Super resolution reconstruction for medical image based on adaptive multi-dictionary learning and structural self-similarity

Fang Zhang\textsuperscript{a,b}, Yue Wu\textsuperscript{b}, Zhitao Xiao\textsuperscript{a,b}, Lei Geng\textsuperscript{a,b}, Jun Wu\textsuperscript{a,b}, Jia Wen\textsuperscript{a,b}, Wen Wang\textsuperscript{a,b} and Ping Liu\textsuperscript{b}

\textsuperscript{a}Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin, China; \textsuperscript{b}School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, China

ABSTRACT
To improve the quality of the super-resolution (SR) reconstructed medical images, an improved adaptive multi-dictionary learning method is proposed, which uses the combined information of medical image itself and the natural images database. In training dictionary section, it uses the upper layer images of pyramid which are generated by the self-similarity of low resolution images. In reconstruction section, the top layer image of pyramid is taken as the initial reconstruction image, and medical image's SR reconstruction is achieved by regularization term which is the non-local structure self-similarity of the image. This method can make full use of the same scale and different scale similar information of medical images. Simulation experiments are carried out on natural images and medical images, and the experimental results show the proposed method is effective for improving the effect of medical image SR reconstruction.

KEYWORDS
Super-resolution reconstruction; medical image; improved adaptive multi-dictionary learning; non-local structural similarity

1. Introduction
With the continuous innovation of medical imaging technology and computer technology, medical image processing is becoming important in clinical diagnosis. It can help doctors fully understand the structure of the internal pathology of patients, and then an accurate treatment plan is developed. High-quality medical images can not only improve the level of clinical diagnosis, but also provide necessary support for medical research, teaching, surgery, etc. Commonly used methods of super-resolution (SR) reconstruction are interpolation methods, such as B-spline and bicubic interpolation etc. Through SR reconstruction, the clarity of medical images and the correctness of disease diagnosis, can be improved. Thereby it can help doctors to perform precise treatment on patients.

A single HR image is generated by one or more low-resolution images in the same scene based on SR method \cite{1}. And for SR method, it is the biggest difficulty to recover high frequency information lost during process of getting LR inputs \cite{2,3}.

Yang et al. \cite{4} propose a SR algorithm based on sparse representation. In this algorithm, FSS (Feature-Sign Search) method plans to generate low-resolution (LR) and high-resolution (HR) dictionaries by using the patch pairs of HR and LR image. LR dictionary produces the spare coefficients of LR image patches, and it uses the coefficients to produce the HR image patches based on the HR dictionary.

For each local patch, pre-learned dictionaries from a dataset of high quality example patches are selected based on Dong \cite{5}. In the process of reconstruction, non-local structure self-similarity (NLSS) and autoregressive model are considered as the regularization term. Dong’s method uses the self-similarity of the image that increases the accuracy and reliability accessed to the information. Therefore, it still uses the image database to train the global dictionary, and it cannot be sparse representation for all image patches. In order to further improve the reconstruction effect, A SR method based on adaptive multi-dictionary learning (MDL) was proposed by Pan \cite{6}. Firstly, patches of LR image pyramid which are packed into some categories are classified into one of these groups, if they meet special conditions. And then, corresponding dictionaries are generated for each group based on MDL.

This paper proposes an improved method based on adaptive MDL and structural self-similarity. In this method, we use the pyramid generated by structural self-similarity upper layer images as samples instead of...
Pan’s method. It uses the pyramid decomposition to obtain samples when training a dictionary, which will achieve more self-similar information of the different scales. In addition, the paper uses the top layer image of pyramid as the initial reconstruction image. SR image is reconstructed by regularization term using the non-local structure self-similarity of image. It also employs self-similar information of the same scale to reconstruct image. Experimental results demonstrate that the HR medical image reconstructed by the proposed method has good effect for medical images.

