on the server with K80, and the image reconstruction test is completed on MATLAB 2010b under the Caffe environment.

4.1. Data set

In all experiments, training different methods using the same training set, which contains 91 images from the SCSR algorithm of Yang et al. and 200 additional images from the Berkeley segmentation dataset. All of the comparative experiments in this section use a unified testing set, including remote sensing images from satellites such as QuickBird, infrared images from NASA's official website, and natural scene images from the MIT image library, B100 and Urban100 image sets. In our experiments, we amplified the input low-resolution image by 2 and 4 to achieve super-resolution reconstruction.

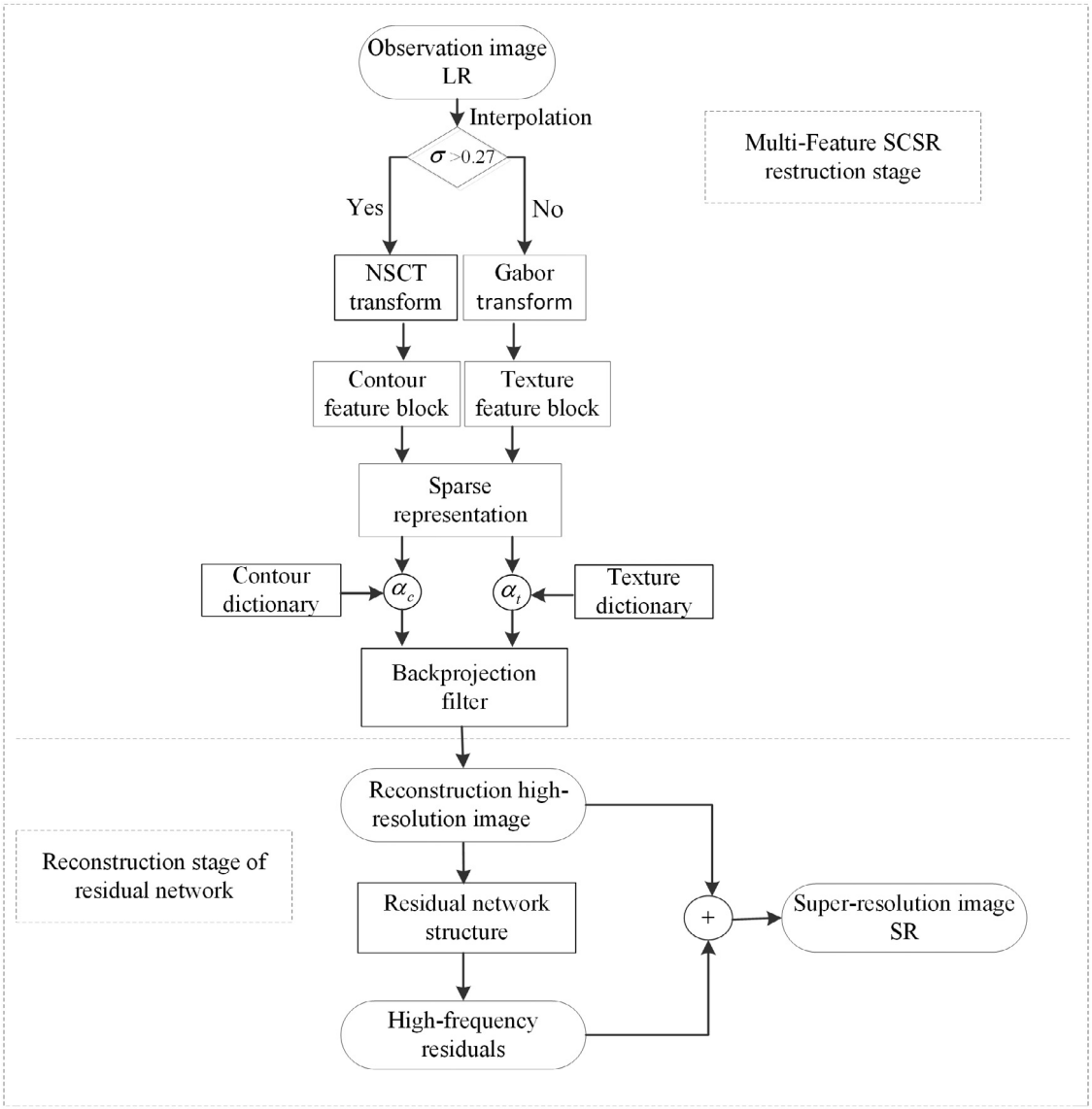

Fig. 6. The flow chart of the MRMFSCSR algorithm.

