Neurocomputing 422 (2021) 62–84

Contents lists available at ScienceDirect

Neurocomputing

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

Joint image fusion and super-resolution for enhanced visualization via semi-coupled discriminative dictionary learning and advantage embedding

Huafeng Li [a,b,1], Moyuan Yang [a,b,1], Zhengtao Yu [a,b,*]

[a] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, Kunming 650500, PR China
[b] Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, PR China

ARTICLE INFO

Article history:
Received 24 January 2020
Revised 23 July 2020
Accepted 8 September 2020
Available online 28 September 2020
Communicated by Pingkun Yan

Keywords:
Image fusion
Super-resolution
Dictionary learning
Low-rank decomposition
Structure information compensation

ABSTRACT

In recent years, image fusion has attracted more and more attention, and many excellent methods have emerged. However, only a few studies on joint image fusion and super-resolution have been carried out, and the performance of existing methods is far from that of simple image fusion. To tackle such problem, we propose a novel joint fusion and super-resolution framework based on discriminative dictionary learning. Specifically, we first jointly learn two pairs of low-rank and sparse dictionaries (LRSD) and a conversion dictionary. One pair is used to represent the low-rank and sparse components of low-resolution input images, and the other is used to reconstruct high-resolution fused result; the conversion dictionary is used to establish the relationship between coding coefficients of low-resolution image and high-resolution image. To compensate for the loss of details, structure information compensation dictionary (SICD) is also learned, and the lost information is compensated by SICD and thus visualization of final results is enhanced. To integrate advantages of excellent image fusion methods into the fused and reconstructed results, we propose a deconvolution-based advantage embedding scheme. The experimental results verify the effectiveness and advantages of our method over other competitive ones.

1. Introduction

Image fusion can integrate the complementary information about the same scene acquired by different sensors into a single image that can provide a more comprehensive and accurate description of this scene, thereby contributing to the identification of events and objects. Such technology has attracted an increasing attention of researchers and made significant research progress in recent years.

The existing image fusion methods can be roughly classified into three categories, namely, multiscale transform (MST) based methods [1–5], dictionary learning based methods [6–10], and deep learning based methods [11–15]. In MST based methods, the common used MSTs include wavelet transform [2,16,17], dual tree complex wavelet transform (DTCWT) [1,18], Shearlet Transform [19,20], curvelet transforms [21], contourlet transform [22], and nonsubsampled contourlet transform (NSCT) [23]. Usually, the bases of MST are fixed and show weak sparsity, thus MST cannot represent the local information of image adaptively and sparsely. Different from MST methods, dictionary leaning (DL) based sparse representation (SR) technique can effectively avoid the defects of the MST-based methods, and exhibit promising fusion performance. With the development of deep learning, image fusion based on deep learning has attracted more and more attention, accordingly, some excellent fusion methods have emerged [24–27]. However, the methods discussed above have an excellent performance only when the source images are of high resolution. If the input images are of low-resolution, the fused result will also be low in resolution, which hinders the application of the fused result. To improve the resolution, a separate processing procedure is a commonly used strategy. However, such method may introduce the artifacts created in the first step into the next step, degrading the visual quality of the final result.

To address this problem, Yin et al. [28] developed a sparse representation based method for simultaneous image fusion and super-resolution. However, this method does not embed dictionary learning and super-resolution reconstruction into a joint learning framework, and the super-resolution fusion can be only achieved

* Corresponding author at: Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, Kunming 650500, PR China.
E-mail addresses: ztyu@hotmail.com, ztychina99@126.com (Z. Yu).
[1] Equal contribution

https://doi.org/10.1016/j.neucom.2020.09.024
0925-2312/

by fusing the interpolated image and supplementing high-frequency information [28]. In [29], although low-resolution and high-resolution dictionary pairs are learned in a jointly trained framework, but such method assumes that the high-resolution image patches and their corresponding low-resolution versions share the same representation coefficients. This assumption is too strong to characterize the difference between high-resolution image and its corresponding low-resolution image [30]. Different from the above methods, Li et al. [31] developed a joint processing framework based on fractional differential and variational method for image fusion and super-resolution. However, such method cannot extract the complete information of the source image and inject it into the fused image, which results in poor visual quality of fused result.

In this paper, we propose a simple yet effective model to solve the problems of joint image fusion and super-resolution. The whole procedure of dictionary learning is illustrated in Fig. 1. Specifically, we first assume that the natural images can be modeled as a superposition of components with different spatial morphologies, and then decompose the input image into low-rank and sparse components. To characterize them more effectively, we propose to utilize different dictionaries to represent different components, and develop a semi-coupled discriminative dictionary joint learning framework to learn a pair of low-resolution representation dictionary, a pair of high-resolution representation dictionary and a conversion dictionary. The representation dictionaries are used to characterize the different components of low-resolution and high-resolution images, the conversion dictionary (CD) is used to reveal the intrinsic relationship between the coding coefficients of high-resolution image and its corresponding low-resolution image.

In addition, to compensate for the lost information as well as enhance the visualization of fusion and super-resolution results, we develop a novel model to learn SICD from the difference between reconstructed high resolution image and original image. In the learning of SICD, we assume that the information compensation components share the same coding coefficients as the sparse components of the high-resolution image. Throughout the training process, two pairs of discriminative representation dictionary are not fully coupled, allowing some deviations between the image patches of high-resolution image and corresponding low-resolution image. For this reason, the discriminative dictionary learning developed in this paper is called semi-coupled discriminative dictionary learning.

To obtain the coding coefficients of different components of input, we construct a sparse and low-rank separation model according to the morphological priori of low-rank and sparse components. The model can effectively decompose the input image into low-rank and sparse components under different morphological dictionaries, so that we can construct high-resolution fusion image by different dictionaries. With the learned coding coefficients, the corresponding high-resolution coefficients can be constructed according to the intrinsic relationship between coding coefficients of low-resolution and its corresponding high-resolution patches. After fusion and super-resolution reconstruction are completed, we propose to employ the deconvolution operation to integrate the advantages of mature and excellent image fusion algorithms into the fusion and super-resolution results so as to make further improve its visual quality. Moreover, to avoid losing the edge details of the source image and enhance the visualization of the fused results, we propose to combine the details compensation dictionary with the reconstructed coding coefficients of high-resolution sparse components to construct the detail compensation components, and inject them into the final reconstructed results. The overview of the developed fusion and super-resolution for two source images is shown in Fig. 2. The contributions and the major advantages of the above design are as follows:

(1) We proposed an efficient joint semi-coupled discriminative dictionary-learning model, which can learn a conversion dictionary and two pairs of discriminative dictionaries jointly for representing the low-rank and sparse components of low-resolution and high-resolution image pairs. Moreover, the relationship between the coding coefficients of low-resolution image and its corresponding high-resolution version is revealed by the CD, so our method can achieve image fusion and super-resolution simultaneously.