2. SR model
2.1. SR model based on sparse representation

Supposing $\Phi \in \mathbb{R}^{N \times M}$ ($N<M$) is an over-complete dictionary of $M$ prototype signal-atoms, a sparse linear combination of these atoms can represent a signal. The signal vector can be expressed as:

$$x = \Phi \alpha$$

(1)

$\alpha \in \mathbb{R}^M$ is the sparse coefficient with very few ($\ll M$) nonzero term. Solving an $l_0$-minimization problem can calculate the sparse decomposition coefficient, and it can be formulated as:

$$\alpha = \arg \min_{\alpha} \| \alpha \|_0, \text{s.t.} \| x - \Phi \alpha \|_2 \leq \varepsilon$$

(2)

where $\| \cdot \|_0$ is a pseudo norm that counts the number of non-zero entries in $\alpha$, and approximation error is influenced by a small constant $\varepsilon$. Minimization $l_0$ is converted to the convex $l_1$ minimization. It is an issue about NP-hard combinatorial optimization. According to sparse coding problem, $l_1$-norm is summarized as follows:

$$\alpha = \arg \min_{\alpha} \| \alpha \|_1, \text{s.t.} \| x - \Phi \alpha \|_2 \leq \varepsilon$$

(3)

Eq. (3) also can be written as the form of Lagrange polynomials:

$$\alpha = \arg \min_{\alpha} \left\{ \| x - \Phi \alpha \|_2^2 + \lambda \| \alpha \|_1 \right\}$$

(4)

where $\lambda$ represents the regularization parameter, it is a constant.

The single image SR problem can be presented as: a HR image $X$ of the same scene can be recovered by giving a LR image $Y$. Assuming the LR image is the version of the solution $X$. And the version is blurred and down-sampled:

$$Y = DBX$$

(5)

where $B$ is a fuzzy filter, and $D$ is a lower sampling operator. The process of solving $X$ by $Y$ is the process of image SR reconstruction, which is to solve the following least squares problem:

$$X = \arg \min_x \| Y - DBX \|_2^2$$

(6)

The sparse representation prior on small patches $x$ of $X$ is used to make HR images $X$ satisfy the reconstruction constraint for a given LR input $Y$.

$x$ is an image column vector reshaped from image $X$. $R_l$ is an extracting-matrix, whose function is extracting patch $x_i$ from $X$. $x_i = R_l X$, $i = 1, 2, \ldots, N$ is the $i$th patch (size: $\sqrt{n} \times \sqrt{n}$) vector of $x$.\{$\Phi_k$\}, $k = 1, 2, \ldots, M$ is defined to be a set of orthonormal sub-dictionaries. $x_i = \Phi_k \alpha_i$ can approximately express the sparse coding $x_i$, $\| \alpha_i \|_1 \leq T$ and the reconstructed patches $\hat{x}_i$ are averaged to reconstruct the whole image $X$. And it can be described as:

$$\hat{X} = \Phi^* \alpha = \left( \sum_{i=1}^{N} R_l^T R_l \right)^{-1} \sum_{i=1}^{N} \left( R_l^T \Phi_k \alpha_i \right)$$

(7)

where $\alpha$ represents a collection of all $\alpha_i$. And $\Phi$ represents a collection of all sub-dictionaries $\{\Phi_k\}$.

For $Y = DBX$, recovering the original image $X$ from $Y$ is our goal. A super-resolution reconstruction model based on sparse representation through Eqs. (6) and (7) can be summarized as follows:

$$\hat{x} = \arg \min_{\alpha} \left\{ \| Y - DB \Phi^* \alpha \|_2^2 + \lambda \| \alpha \|_1 \right\}$$

(8)

$l_1$-norm sparsity regularization term $\| \alpha \|_1$ is weighted by the constant $\lambda$. And the $l_0$-norm sparsity is closer to the reweighted $l_1$-norm sparsity than using a constant weight according to Ref [13]. Therefore, Eq. (8) can be summed up as:

$$\hat{x} = \arg \min_{\alpha} \left\{ \| Y - DB \Phi^* \alpha \|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{ij} \| \alpha_{ij} \| \right\}$$