4.2. Experimental results

We apply all the comparative methods to generic images such as flowers, human faces and architectures; remote sensing images such as Mountain and ship; and Infrared images to complete image super-resolution reconstruction. Figs. 7–9 compare the outputs of our method with those of the Bicubic, SCSR, SRCNN and VDSR in visual. We also show the qualitatively measurement indicators (PSNR and SSIM) in Table 1. All experimental results are the average of each type of test images which totally contain more than 200 images.

Figs. 7 and 8 compares our method with the other four methods on two test images which contain a natural image of peppers and a remote sensing image of an island. In both cases, our method gives sharper edges and reconstructs more clearly the details of the scene. There are noticeable differences in the edge of the pedicle on the peppers and in the texture of the island platform.

According to the experimental results in Table 1, for the average of all the test images, compared with the traditional learning method, our method is 2 dB and 0.122 higher than the original SCSR in PSNR and SSIM at upscaling of 2, and 1.882 dB and 0.066 higher at upscaling of 4.

Compared with the deep learning method of VDSR, our method is 0.544 dB and 0.074 higher in PSNR and SSIM at upscaling of 2, and 0.5 dB and 0.0478 higher at upscaling of 4. Meanwhile, the MR network of our method is a 4-layer network with network parameters of about 110k, while the MFSCSR algorithm contains two pairs of high- and low-resolution dictionaries with about 130k parameters, and Table 2 shows the comparison of the network depth and the parameter amount of different methods, which shows our method has less parameters than VDSR. Therefore, compared with the current mainstream traditional learning method or deep learning method, our method has a better performance.

4.3. Results analysis

(a) Effects of dictionary size

In our method, the better output of MFSCSR, the better final reconstructed SR image. In the above experimental results, we fix the dictionary size to be 512 of MFSCSR. In this section, we evaluate the effect of dictionary size on generic image super-resolution. we train four dictionaries of size 256, 512, 1024 and 2048, and apply them to the same input image. The results are evaluated quantitatively in PSNR.

Fig. 7. Results of the peppers image magnified by a factor of 2. Left to right: Bicubic interpolation, SCSR, SRCNN, VDSR, our method and the original HR image.