(2) To avoid the loss of edge details, we design a learning model to achieve the SICD. With this dictionary, the reverse compensation of lost information is realized. This design not only prevents the structural details from lossing, but also enhances the visualization of the final fusion and reconstruction results. A large number of experiments verified the superiority of this algorithm over the traditional methods.

(3) In order to integrate the excellent performance of existing image fusion algorithms into the results of our fusion and super-resolution, a deconvolution strategy is proposed. Benefiting from above design, our method can not only achieve image fusion and super-resolution simultaneously, but also effectively avoid the loss of detailed information and raise the visualization level of fusion and reconstruction results.

The rest part of this paper is arranged as follows: Section 2 briefly reviews the related researches about sparse representation, low-rank decomposition, and dictionary learning; the proposed method together with dictionary learning, image decomposition and optimization is presented in Section 3; experimental results and discussion are reported in Section 4; major conclusions are drawn in Section 5.

2. Related work

2.1. Low-rank and sparse representation in image fusion

Low-rank representation (LLR) and sparse representation (SR) are two types of representation learning. SR can model a signal as the liner combination of columns of an overcomplete dictionary to better express the significant information of the signal; while LRR can recovery the underlying low-rank structures from its noisy observations. Thus, LRR and SR have been widely used in image restoration [32], image classification [33], image fusion and denoising [6] and image super-resolution [34]. For image fusion, Li et al. [35] proposed multi-source image fusion based on SR. Subsequently, many other SR-based image fusion methods have been put forward and shown excellent performances [7,36–38,9,39].

In multi-scale transformation domain, Liu et al. [40] developed a SR-based method image fusion framework to merge the low-frequency information. To achieve a robust sparse representation, Li et al. [39] proposed a dictionary learning model based on group-sparse representation for multi-modality medical image fusion. For image fusion and denoising, Li et al. [8] separated the input image into coarse-scale components and the fine-scale components, and learned two discriminative dictionaries from a set of training images. During this process, the morphological priority of different components is used to regularize the decomposed results. However, it should be pointed out that these methods generally assume that the input source image has high resolution. This method can have satisfactory fusion results only when the assumed condition is met. When the resolution of the source image is low, the resolution of the fused results is low as well. To address this issue, image fusion and super-resolution methods are proposed in [28,31,29]. However, these methods

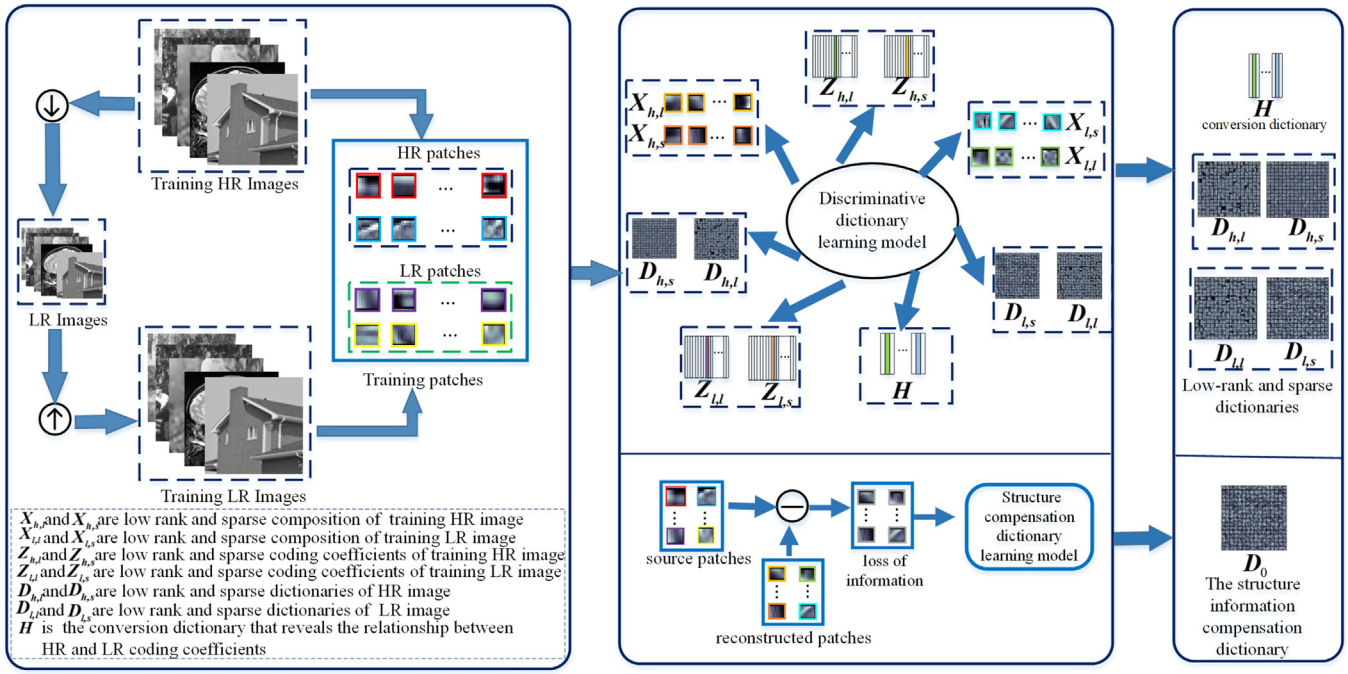

Fig. 1. Procedures of discriminative dictionary and structure compensation dictionary learning. HR: High resolution, LR: Low resolution.