(9)

where $\lambda_{ij}$ represents the weight assigned to $\alpha_{ij}$, $\alpha_{ij}$ represents the coefficient associated with the $j$th atom of $\Phi_k$. $\lambda_{ij}$ that can be obtained by the following formula:

$$\lambda_{ij} = \frac{c}{|\alpha_{ij}^{(l)}| + \varepsilon}$$

(10)

where $c$ is a constant, $\alpha_{ij}^{(l)}$ is an estimated value of the first $l$ iteration of $\alpha_{ij}$ and $\varepsilon$ is a small constant.

2.2. SR model combined with structure self-similarity

In general, many duplicate images contain a lot of redundant information. And the quality of reconstructed images is significantly improved. Therefore, a
non-local self-similar regularization term based on sparsity of the image SR framework is proposed. In the whole image, similar patches \( x_i'(l = 1, 2, \ldots, L) \) are found for each local patch \( x_i \). If \( e_i' = \| x_i - x_i' \|_2^2 \leq t \), similar patch will choose a patch \( x_i' \). \( x_i, x_i' \) are the current estimated value based on the preset threshold \( t \). And the patch \( x_i' \) within the first \( L \) closest patches is also selected to \( x_i \), where \( L \) is also a preset threshold. The center pixel of patch \( x_i \) and \( x_i' \) are \( \chi_i \) and \( \chi_i', \) respectively. The weighted average of \( \chi_i' \) is predicted based on \( \chi_i \). The weight \( b_i' \), which is allocated to \( \chi_i' \), is assumed as \( b_i' = \exp(-e_i'/h)/\sum_{i=1}^{L} \exp(-e_i'/h) \), where \( h \) is a weight control factor. In this paper, the difference between \( \chi_i \) and \( \chi_i' \) is used as a regularization term, i.e. \( \| \chi_i - \sum_{i=1}^{L} b_i' \chi_i' \|_2^2 \). So the sparse representation and structure self-similarity are combined to obtain the image super resolution reconstruction model:

\[
\hat{x} = \arg \min_{\alpha} \left\{ \| Y - DB\Phi^x \alpha \|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{ij} \| x_{ij} \| + \eta \cdot \sum_{i=1}^{N} \| \chi_i - \sum_{i=1}^{L} b_i' \chi_i' \| \right\} 
\]

where \( \eta \) is to balance the contribution of non-local regularization, it is a constant.

Then by merging the first term and third term, the simplified form of Eq. (11) can be expressed as

\[
\hat{x} = \arg \min_{\alpha} \left\{ \| \tilde{Y} - F \Phi^x \alpha \|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{ij} \| x_{ij} \| \right\}
\]

where

\[
\tilde{Y} = \begin{bmatrix} Y \\ 0 \end{bmatrix}, \quad F = \begin{bmatrix} DB \\ \eta \cdot (I - A) \end{bmatrix},
\]

A expresses the weight matrix, and \( I \) expresses the identity matrix.

3. Acquisition and training of adaptive multi-dictionary samples

3.1. Selection of samples

In order to get the missing high-frequency information, it is necessary to obtain the dictionary that contains the high-frequency information. In this paper, the dictionary training samples is consisted of image itself to be processed and the natural image database. The natural image database is mainly to select high resolution images with rich texture details. For the image itself, we generate the image pyramid by self-similarity to produce the training samples. It contains the image’s own information, and the upper layer images of pyramid are selected as the samples of a dictionary. The principle of generating the image pyramid by self-similarity is as follows.