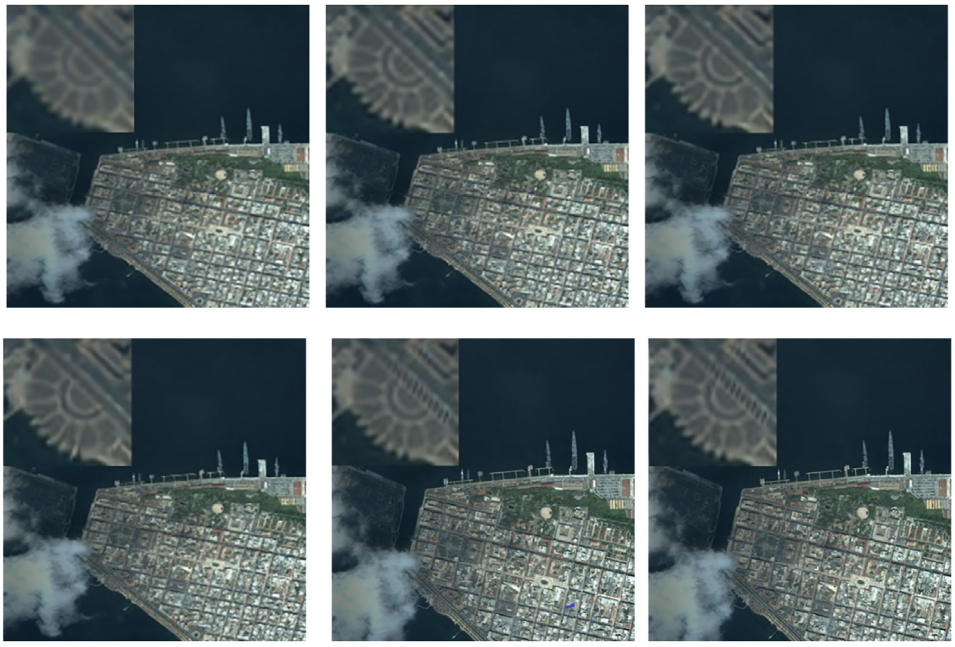

Fig. 8. Results of the remote sensing image magnified by a factor of 4. Left to right: Bicubic interpolation, SCSR, SRCNN, VDSR, our method and the original HR image.

Fig. 10 shows the reconstructed results for the building image using dictionaries of different sizes.

It can be seen from the PSNR of the experimental results in Table 3 that when the dictionary size belongs to [256, 512], the PSNR is increased by about 0.2 dB. When it belongs to the range of [512, 2048], the PSNR tends to be smooth, that means the reconstruction effect is not improved. Based on the above results, combined with the training time and PSNR indicator factors, the dictionary size is chosen to be 512 in our proposed method.

(b) Feature extraction methods

In this section, we evaluate the effect of different feature extraction methods on generic image super-resolution. we train dictionaries of SCSR, SCSR using Gabor transform, SCSR using NSCT transform and MFSCSR, and apply them to the same input image. In this experiment, the size of all trained dictionaries is 512, the training image block size is set to 5, and the multiple of image reconstruction is 2 times. The results are evaluated quantitatively in PSNR and SSIM, showing in Fig. 11 and Table 4.

The experimental results in Table 4 shows under the same conditions, the MFSCSR algorithm makes the reconstructed image closer to the real image, which can increase the PSNR of the reconstructed super-resolution image by about 0.83 dB compared to the SCSR algorithm. In the bird picture, the junction between the guanine and the sunlight in the picture reconstructed using the MFSCSR algorithm is the clearest. In the baby picture, it can be seen that the MFSCSR algorithm performs better for the detail reconstruction of the eyelashes and pupils, and this situation also occurs in the headgear of the protagonist in the woman figure.

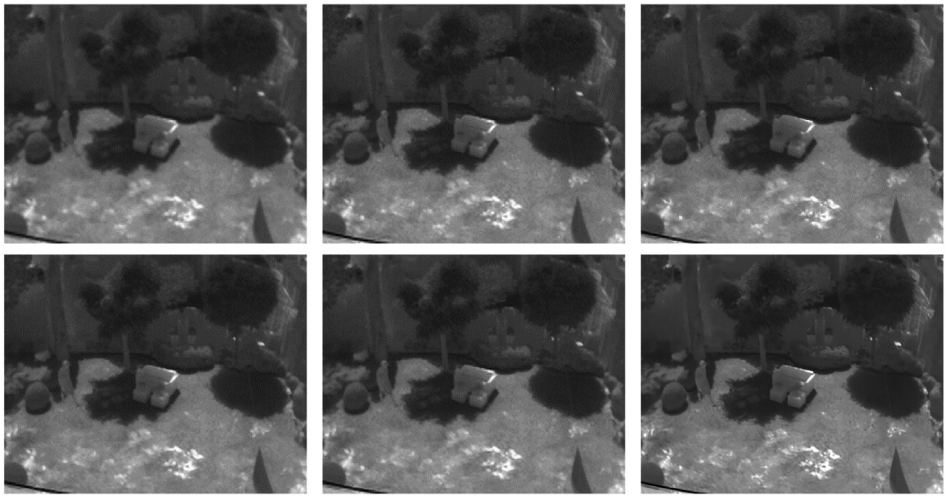

Fig. 9. Results of the Infrared image magnified by a factor of 2. Left to right: Bicubic interpolation, SCSR, SRCNN, VDSR, our method and the original HR image.