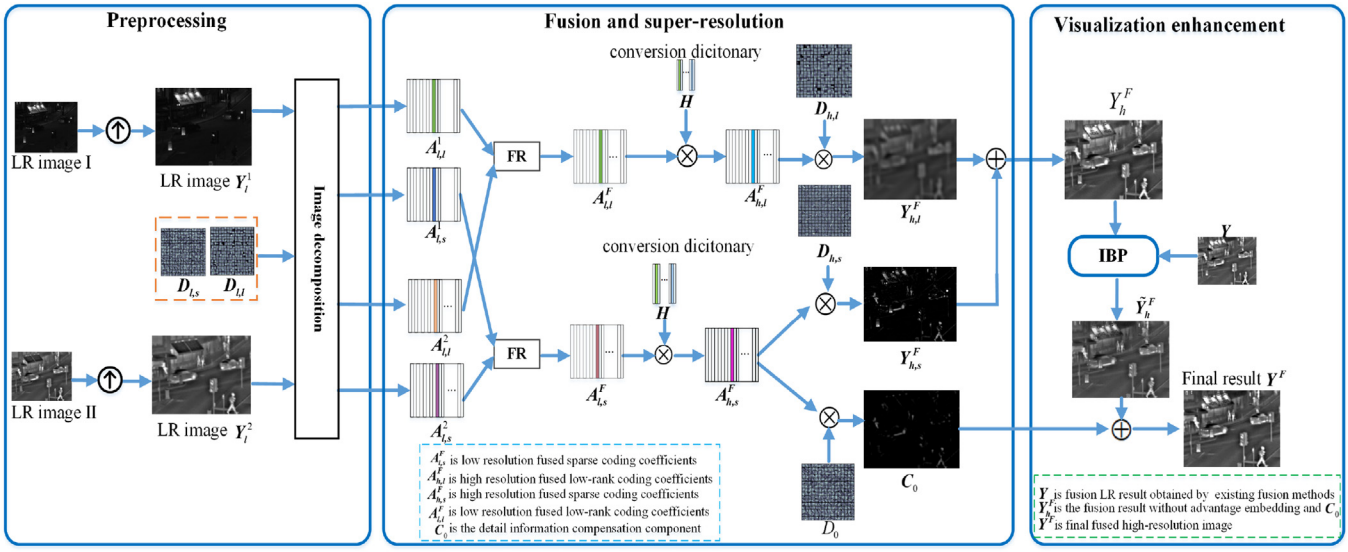

Fig. 2. Procedures of image fusion and super-resolution. FR: Fusion rule. IBP: Iterative back projection.

usually produce poor visual quality due to the defects of algorithm design.

2.2. Dictionary learning based image fusion

In recent years, dictionary learning has been applied to various practical cases, including image super-resolution [30,34,41], recognition [42,43], and classification [44,45], and shown impressive performance. In image fusion, dictionary learning has also received extensive attention, and a large number of effective algorithms have been proposed [7,6,9,36,38,46,47]. Particularly, Kim et al. [38] developed an efficient dictionary learning method based on patch joint clustering for multi-modal image fusion. Zhu et al. [7] proposed a local density peaks based clustering dictionary algorithm for multi-modality medical image fusion. Singh et al. [47] proposed to learn sparse K-SVD dictionary in nonsubsampled shearlet domain [48] to construct multi-modal medical image fusion model. Yin et al. [28] developed low-frequency and high-frequency component dictionary pairs learning for simultaneous image fusion and super-resolution. Subsequently, Iqbal [29] proposed a jointly learning method to constructer the low-resolution and high-resolution dictionary pairs for image fusion and super-resolution.

In the above mentioned methods, the whole image is represented by a fixed over-complete dictionary. However, different components of an image usually have different spatial morphological features, so it is difficult for one dictionary to describe all the components. This problem can be solved by using a larger size

dictionary, yet it will increase the computational complexity of the algorithm and reduce its efficiency. Aiming at this problem, Jiang et al. [49] proposed an image fusion method based on morphological component analysis, in which the DCT and Curvelet dictionaries are used respectively to express the texture and cartoon components of input images. However, the dictionaries analytically designed lack the adaptivity to image local structures. Although the existing researches have realized this defect and proposed to learn a compact dictionary from a set of training samples [9,10,36,47,50], such method fails to consider the resolution of the input images, and is unable to realize image fusion, super-resolution and visualization enhancement simultaneously. In this paper, we propose a novel semi-coupled discriminative dictionary learning and advantage embedding framework to address these problems.

2.3. Image super-resolution

Single image super-resolution (SISR) aims to construct the clean high-resolution from its low-resolution version. Because of its wide practical significance, it has attracted the attention of researchers. In deep learning based methods, convolutional neural network (CNN) is an important method for image super-resolution (SR). Specifically, Zhang et al. [51] proposed deep residual channel attention networks in the CNN framework, which can resist hindering the representational ability of CNNs caused by treating equally across different feature channels. Zhang et al. [52] designed a novel principled formulation and framework for single image super-resolution with arbitrary blur kernels, where the SISR degradation model is developed so as to utilize the advantage of existing blind deblurring approach for blur kernel estimation. To deploy the deep learning model to the mobile phone, Li et al. [53] presented a super lightweight SR network, and termed it as s-LWSR. Nazeri et al. [54] developed a "edge-informed" approach, where the SISR is formulated as an image inpainting task. The existing image super-resolution methods can improve the visual quality of images, but they cannot integrate complementary information from different sensors into one image. Although a large number of literatures on image fusion and SISR offer solutions separately for this issue, a unified framework is more viable and it has not been explored deeply. In this paper, we attempt to develop a novel simultaneous image fusion and super-resolution framework, and it can take advantage of existing image fusion methods for visualization enhancement of fusion result.

3. The proposed method

The proposed method is mainly composed of discriminative dictionary learning, low-rank sparse decomposition, image fusion and dominance embedding. In this section, we will first introduce the acquisition of discriminant dictionaries for different components.

3.1. Discriminative dictionary learning model

3.1.1. Discriminative representation dictionary learning

In multi-component analysis, dictionary learning of different components plays an important role in improving the performance of image decomposition. In this section, we develop a new dictionary learning method for decomposing the input image into low-rank and sparse components. To achieve fusion and super-resolution simultaneously, a pair of HR dictionaries and a pair of LR dictionaries are jointly learned. Let $\boldsymbol{D}_{h,l}=[\boldsymbol{d}^h_{l,1},\ldots,\boldsymbol{d}^h_{l,K}]\in\mathbb{R}^{M\times K}$ and $\boldsymbol{D}_{h,s}=[\boldsymbol{d}^h_{s,1},\ldots,\boldsymbol{d}^h_{s,K}]\in\mathbb{R}^{M\times K}$ be low-rank and sparse dictionaries used to represent the low-rank and sparse components of HR image respectively. $\boldsymbol{X}_h$ is the high-resolution training sample set and its low-resolution version is denoted by $\boldsymbol{X}_l$. Besides, the low-resolution dictionaries corresponding to $\boldsymbol{D}_{l,l}$ and $\boldsymbol{D}_{l,s}$ are denoted by $\boldsymbol{D}_{l,l}=[\boldsymbol{d}^l_{l,1},\ldots,\boldsymbol{d}^l_{l,K}]\in\mathbb{R}^{M\times K}$ and $\boldsymbol{D}_{l,s}=[\boldsymbol{d}^l_{s,1},\ldots,\boldsymbol{d}^l_{s,K}]\in\mathbb{R}^{M\times K}$, respectively.