Figure 1 shows the input image \( I_{in} \), and we generate LR images \( I_{in-k} \) \((k = 1, 2, \ldots, m)\) as the lower layer images of the pyramid firstly. In the lower-resolution images \( I_{in-1} \) or \( I_{in-2} \), some similar patches \((P_1 \text{ and } P_2)\) are searched where \( P_1 \) is a source patch in \( I_{in} \). A corresponding region \( Q_1 \) or \( Q_2 \) in \( I_{in} \) are determined for each patch \((P_1 \text{ or } P_2)\). The corresponding region \( D_1 \) or \( D_2 \) depends on two factors: (1) the area of source patch \( P_1 \), and (2) the layer index of the found patch \((-1 \text{ of } I_{in-1}, \text{ or } -2 \text{ of } I_{in-2})\). In the end, \( Q_1 \) is moved to the corresponding position of the enlarged area \( D_1 \), and \( Q_2 \) is also the case. \( P_1 \) is not exactly like \( P_1 \) and \( Q_k \) is not exactly like \( D_k \). So \( D_k \) can be obtained by \( Q_k \) weighted, and the formula is as follows:

\[
D_k = \sum \omega_k Q_k
\]

where their weights are

\[
\omega_k = \exp \left(-\| P_i - P_k \|^2 / \sigma^2 \right)
\]
in $l_p$). Uncovered area will be filled with the back projection to improve image resolution based on method [6].

3.2. Training of adaptive multi-dictionary

As mentioned above, the dictionary has an important impact on the reconstruction results. In order to enrich the high frequency information contained in the dictionary, this paper adopts the method of combining the image itself and the natural images database. Firstly, we take the upper layer images of pyramid as the training samples $S$. The semantic information is conveyed by the edge of image which is sensitive to human visual system. The high-pass filtering is used to obtain the feature of each patch with size $\sqrt{n} \times \sqrt{n}$ for the training samples $S$ of clustering. Blocks only with a certain edge structure are used in dictionary learning to exclude the smooth patches from training. We select the intensity variance of image patch which is better than a threshold $\Delta$. The selected image patch is denoted by $\tilde{s}_i, i=1,2,\ldots,n$, and $\tilde{g}_i, i=1,2,\ldots,n$ is the high frequency component of $\tilde{s}_i$. Then we compute the distance between $\tilde{g}_i$ and the class centers $\mu_k$, which is denoted by $d_k^i$. If $\min(d_k^i) < \delta_k$, $\tilde{g}_i$ is added to $C_k$, otherwise the image block is abandoned, where $\delta$ is a parameter that is used to control the similarity degree between the image blocks and the class centers. After the above operation, the samples are expanded, and each type of expanded sample is recorded as $C_k = \{g^k_1, g^k_2, \ldots, g^k_{q_k}\}$, where $q_k$ is the number of samples after the expansion.

Then, according to the high frequency component samples $C_k$, we can find the corresponding image block sample matrix $S_k = [s^k_1, s^k_2, \ldots, s^k_{q_k}] \quad k = 1,2,\ldots,K$. The following objective function can formulate the design of sub-dictionary $\Phi_k$.

$$\left(\Phi_k, \Lambda_k\right) = \arg\min_{\Phi_k, \Lambda_k} \left\{ \| S_k - \Phi_k \Lambda_k \|^2_F + \lambda \| \Lambda_k \|_1 \right\} \quad (16)$$

where $\Lambda_k$ is the representation coefficient matrix, and $\lambda$ is the parameter controlling the degree of sparsity. In order to reduce the amount of computation, the principal component analysis (PCA) is used to calculate the main parts for each sub-dataset $S_k$. An orthogonal transformation matrix $P_k$, which is generated by applying PCA, is set the dictionary, and it meets $Z_k = P_k^T S_k$.

Then we will have $\| S_k - P_k Z_k \|^2_F = \| S_k - P_k P_k^T S_k \|^2_F = 0$.