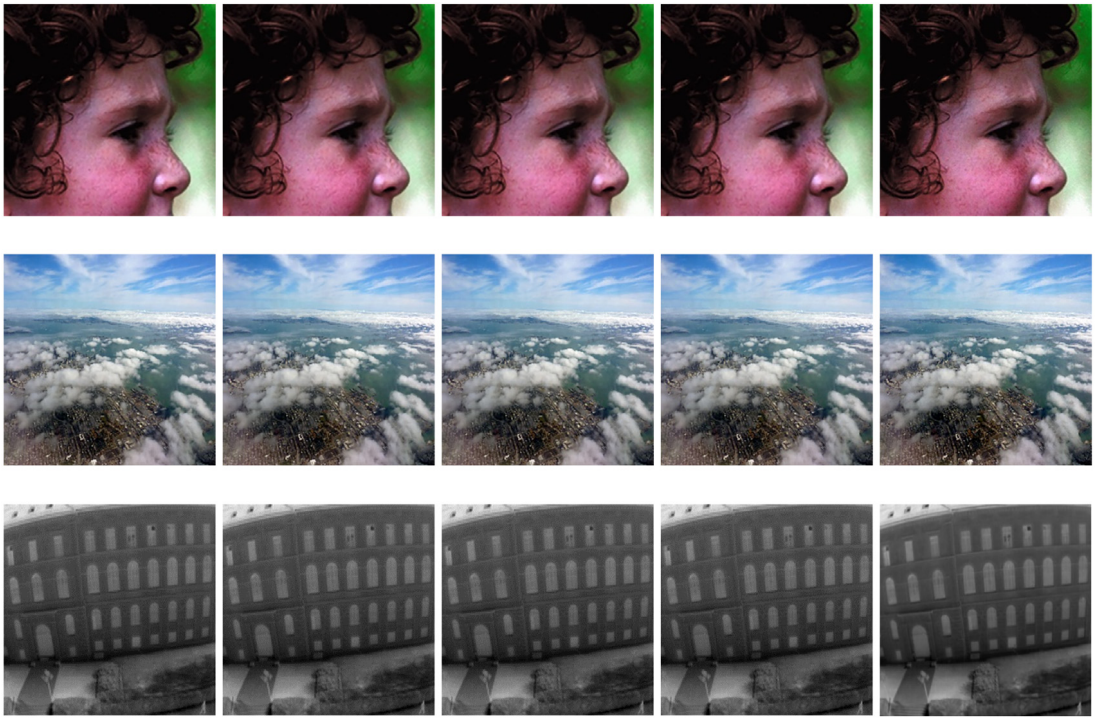

Fig. 10. The effects of dictionary size on the super-resolution reconstruction of Nature image, Remote sensing image and Infrared image. From left to right: dictionary size 256, 512, 1024, 2048 and the original HR image respectively.

Table 1
The PSNRs and SSIMs of the reconstructed images using super-reconstruction algorithm of different methods.

Algorithm\Image set	Scale	Natural image PSNR/SSIM	Remote sensing image PSNR/SSIM	Infrared image PSNR/SSIM	B100 PSNR/SSIM	Urban00 PSNR/SSIM
Bicubic	×2	33.86/0.9299	36.52/0.9534	36.36/0.9414	28.56/0.8431	27.28/0.8403
	×4	28.35/0.8104	32.65/0.8927	31.77/0.8952	25.16/0.6675	23.34/0.6577
SCSR	×2	35.49/0.9537	37.62/0.9604	37.71/0.9647	30.21/0.8763	28.70/0.8836
	×4	29.28/0.8503	33.02/0.9152	32.96/0.9061	25.83/0.7024	23.92/0.7073
SRCNN	×2	36.49/0.9537	38.65/0.9764	38.82/0.9815	31.08/0.8855	29.54/0.8902
	×4	30.31/0.8619	34.24/0.9358	34.53/0.9291	26.34/0.7156	24.79/0.7374
VDSR	×2	36.92/0.9642	39.02/0.9835	39.26/0.9842	31.36/0.8879	29.86/0.8956
	×4	30.48/0.8628	34.78/0.9423	34.90/0.9338	26.90/0.7201	24.63/0.7221
Ours	×2	37.43/0.9587	39.53/0.9912	39.71/0.9924	31.80/0.8960	30.67/0.9140
	×4	31.25/0.8838	35.36/0.9532	35.09/0.9405	27.19/0.7251	25.11/0.7524