In our method, the discriminative dictionary learning framework is formulated as:

$$\begin{aligned}\{\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s},\boldsymbol{D}_{h,l},\boldsymbol{D}_{h,s}\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H}\}=\arg\min_{\substack{\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s},\boldsymbol{D}_{h,l},\boldsymbol{D}_{h,s}\\ \boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H}}}\\ \left\{\sum_{i=l,h}\|\boldsymbol{X}_i-\boldsymbol{D}_{i,s}\boldsymbol{Z}_{i,s}-\boldsymbol{D}_{i,l}\boldsymbol{Z}_{i,l}\|_F^2+\Phi(\boldsymbol{D}_{l,l},\boldsymbol{D}_{h,l},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,s})+\Psi(\boldsymbol{H},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s})\right\}\\ s.t\|\boldsymbol{d}^l_{l,k}\|_2^2\leqslant\varepsilon_1,\|\boldsymbol{d}^h_{l,k}\|_2^2\leqslant\varepsilon_2,\|\boldsymbol{d}^l_{s,k}\|_2^2\leqslant\varepsilon_3,\|\boldsymbol{d}^h_{s,k}\|_2^2\leqslant\varepsilon_4,\forall k\end{aligned}\tag{1}$$

where $\varepsilon_1,\varepsilon_2,\varepsilon_3$ and ε_4 are constants that control the amplitude of each atom in different dictionaries. Usually, these parameters are set to 1; $\boldsymbol{Z}_{l,l}$ and $\boldsymbol{Z}_{l,s}$ are coding coefficients of low-rank and sparse components of $\boldsymbol{X}_l$, and $\boldsymbol{Z}_{h,l}$ and $\boldsymbol{Z}_{h,s}$ are coding coefficients of low-rank and sparse components of $\boldsymbol{X}_h$; $\boldsymbol{H}$ is the conversion dictionary that reveals the relationship between $\boldsymbol{Z}_{l,i}$ and $\boldsymbol{Z}_{h,i}(i=l,s)$; $\Psi(\boldsymbol{H},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s})$ and $\Phi(\boldsymbol{D}_{l,l},\boldsymbol{D}_{h,l},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,s})$ are discrimination promotion terms, which ensure the discrimination power of the learned dictionaries $\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s},\boldsymbol{D}_{h,l}$, and $\boldsymbol{D}_{h,s}$.

In image super-resolution, we can assume that the low-resolution image is generated by down-sampling a high-resolution image. Since the high-resolution image and its low-resolution version indicate the same scene, it is reasonable to assume that the coding coefficients of high-resolution and its corresponding low-resolution image patches can be converted to each other. Based on this assumption, some image super-resolution methods assume that the coding coefficients of the high-resolution and low-resolution image pairs are the same. However, such constraint not only limits the flexibility between high-resolution and low-resolution dictionary pairs, but also makes it difficult to determine the overlap size of the recovered high-resolution image patches.

In our method, we relax the restrictions of this assumption and propose a semi-coupled discriminative dictionary learning to construct the discriminative representation dictionaries. In this process, we employ a conversion dictionary $\boldsymbol{H}$ to reveal the relationship between the coding coefficients of high-resolution and low-resolution image patches. To facilitate the processing, the size of low resolution image is reset to the same as that of high resolution image via Bicubic interpolation. Then the relationship between the coding coefficients of low-resolution image and high-resolution image can be described as:

$$\begin{aligned}\Psi(\boldsymbol{H},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s})=\lambda_1\|\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}-\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2+\lambda_5\|\boldsymbol{H}\|_F^2\\ +\lambda_2\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}-\boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2,\end{aligned}\tag{2}$$

where λ_1,λ_2 and λ_5 are regularization parameters.

In Eq. (2), $\|\boldsymbol{H}\|_F^2$ is used to avoid over-fitting, $\|\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}-\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2$ is a relationship transformation term, and it is used to reveal the relationship between the coding coefficients of sparse components in low-resolution image and those of their corresponding high resolution image; and $\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}-\boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2$ is used to establish the relationship between $\boldsymbol{Z}_{h,l}$ and $\boldsymbol{Z}_{l,l}$. As we know, there is a strong linear correlation between low-rank components, and each element in the same coding coefficient vector has similar values. Based on this fact, we introduce an all 1 matrix $\boldsymbol{A}\in\mathbb{R}^{K\times K}$ and utilize regularization term $\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}-\boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2$ to promote the low-rankness of coding coefficients.

To promote the discrimination power of the learned dictionaries, we define:

$$\Phi(\boldsymbol{D}_{l,l},\boldsymbol{D}_{h,l},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,s}) = \lambda_3\left(\|\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}\|_* + \|\boldsymbol{D}_{l,l}\boldsymbol{Z}_{l,l}\|_*\right) + \lambda_4\left(\|\boldsymbol{Z}_{h,s}\|_1 + \|\boldsymbol{Z}_{l,s}\|_1\right), \tag{3}$$

where λ_3 and λ_4 are regularization parameters. $\|\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}\|_*$ and $\|\boldsymbol{D}_{l,l}\boldsymbol{Z}_{l,l}\|_*$ are used to ensure that $\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}$ and $\boldsymbol{D}_{l,l}\boldsymbol{Z}_{l,l}$ separated from the input image pairs are of low-rank. After the low-rank components are obtained, the sparse components can be also obtained by $\boldsymbol{X}_i - \boldsymbol{D}_{i,l}\boldsymbol{Z}_{i,l}$, where $\boldsymbol{X}_i$ denotes the input source data. Thus, the objective function for discriminative representation dictionary learning can be formulated as:

$$\begin{Bmatrix} \boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s} \\ \boldsymbol{D}_{h,l},\boldsymbol{D}_{h,s} \\ \boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s} \\ \boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H} \end{Bmatrix} = \arg \min_{\substack{\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s},\boldsymbol{D}_{h,l},\\ \boldsymbol{D}_{h,s},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s}\\ \boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H}}} \Bigg\{ \sum_{i=l,h}\|\boldsymbol{X}_i - \boldsymbol{D}_{i,s}\boldsymbol{Z}_{i,s} - \boldsymbol{D}_{i,l}\boldsymbol{Z}_{i,l}\|_F^2 + \lambda_5\|\boldsymbol{H}\|_F^2 + \lambda_1\|\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s} - \boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2 + \lambda_2\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l} - \boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2 + \lambda_3\left(\|\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}\|_* + \|\boldsymbol{D}_{l,l}\boldsymbol{Z}_{l,l}\|_*\right) + \lambda_4\left(\|\boldsymbol{Z}_{h,s}\|_1 + \|\boldsymbol{Z}_{l,s}\|_1\right)\Bigg\} \; s.t \|\boldsymbol{d}_{l,k}^l\|_2^2 \leqslant \varepsilon_1, \|\boldsymbol{d}_{l,k}^h\|_2^2 \leqslant \varepsilon_2, \|\boldsymbol{d}_{s,k}^l\|_2^2 \leqslant \varepsilon_3, \|\boldsymbol{d}_{s,k}^h\|_2^2 \leqslant \varepsilon_4, \; \forall k, \tag{4}$$