In order to obtain a dictionary $\Phi_k = [p^k_1, p^k_2, \ldots, p^k_r]$, the first $r$ which is the most important eigenvectors in $P_k$ is separated. In this way, the relationship of $l_1$-norm regularization term and $l_2$-norm approximation term in Eq. (16) is more balanced. The reconstruction error $\| S_k - \Phi_k \Lambda_k \|^2_F$ will increase the decrease of $r$ according to Eq. (16). Considering the term $\| \Lambda_k \|_1$ will reduce, the best value $r_0$ of $r$ can be prescribed for:

$$r_0 = \arg\min_{r} \left\{ \| S_k - \Phi_k \Lambda_r \|^2_F + \lambda \| \Lambda_r \|_1 \right\} \quad (17)$$

where $\lambda$ is the weight of the $l_1$ norm sparse regularization term $\| \alpha \|_1$. In the end, the sub-dictionary which is learned from sub-dataset $S_k$ is $\Phi_k = [p^k_1, p^k_2, \ldots, p^k_r] \quad k = 1,2,\ldots,K$.

Finally, we seek an adaptive dictionary for the image patches to be reconstructed. Considering spanning the adaptive sparse domain, a sub-dictionary is assigned adaptively to each partial patch of $x$. $X$ needs
to be estimated in advance, because it is not known at the beginning. We can take the top layer image of pyramid as the initial estimation of $\hat{X}$. $\hat{X}$ represents the estimate of $X$, and $\hat{x}_i$ represents the local patch of $\hat{X}$. For the centroid $\mu_k$, considering the high-pass filtered patch of $\hat{g}_i$, the best fitted sub-dictionary to $\hat{x}_i$ is generated, which is of each cluster available. For $\hat{x}_i$, the dictionary is influenced by the minimum distance between $\hat{g}_i$ and $\mu_k$, i.e.

$$k_i = \text{argmin}_k \| \hat{g}_i - \mu_k \|_2$$

(18)

### 3.3. Algorithm Flow

In the paper, in order to solve Eq. (12), we use the iterative shrinkage algorithm [7]. The algorithm flow chart is shown in Figure 2.

The overall algorithm process is as follows:

1. **Initialization:**
   a. By the descending process of the image, we determine the down-sampling matrix $D$ and the blurring matrix $B$;
   b. The sub-dictionary $\Phi_k$ is established by PCA, in which samples from the natural images and the upper layer images of the pyramid by structural self-similarity property are established;
   c. The weight matrix $A$ of the non-local structure self-similarity is obtained by searching the similar image patches of the same scale in the image;
   d. The termination error $\epsilon$, the maximal iteration number $\text{MaxIter}$, the constant of controlling non-local regularization term $\eta$, and the
2. It is iterated on $k$ until $k \geq \text{MaxIter}$ or $\| \tilde{X}^{(k)} - \tilde{X}^{(k-1)} \|_2^2/N \leq \epsilon$ is satisfied.
   a. The current estimation is updated using Eq. (19):
      \[
      \tilde{X}^{(k)} = \tilde{X}^{(k-1)} + F^T \left( \tilde{Y} - F \tilde{X}^{(k-1)} \right)
      \] (19)
   b. The sparse representation coefficients are updated using Eq. (20) and Eq. (21):
      \[
      \alpha_{ij}^{(k+1/2)} = \Phi_k R \tilde{X}^{(k+1/2)}
      \] (20)
      \[
      \alpha_{ij}^{(k+1)} = \text{soft} \left( \alpha_{ij}^{(k+1/2)}, \tau_{ij} \right)
      \] (21)
   where $\text{soft}(-, \tau)$ is a soft function of thresholding, and its threshold is $\tau$:
      \[
      \text{soft}(x, \tau) = \text{sgn}(x) \max(|x| - \tau, 0)
      \] (22)
   where $\text{sgn}(x)$ is a sign function;
   c. Re-update the current estimation:
      \[
      \tilde{X}^{(k+1)} = \Phi \tilde{X}^{(k)};
      \]

\begin{table}[ht]
\centering
\caption{Comparison of PSNR (dB)/SSIM of the medical images ($\sigma = 0$) based on various methods.}
\begin{tabular}{|c|c|c|c|}
\hline
Test Images & Bicubic & ASDS & Our \\
\hline
Head1 & 22.26 & 27.14 & 27.40 \\
& 0.7371 & 0.8906 & 0.8970 \\
Head2 & 24.15 & 29.46 & 29.86 \\
& 0.8005 & 0.9136 & 0.9166 \\
Chest & 21.23 & 24.16 & 24.66 \\
& 0.4490 & 0.5597 & 0.5666 \\
\hline
\end{tabular}
\end{table}