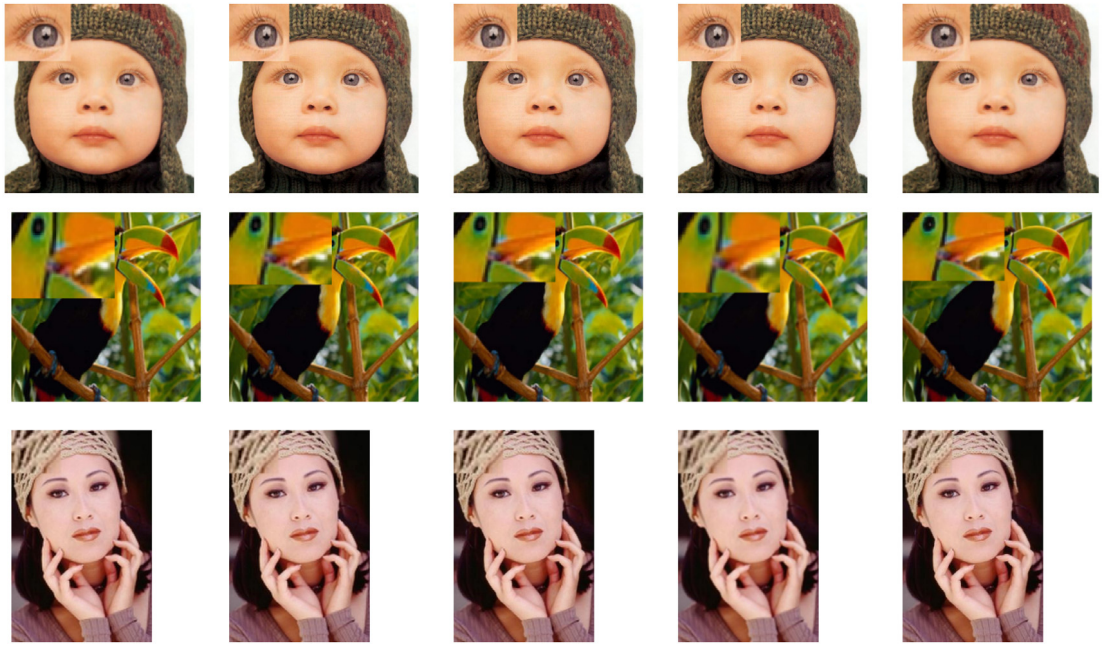

Fig. 11. The effects of feature extraction methods on the super-resolution reconstruction. From left to right: SCSR,SCSR using Gabor transform, SCSR using NSCT transform, MFSCSR and the original HR image.

Table 2
Comparison of network layers and parameters of four algorithms.

	Depth	Parameters
SCSR	–	64k
SRCNN	3	57k
VDSR	20	665k
MRMFSCSR	4	240k

Table 4
The PSNRs of the reconstructed images using different feature extraction methods.

Image\Input	SCSR		Gabor		NSCT		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Baby	38.27	0.963	38.26	0.962	38.36	0.963	38.38	0.963
Woman	33.61	0.959	33.70	0.959	34.26	0.961	34.61	0.965
Bird	38.47	0.980	38.64	0.980	39.30	0.981	39.85	0.983
Average	36.78	0.967	36.86	0.967	37.30	0.968	37.61	0.970

(c) Residual structure

The MRMFSCSR algorithm introduces the feature fusion idea, and uses multiple residual structure to fuse the features of layers 1, 2, 3, and 4, enriching high-frequency detail information. This section compares two network structures, where our method represents the MRMFSCSR algorithm, and Ours_no represents a simplified network of our proposed network which only has one traditional residual structure as VDSR. The parameters are unified except for the residual structure settings. Fig. 12 shows a comparison of reconstructed high-frequency detail residuals, and Table 5 gives objective indicators of reconstructed images.

It can be seen from the experimental results that under the same conditions of other parameters in the network model, introduction of feature fusion makes the reconstructed residual image is closer to the real residual image, which can increase the PSNR of the reconstructed super-resolution image by about 0.22 dB.