3.1.2. Structure compensation dictionary learning

Inevitably, there are some deviations between the reconstructed high resolution result $\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}$ + $\boldsymbol{D}_{h,l}\boldsymbol{H}\boldsymbol{Z}_{l,l}$ and the real high-resolution input $\boldsymbol{X}_h$, because our objective function in Eq. (4) cannot guarantee $\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s} + \boldsymbol{D}_{h,l}\boldsymbol{H}\boldsymbol{Z}_{l,l}$ and $\boldsymbol{X}_h$ are equal. Therefore, some structural details of the source image may be lost during the reconstruction. To alleviate this problem and enhance the visualization of reconstructed results, we develop a SICD learning model to compensate for the lost information. Considering the difference between reconstructed results and real results, we formulate the SICD learning model as:

$$\boldsymbol{D}_0 = \arg\min_{\boldsymbol{D}_0}\left\{\|\boldsymbol{X}_h - \boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s} - \boldsymbol{D}_{h,l}\boldsymbol{H}\boldsymbol{Z}_{l,l} - \boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2 + \lambda_0\|\boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_1\right\} s.t\|\boldsymbol{d}_{0,k}\|_2^2 \leqslant \varepsilon_0, \; \forall k, \tag{5}$$

where $\boldsymbol{d}_{0,k}$ denotes the k-th column of $\boldsymbol{D}$; ε_0 is a positive constant, and is set to 1; $\boldsymbol{D}_0$ denotes the SICD; λ_0 is a scalar constant; and $\boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}$ represents the structure information compensation components for reconstruction of HR image. Since the components we want to supplement are sparse, we force $\boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}$ to be sparse by minimizing $\|\boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_1$. As for the model in Eq. (5), we assume that the compensation components and the sparse components share the same coding coefficients, so as to make it easier to learn the structure dictionary $\boldsymbol{D}_0$.

3.2. Image decomposition

With the learned low-rank dictionary $\boldsymbol{D}_{l,l}$ and sparse dictionary $\boldsymbol{D}_{l,s}$, the input LR image patches $\boldsymbol{Y}_l$ can be decomposed into low-rank and sparse components. To this end, one of the most straightforward ways is to obtain the coefficients of different components of input $\boldsymbol{Y}_l$ by:

$$\{\boldsymbol{A}_{l,l},\boldsymbol{A}_{l,s}\} = \arg\min_{\boldsymbol{A}_{l,l},\boldsymbol{A}_{l,s}}\left\{\|\boldsymbol{Y}_l - \boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s} - \boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_F^2 + \beta_2\|\boldsymbol{A}_{l,l}\|_1 + \beta_1\|\boldsymbol{A}_{l,s}\|_1\right\}, \tag{6}$$

where $\boldsymbol{A}_{l,l}$ and $\boldsymbol{A}_{l,s}$ are the coefficients of low-rank and sparse component of $\boldsymbol{Y}_l$ over the corresponding dictionaries $\boldsymbol{D}_{l,l}$ and $\boldsymbol{D}_{l,s}$, respectively. Then the low-rank and sparse components can be generated by $\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}$ and $\boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}$.

In this way, the performance of image decomposition is completely determined by the discriminative ability of the learned dictionaries $\boldsymbol{D}_{l,l}$ and $\boldsymbol{D}_{l,s}$. To ensure the low-rank sparse components are completely separated from the input image, nuclear norm and l_1 norm regularization are developed and incorporated into this model. For sparse components, we propose to minimize $\|\boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}\|_1$ to ensure that the separated $\boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}$ is sparse; for low-rank components, we guarantee that $\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}$ are low-rank by minimizing $\|\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_*$. Therefore, the proposed image decomposition model can be formulated as:

$$\{\boldsymbol{A}_{l,l},\boldsymbol{A}_{l,s}\} = \arg\min_{\boldsymbol{A}_{l,l},\boldsymbol{A}_{l,s}}\left\{\|\boldsymbol{Y}_l - \boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s} - \boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_F^2 + \beta_2\|\boldsymbol{A}_{l,l}\|_1 + \beta_1\|\boldsymbol{A}_{l,s}\|_1 + \beta_3\|\boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}\|_1 + \beta_4\|\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_*\right\}, \tag{7}$$

Since the low-rank components possess similar spatial features, the coding coefficient should be linearly interrelated. Moreover, similar low-rank vectors in low-rank components should have the same representation coefficients, so the above requirements cannot be met by minimizing $\|\boldsymbol{A}_{l,l}\|_1$. Furthermore, as stated in [55], minimizing trace norm $\|\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_*$ cannot allow the rank of $\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}$ change along with $\|\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_*$. To this end, we propose the following objective function for low-rank and sparse decomposition:

$$\{\boldsymbol{A}_{l,l},\boldsymbol{A}_{l,s}\} = \arg\min_{\boldsymbol{A}_{l,l},\boldsymbol{A}_{l,s}}\left\{\|\boldsymbol{Y}_l - \boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s} - \boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_F^2 + \beta_2\|\boldsymbol{A}_{l,l}\|_{2,1} + \beta_1\|\boldsymbol{A}_{l,s}\|_1 + \beta_3\|\boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}\|_1 + \beta_4\|\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_*\right\}, \tag{8}$$

where $\|\boldsymbol{A}_{l,l}\|_{2,1} = \sum_j\sqrt{\sum_i \boldsymbol{A}_{l,l}(i,j)}$, $\boldsymbol{A}_{l,l}(i,j)$ is the (i,j)-th entry of $\boldsymbol{A}_{l,l}$. If the entries in each row in $A_{l,l}$ are taken as a group, then minimizing $\|\boldsymbol{A}_{l,l}\|_{2,1}$ will encourage the entries in each row of $\boldsymbol{A}_{l,l}$ to be the same. From inequality $\text{rank}(\boldsymbol{AB}) \leqslant \min\{\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B})\}$, it can be known that minimizing $\|\boldsymbol{A}_{l,l}\|_{2,1}$ can further guarantee the low-rankness of $\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}$, and avoid some defects caused by minimizing the trace norm $\|\boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_*$.

3.3. Optimization and algorithm

In this section, the ways for solving the objective functions in Eqs. (4), (5) and (8) are introduced. Obviously, these functions are non-convex jointly for $\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s},\boldsymbol{D}_{h,l},\boldsymbol{D}_{h,s},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H},\boldsymbol{D}_0$, $\boldsymbol{A}_{l,l}$ and $\boldsymbol{A}_{l,s}$. However, in the case that we solve one variable and keep others fixed, these sub-problems are convex. Therefore, these problems are solved via standard alternating iteration. In this way, the convergence of each subproblem with respect to a variable is guaranteed.