\begin{table}[ht]
\centering
\caption{Comparison of PSNR (dB)/SSIM of the medical images ($\sigma = 5$) based on various methods.}
\begin{tabular}{|c|c|c|c|}
\hline
Test Images & Bicubic & ASDS & Our \\
\hline
Head1 & 22.09 & 25.03 & 25.51 \\
& 0.5662 & 0.6310 & 0.6385 \\
Head2 & 23.87 & 27.13 & 27.51 \\
& 0.6046 & 0.6288 & 0.6306 \\
Chest & 21.09 & 23.40 & 23.83 \\
& 0.3742 & 0.4254 & 0.4391 \\
\hline
\end{tabular}
\end{table}

Figure 6. Comparison of the reconstructed images of various methods for noisy medical image (1) (a) Original image (b) LR image (c) Bicubic (d) ASDS (e) Our.

Figure 7. Comparison of the reconstructed images of various methods for noisy medical image (2) (a) Original image (b) LR image (c) Bicubic (d) ASDS (e) Our.

Figure 8. Comparison of the reconstructed images of various methods for noisy medical image (3).
d. If \( \text{mod}(k, P) = 0 \), the improved estimate \( \hat{X}^{(k+1)} \) is used to update the matrix \( A \) and the adaptive sparse domain of \( X \).

4. Experimental results and analysis

This section mainly presents the results of the SR reconstruction of medical images by this method. It also performs three sets of experiments in both noiseless and noisy cases. The three sets of experimental images (Head1, Head2, and Chest) are actually taken by the hospital, including MRI images and CT images. In the experiments, the degraded LR image is generated by a \( 7 \times 7 \) gaussian kernel and an original image with a standard deviation of 1.6 make the degraded LR image. Then the image is down sampled by a factor of 3. Firstly, the low resolution color images are converted to ycbcr format. Finally, the bicubic interpolator is applied to the color components. We compare our method with two most advanced methods: the bicubic interpolation method and the ASDS method [8]. The statistical experimental results are analyzed and discussed to verify the advantages of this method for medical image’s SR reconstruction. The results with noiseless are shown in Figures 3–5 and Table 1. The results with noise are shown in Figures 6–8 and Table 2.

Firstly, Figures 3–8 show the reconstructed results for the three images of Head1, Head2, and Chest. The texture of the head and the texture of lungs in the chest can be observed clearly. The blur and noise conditions are greatly improved. Compared with the bicubic algorithm and the ASDS method, the medical images reconstructed by our method are of better quality, and more detailed.

Secondly, we compare our method with the two method for medical images respectively for quantitatively prove the advantages of our method. Evaluating the experimental results based on various methods use the SSIM and PSNR [9]. The greater value of the PSNR indicates the quality of the reconstructed image and the reconstructed results are better. SSIM evaluates the structure differences between the reconstructed image and the high resolution image by SR. The greater the value of SSIM is, the better the reconstructed result will be. Table 1 and Table 2 list the SSIM and PSNR results, and it can be seen from the results in the table that the method presented in this paper has good results.

In addition, we compare the complexity and time consumption of the bicubic interpolation method, the ASDS method and our method. The calculation formula of time complexity is \( T(n) = O(f(n)) \), where \( f(n) \) is the frequency of the sentence with the largest frequency in the algorithm. The program of time consuming test method is run on Matlab R2014b, and the average time consumption of the 512 × 512 pixel test chart is calculated. The test results are shown in Table 3.