Table 5
The PSNRs of the reconstructed images using residual structure of different number.

Image\Input	Ours_no		Ours	
	PSNR	SSIM	PSNR	SSIM
ppt	32.240	0.918	32.411	0.926
Bridge	31.339	0.904	31.642	0.910
Coastguard	32.110	0.913	32.299	0.917
Average	31.896	0.912	32.117	0.918

5. Conclusion

Based on the SCSR algorithm and VDSR network, in order to further improve the image reconstruction quality, the image super-resolution reconstruction algorithm MRMFSCSR is proposed, which combines with MR network and MFSCSR. Considering the feature diversity of image blocks and the advantages of semantic enhancement of image feature fusion, so the algorithm can effectively maintain the image geometric features, and the performance is significantly improved. Compared with the SCSR algorithm and the VDSR algorithm, the PSNR values are improved 2 dB and 0.544 dB respectively at the commonly upscaling of 2.

Table 3
The PSNRs of the reconstructed images using dictionaries of different sizes.

Dictionary size\Image	Nature image		Remote sensing image		Infrared image	
	SCSR	Our method	SCSR	Our method	SCSR	Our method
D256	35.891	36.353	37.103	38.876	37.114	38.095
D512	35.924	36.576	37.352	38.051	37.121	38.167
D1024	35.012	36.772	37.378	38.103	37.124	38.172
D2048	35.090	36.889	37.463	38.118	37.127	38.208
Average	35.480	36.648	37.324	38.287	37.120	38.139

Fig. 12. The effects of Residual structure from network. From left to right, the residual between upsample image and original HR image, the residual result of MR network which has only one block , the residual result of MR network which has three blocks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61801359, Grant 61571345.

References

[1] H.S. Hou, H.C. Andrews, Cubic spline for image interpolation and digital filtering, IEEE Trans. Signal Process. 26 (1978) 508–517.
[2] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, Soft edge smoothness prior for alpha channel super-resolution, in: IEEE Conference on Computer Vision and Pattern Classification (CVPR), 2007, pp. 1–8.
[3] J. Sun, Z. Xu, H. Shum, Image super-resolution using gradient profile prior, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[4] R.C. Hardie, K.J. Barnard, E.A. Armstrong, Joint map registration and high-resolution image estimation using a sequence of undersampled images, IEEE Trans. Image Process. 6 (1997) 1621–1633.
[5] S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super-resolution, IEEE Trans. Image Process. 13 (2004) 1327–1344.
[6] M.E. Tipping, C.M. Bishop, Bayesian image super-resolution, in: Advances in Neural Information and Processing Systems 16 (NIPS), 2003.
[7] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Trans. Pattern Anal. Mach. Intell. 24 (9) (2002) 1167–1183.
[8] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[9] F. Yeganli, M. Nazzal, M. Unal, H. Ozkaramanli, Image super resolution via sparse representation over coupled dictionary learning based on patch sharpness, in: Proc. EMS, Prague, Czech Republic, Oct. 2014, pp. 203–208.
[10] Y. Zhang, W. Wu, Y. Dai, X. Yang, B. Yan, W. Lu, Remote sensing images super-resolution based on sparse dictionaries and residual dictionaries, in: Proc. DASC, Chengdu, China, Dec. 2013, pp. 318–323.
[11] C.-H. Fu, H. Chen, H. Zhang, Y.-L. Chan, Single image super resolution based on sparse representation and adaptive dictionary selection, in: Proc. DIP, Hong Kong, Aug. 2014, pp. 449–453.
[12] 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) (2011) 1838–1857.
[13] R. Timofte, V. De Smet, L. Van Gool, A+: Adjusted anchored neighborhood regression for fast super-resolution, in: Asian Conference on Computer Vision (ACCV), Springer, 2014, pp. 111–126.
[14] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European Conference on Computer Vision (ECCV), Springer, 2014, pp. 818–833.
[15] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in: International Conference on Learning Representations (ICLR), 2016.
[16] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.
[17] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: IEEE International Conference on Computer Vision, 2015.
[18] Deeply-recursive convolutional network for image super resolution, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[19] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[20] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.