3.3.1. Optimization of discriminative dictionary pairs learning

To facilitate the optimization, we introduce four auxiliary variables $\boldsymbol{X}_{h,s},\boldsymbol{X}_{h,l},\boldsymbol{X}_{l,s}$ and $\boldsymbol{X}_{l,l}$, and relax the optimization problem in Eq. (4) into:

$$\left\{\begin{array}{c}\boldsymbol{X}_{l,l},\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s}\\ \boldsymbol{X}_{l,s}\boldsymbol{D}_{h,l},\boldsymbol{D}_{h,s}\\ \boldsymbol{X}_{h,l},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s}\\ \boldsymbol{X}_{h,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H}\end{array}\right\}=\arg\min_{\begin{array}{c}\boldsymbol{X}_{l,l},\boldsymbol{D}_{l,l},\boldsymbol{D}_{l,s}\\ \boldsymbol{X}_{l,s}\boldsymbol{D}_{h,l},\boldsymbol{D}_{h,s}\\ \boldsymbol{X}_{h,l},\boldsymbol{Z}_{l,l},\boldsymbol{Z}_{l,s}\\ \boldsymbol{X}_{h,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{h,s},\boldsymbol{H}\end{array}}$$

$$\begin{aligned}&\left\{\sum_{i=l,h}\|\boldsymbol{X}_i-\boldsymbol{X}_{i,s}-\boldsymbol{X}_{i,l}\|_F^2+\lambda_5\|\boldsymbol{H}\|_F^2+\lambda_1\|\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}-\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2\right.\\&+\lambda_2\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}-\boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2+\lambda_3\left(\|\boldsymbol{X}_{h,l}\|_*+\|\boldsymbol{X}_{l,l}\|_*\right)\\&+\lambda_4\left(\|\boldsymbol{Z}_{h,s}\|_1+\|\boldsymbol{Z}_{l,s}\|_1\right)+\|\boldsymbol{X}_{i,s}-\boldsymbol{D}_{i,s}\boldsymbol{Z}_{i,s}\|_F^2+\|\boldsymbol{X}_{i,l}\\&\left.-\boldsymbol{D}_{i,l}\boldsymbol{Z}_{i,l}\|_F^2\right\}\ s.t\|\boldsymbol{d}_{l,k}^l\|_2^2\\&\leqslant\varepsilon_1,\|\boldsymbol{d}_{l,k}^h\|_2^2\leqslant\varepsilon_2,\|\boldsymbol{d}_{s,k}^l\|_2^2\leqslant\varepsilon_3,\|\boldsymbol{d}_{s,k}^h\|_2^2\leqslant\varepsilon_4,\forall k,\end{aligned}\tag{9}$$

1) **Updating $\boldsymbol{X}_{h,s},\boldsymbol{X}_{h,l}$ and coding coefficients $\boldsymbol{Z}_{h,s}$, and $\boldsymbol{Z}_{h,l}$.** First, we update $\boldsymbol{X}_{h,s}$ by fixing all the other variables. Then the objective function Eq. (9) is reduced to:

$$\boldsymbol{X}_{h,s}=\arg\min_{\boldsymbol{X}_{h,s}}\{\|\boldsymbol{X}_h-\boldsymbol{X}_{h,s}-\boldsymbol{X}_{h,l}\|_F^2+\|\boldsymbol{X}_{h,s}-\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}\|_F^2\},\tag{10}$$

and it has the following closed-form solution:

$$\boldsymbol{X}_{h,s}=\frac{1}{2}(\boldsymbol{X}_h-X_{h,l}+\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}),\tag{11}$$

Similarly, we can obtain optimal $\boldsymbol{X}_{h,l}$ by solving

$$\begin{aligned}\boldsymbol{X}_{h,l}=\arg\min_{\boldsymbol{X}_{h,l}}\{&\|\boldsymbol{X}_h-\boldsymbol{X}_{h,s}-\boldsymbol{X}_{h,l}\|_F^2+\|\boldsymbol{X}_{h,l}-\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}\|_F^2\\&+\lambda_3\|\boldsymbol{X}_{h,l}\|_*\}.\end{aligned}\tag{12}$$

This problem can be easily solved through the singular value thresholding algorithm (SVT) [56].

With the updated $\boldsymbol{X}_{h,s}$ and $\boldsymbol{X}_{h,l}$, we can update the coding coefficients $\boldsymbol{Z}_{h,s}$ and $\boldsymbol{Z}_{h,l}$ by solving:

$$\boldsymbol{Z}_{h,s}=\arg\min_{\boldsymbol{Z}_{h,s}}\{\|\tilde{\boldsymbol{X}}_{h,s}-\tilde{\boldsymbol{D}}_{h,s}\boldsymbol{Z}_{h,s}\|_F^2+\lambda_4\|\boldsymbol{Z}_{h,s}\|_1\},\tag{13}$$

and

$$\boldsymbol{Z}_{h,l}=\arg\min_{\boldsymbol{Z}_{h,l}}\left\{\|\boldsymbol{X}_{h,l}-\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}\|_F^2+\lambda_2\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}-\boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2\right\},\tag{14}$$

where $\tilde{\boldsymbol{X}}_{h,s}=[\boldsymbol{X}_{h,s};\sqrt{\lambda_1}\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}]$, and $\tilde{\boldsymbol{D}}=[\boldsymbol{D}_{h,s};\sqrt{\lambda_1}\boldsymbol{D}_{h,s}]$. Obviously, Eq. (13) is convex and is a typical l_1-minimization problem, and we can solve it by using iterative shrinkage algorithm (ISA) [57], fast iterative shrinkage thresholding algorithm (FISTA) [58] or two-step iterative shrinkage/thresholding algorithms (TwIST) [59]. For Eq. (14), it has the following closed-form solution:

$$\boldsymbol{Z}_{h,l}=\left(\boldsymbol{D}_{h,l}^T\boldsymbol{D}_{h,l}+\frac{\lambda_2}{K^2}\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\left(\boldsymbol{D}_{h,l}^T\boldsymbol{X}_{h,l}+\frac{\lambda_2}{K}\boldsymbol{A}^T\boldsymbol{H}\boldsymbol{Z}_{l,l}\right),\tag{15}$$