Furthermore, we give the experimental results on the public image library for fully prove the advantages of our method. We randomly selected 20 high quality natural images from the sample database. Figure 9 shows the results of comparing our method with two most advanced methods: the bicubic interpolation method and the Yang method. From the result we can see the proposed method results achieve the most visual effect, which is more clearly compared

| Test Images | Bicubic | Yang | Our |
|-------------|---------|------|-----|
| hat         | 28.27/0.8005 | 29.09/0.8263 | 31.44/0.8793 |
| castle      | 25.61/0.7712 | 26.12/0.7935 | 26.98/0.8292 |
| Input3      | 27.36/0.8892 | 28.51/0.9163 | 33.03/0.9643 |
| average     | 27.08/0.8203 | 27.91/0.8454 | 30.48/0.8909 |

Figure 9. Comparison of the reconstructed images of various methods for Butterfly image (a) Original image (b) Bicubic (c) Yang (d) Our.
with the other algorithms because more details are restored. Secondly, for quantitatively demonstrate the advantages of our method, we use PSNR and SSIM to evaluate the experimental results based on various methods. Table 4 list the SSIM and PSNR consequence of these algorithms, where it can be seen that our method is more effective than others. The average PSNR gains over other methods are up to 3.4 dB and 2.57 dB. Compared with other methods, the average value of the SSIM gains up to 0.0706 and 0.0455, respectively.

5. Conclusions
In this paper, we introduce a single image super resolution reconstruction method based on adaptive MDL and structural self-similarity. The proposed method is the improvement for the existing adaptive MDL method. In this method, the image pyramid is generated by structural self-similarity, and the pyramid upper layer images are as samples, which will contain more self-similar information of the different scales than directly conducting the pyramid decomposition for the image itself. In addition, the paper uses the top layer image of pyramid as the initial reconstruction image and the image SR reconstruction uses the non-local structure self-similarity of the image as the regularization term. It also added the self-similar information of the same scale to the reconstruction image. Experimental results demonstrate that our method achieves better results in terms of all the visual perception and the parameters of PSNR and SSIM compared with multiple super resolution reconstruction methods.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This study was sponsored by the National Natural Science Foundation of China (NSFC) (grant no. 61601325), the Tianjin Natural Science Foundation (grant no. 17JCQNJC01400), and Supported by the Program for Innovative Research Team in University of Tianjin (grant no. TD13-5034).

References
[1] Xie B, Duan ZM, Ma PG, et al. SR reconstruction algorithm of infrared image based on dynamic pyramid model[J]. Infrared and Laser Eng. 2018;001(1):277–282.
[2] Pan ZX, Yu J, Hu SX, et al. Super-resolution based on compressive sensing and structural self-similarity for remote sensing images. IEEE Transac Geosci Remote Sens. 2013;51(9):4864–4876.
[3] Zhao H, Wei JX, Pang ZH, et al. Wave-front coded super-resolution imaging technique. Infrared and Laser Eng. 2016;45(4):227–236.
[4] Yang J, Wright J, Huang TS, et al. Image super-resolution via sparse representation. IEEE Transac Image Proc. 2010;19(11):2861.
[5] Dong WS, Zhang L, Shi GM, et al. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Transac Image Proc. 2011;20(7):1838–1857.
[6] Pan ZX, Yu J, Xiao CB, et al. Single image super resolution based on adaptive multi-dictionary learning. Acta Electronica Sinica. 2015;43(2):209–216.
[7] Dong WS, Zhang L, Lukac R, et al. Sparse representation based image interpolation with nonlocal autoregressive modeling. IEEE Transac Image Proc. 2013; 22(4):1382–1394.
[8] Zhang A, Guan C, Jiang H, et al. An image super-resolution scheme based on compressive sensing with PCA sparse representation. In: Shi YQ, Kim HJ, editors. The international workshop on digital forensics and watermarking 2012. Berlin Heidelberg: Springer; 2013.
[9] Zhang X, Zhou W, Duan Z. Image super-resolution reconstruction based on fusion of K-SVD and semi-coupled dictionary learning. IEEE Signal and Information Processing Association Summit and Conference; 2017 December 13–16, Jeju, South Korea.