2) **Updating $\boldsymbol{X}_{l,s},\boldsymbol{X}_{l,l}$ and coding coefficients $\boldsymbol{Z}_{l,s}$ and $\boldsymbol{Z}_{l,l}$.** Then, we optimize $\boldsymbol{X}_{l,s}$, and $\boldsymbol{X}_{l,l}$ while keep others fixed. As a result, $\boldsymbol{X}_{l,s}$ and $\boldsymbol{X}_{l,l}$ can be updated by:

$$\boldsymbol{X}_{l,s}=\frac{1}{2}\left(\boldsymbol{X}_l-\boldsymbol{X}_{l,l}+\boldsymbol{D}_{l,s}\boldsymbol{Z}_{l,s}\right).\tag{16}$$

and

$$\tilde{\boldsymbol{X}}_{l,l}=\arg\min_{\boldsymbol{X}_{l,l}}\left\{\left\|\tilde{\boldsymbol{X}}_l-\tilde{\boldsymbol{X}}_{l,l}\right\|_F^2+\lambda_3\left\|\tilde{\boldsymbol{X}}_{l,l}\right\|_*\right\},\tag{17}$$

where $\tilde{\boldsymbol{X}}_l=[X_l-X_{l,s};\boldsymbol{D}_{l,l}\boldsymbol{Z}_{l,l}]$, and $\tilde{\boldsymbol{X}}_{l,l}=[\boldsymbol{X}_{l,l};\boldsymbol{X}_{l,l}]$. This problem can be easily solved like Eq. (12). Next, we update $\boldsymbol{Z}_{l,s}$ and $\boldsymbol{Z}_{l,l}$ while keeping the other variables fixed, then we have:

$$\boldsymbol{Z}_{l,s}=\arg\min_{\boldsymbol{Z}_{l,s}}\left\{\left\|\tilde{\boldsymbol{X}}_{l,s}-\tilde{\boldsymbol{D}}_{l,s}\boldsymbol{Z}_{l,s}\right\|_F^2+\lambda_4\left\|\boldsymbol{Z}_{l,s}\right\|_1\right\},\tag{18}$$

and

$$\boldsymbol{Z}_{l,l}=\left(\boldsymbol{D}_{l,l}^T\boldsymbol{D}_{l,l}+\lambda_2\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\left(\boldsymbol{D}_{l,l}^T\boldsymbol{X}_{l,l}+\frac{\lambda_2}{K}\boldsymbol{H}^T\boldsymbol{A}\boldsymbol{Z}_{h,l}\right).\tag{19}$$

where $\tilde{\boldsymbol{X}}_{l,s}=[\boldsymbol{X}_{l,s};\sqrt{\lambda_1}\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}]$, and $\tilde{\boldsymbol{D}}_{l,s}=[\boldsymbol{D}_{l,s};\sqrt{\lambda_1}\boldsymbol{D}_{h,s}\boldsymbol{H}]$.Similarly, the problem in Eq. (18) can be solved through TwIST as well.

3) **Updating $\boldsymbol{H}$.** With dictionaries and coding coefficients fixed, we optimize $\boldsymbol{H}$ by solving:

$$\boldsymbol{H}=\arg\min_{\boldsymbol{H}}\left\{\lambda_1\left\|\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}-\boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}\right\|_F^2+\lambda_2\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}-\boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2+\lambda_5\|\boldsymbol{H}\|_F^2\right\}.\tag{20}$$

All the terms in Eq. (20) are characterized by Frobenius norm, so we can obtain

$$\begin{aligned}&\lambda_1(\boldsymbol{D}_{h,s}^T\boldsymbol{D}_{h,s})\boldsymbol{H}+\boldsymbol{H}(\lambda_2(\boldsymbol{Z}_{l,l}\boldsymbol{Z}_{l,l}^T)(\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T)^{-1}+\lambda_5(\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T)^{-1})\\&=\lambda_1\boldsymbol{D}_{h,s}^T\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}\boldsymbol{Z}_{l,s}^T(\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T)^{-1}+\frac{\lambda_2}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l}\boldsymbol{Z}_{l,l}^T(\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T)^{-1}.\end{aligned}\tag{21}$$

This is a standard Sylvester equation, which can be easily solved by MATLAB Sylvester function.

4) **Updating discriminative $\boldsymbol{D}_{h,s},\boldsymbol{D}_{h,l},\boldsymbol{D}_{l,s}$ and $\boldsymbol{D}_{l,l}$.** Keeping the other variables fixed, we can use the Lagrange dual approach [60] to obtain the analytical solutions of $\boldsymbol{D}_{h,s}$ and $\boldsymbol{D}_{h,l}$:

$$\boldsymbol{D}_{h,s}=\boldsymbol{X}_{h,s}\boldsymbol{Z}_{h,s}^T((\lambda_1+1)\boldsymbol{Z}_{h,s}\boldsymbol{Z}_{h,s}^T-\lambda_1(\boldsymbol{Z}_{h,s}\boldsymbol{Z}_{l,s}^T\boldsymbol{H}^T+\boldsymbol{H}\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{h,s}^T-\boldsymbol{H}\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T\boldsymbol{H}^T)+\Lambda_1)^{-1},\tag{22}$$

and

$$\boldsymbol{D}_{h,l}=(\boldsymbol{X}_{h,l}\boldsymbol{Z}_{h,l}^T)(\boldsymbol{Z}_{h,l}\boldsymbol{Z}_{h,l}^T+\Lambda_2)^{-1}.\tag{23}$$

Similarly, we can obtain the closed-form solutions to $\boldsymbol{D}_{l,s}$ and $\boldsymbol{D}_{l,l}$:

$$\boldsymbol{D}_{l,s}=(\boldsymbol{X}_{l,s}\boldsymbol{Z}_{l,s}^T)(\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T+\Lambda_3)^{-1},\boldsymbol{D}_{l,l}=(\boldsymbol{X}_{l,l}\boldsymbol{Z}_{l,l}^T)(\boldsymbol{Z}_{l,l}\boldsymbol{Z}_{l,l}^T+\Lambda_4)^{-1},\tag{24}$$

In the above four equations, $\Lambda_1,\Lambda_2,\Lambda_3$ and Λ_4 are diagonal matrices constructed from their respective optimal dual variables.

5) **Algorithm of discriminative dictionary pairs learning.** For ease of understanding, the details of the optimization procedures described above can be summarized in Algorithm 1.

Algorithm 1 Algorithm for discriminative dictionary pair learning

Input : Initial variables $\boldsymbol{D}_{h,s},\boldsymbol{D}_{h,l},\boldsymbol{D}_{l,s},\boldsymbol{D}_{l,l},\boldsymbol{Z}_{h,s},\boldsymbol{Z}_{h,l},\boldsymbol{Z}_{l,s},\boldsymbol{Z}_{l,l},\boldsymbol{H}$, parameters $\lambda_i(i=1,2\ldots,5)$, maximum number of iterations $\mathbb{K}$.

while $\mathbb{K}$ not reached **do**

(a) Fix all other variables, update $\boldsymbol{X}_{h,s}$ and $\boldsymbol{X}_{h,l}$ by solving Eqs. (11) and (12).

(b) Fix other variables, update $\boldsymbol{Z}_{h,s}$ and $\boldsymbol{Z}_{h,l}$ by solving Eqs. (13) and (15).

(c) Fix other variables, update $\boldsymbol{X}_{l,s}$ and $\boldsymbol{X}_{l,l}$ by solving Eqs. (16) and (17).

(d) Fix other variables, update $\boldsymbol{Z}_{l,s}$ and $\boldsymbol{Z}_{l,l}$ by solving Eqs. (18) and (19).

(e) Fix other variables, update $\boldsymbol{H}$ via Eq. (21).

(*continued on next page*)

(f) Fix other variables, update $\boldsymbol{D}_{h,s}, \boldsymbol{D}_{h,l}, \boldsymbol{D}_{l,s}$ and $\boldsymbol{D}_{l,l}$ respectively by solving Eqs. (22)–(24).

end while if the maximum number of iterations has been reached.

Output: $\boldsymbol{D}_{h,s}, \boldsymbol{D}_{h,l}, \boldsymbol{D}_{l,s}, \boldsymbol{D}_{l,l}$ and $\boldsymbol{H}$

3.3.2. Optimization of SICD

To obtain the optimal $\boldsymbol{D}_0$, we introduce a relaxation variable $\boldsymbol{X}_0$, and Eq. (5) can be rewritten as

$$\begin{aligned}\boldsymbol{D}_0 = \arg\min_{\boldsymbol{D}_0}\{&\|\boldsymbol{X}_h - \boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s} - \boldsymbol{D}_{h,l}\boldsymbol{H}\boldsymbol{Z}_{l,l} - \boldsymbol{X}_0\|_F^2 + \|\boldsymbol{X}_0 \\ &- \boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2 + \lambda_0\|\boldsymbol{X}_0\|_1\} s.t \|\boldsymbol{d}_{0,k}\|_2^2 \\ &\leqslant \varepsilon_0, \forall k. \end{aligned} \tag{25}$$

With $\boldsymbol{D}_0$ fixed, $\boldsymbol{X}_0$ can be updated by solving

$$\tilde{\boldsymbol{X}}_0 = \arg\min_{\tilde{\boldsymbol{X}}_0}\{\|\tilde{\boldsymbol{X}} - \tilde{\boldsymbol{X}}_0\|_F^2 + \lambda_0\|\boldsymbol{X}_0\|_1\}, \tag{26}$$

where $\tilde{\boldsymbol{X}} = [\boldsymbol{X}_h - \boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s} - \boldsymbol{D}_{h,l}\boldsymbol{H}\boldsymbol{Z}_{l,l}; \boldsymbol{D}_0\boldsymbol{H}\boldsymbol{Z}_{l,s}]$, and $\tilde{\boldsymbol{X}}_0 = [\boldsymbol{X}_0; \boldsymbol{X}_0]$. Eq. (30)is a typical l_1 minimization problem, which can be solved by using TwIST. When $\boldsymbol{X}_0$ is updated, $\boldsymbol{D}_0$ can be updated by

$$\boldsymbol{D}_0 = (\boldsymbol{X}_0\boldsymbol{Z}_{l,s}^T\boldsymbol{H}^T)(\boldsymbol{H}\boldsymbol{Z}_{l,s}\boldsymbol{Z}_{l,s}^T\boldsymbol{H}^T + \Lambda_0)^{-1}, \tag{27}$$

where Λ_0 is a diagonal matrix constructed from the dual variables.

3.3.3. Optimization of image decomposition

To update $\boldsymbol{A}_{l,s}$ and $\boldsymbol{A}_{l,l}$, the alternative update strategy is utilized again to solve the minimization problem in Eq. (8). To facilitate the optimization, we introduce $\boldsymbol{Y}_{l,s}$ and $\boldsymbol{Y}_{l,l}$ as auxiliary variables, and then Eq. (8) can be converted into:

$$\begin{aligned}\begin{Bmatrix}\boldsymbol{A}_{l,s}, \boldsymbol{A}_{l,l} \\ \boldsymbol{Y}_{l,s}, \boldsymbol{Y}_{l,l}\end{Bmatrix} = \arg\min_{\substack{\boldsymbol{A}_{l,s}, \boldsymbol{A}_{l,l} \\ \boldsymbol{Y}_{l,s}, \boldsymbol{Y}_{l,l}}} &\{\|\boldsymbol{Y} - \boldsymbol{Y}_{l,s} - \boldsymbol{Y}_{l,l}\|_F^2 + \|\boldsymbol{Y}_{l,s} - \boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}\|_F^2 \\ &+ \|\boldsymbol{Y}_{l,l} - \boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_F^2 + \beta_1\|\boldsymbol{A}_{l,s}\|_1 + \beta_2\|\boldsymbol{A}_{l,l}\|_{2,1} \\ &+ \beta_3\|\boldsymbol{Y}_{l,s}\|_1 + \beta_4\|\boldsymbol{Y}_{l,l}\|_*\},\end{aligned} \tag{28}$$

1) **Updating $\boldsymbol{Y}_{l,s}$ and $\boldsymbol{Y}_{l,l}$.** With $\boldsymbol{Y}_{l,l}, \boldsymbol{A}_{l,s}$ and $\boldsymbol{A}_{l,l}$ fixed, we update $\boldsymbol{Y}_{l,s}$ via solving:

$$\begin{aligned}\boldsymbol{Y}_{l,s} = \arg\min_{\boldsymbol{Y}_{l,s}}\{&\|\boldsymbol{Y} - \boldsymbol{Y}_{l,s} - \boldsymbol{Y}_{l,l}\|_F^2 + \|\boldsymbol{Y}_{l,s} - \boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}\|_F^2 \\ &+ \beta_3\|\boldsymbol{Y}_{l,s}\|_1\}.\end{aligned} \tag{29}$$

This problem can be solved by TwIST. Assuming that $\boldsymbol{Y}_{l,s}, \boldsymbol{A}_{l,s}$ and $\boldsymbol{A}_{l,l}$ are fixed, the optimal $\boldsymbol{Y}_{l,l}$ can be obtained by solving

$$\tilde{\boldsymbol{Y}}_{l,l} = \arg\min_{\tilde{\boldsymbol{Y}}_{l,l}}\left\{\|\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{Y}}_{l,l}\|_F^2 + \beta_4\|\tilde{\boldsymbol{Y}}_{l,l}\|_*\right\}. \tag{30}$$

where $\tilde{\boldsymbol{Y}} = [\boldsymbol{Y} - \boldsymbol{Y}_{l,s}; \boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}]$, and $\tilde{\boldsymbol{Y}}_{l,l} = [\boldsymbol{Y}_{l,l}; \boldsymbol{Y}_{l,l}]$. This problem can be effectively solved by SVT. Then we can obtain optimal $\boldsymbol{Y}_{l,l}$ from $\tilde{\boldsymbol{Y}}_{l,l}$.

2)**Updating $\boldsymbol{A}_{l,s}$ and $\boldsymbol{A}_{l,l}$.** Similarly, the optimal $\boldsymbol{A}_{l,s}$ and $\boldsymbol{A}_{l,l}$ can be obtained by solving the following minimization problems alternately:

$$\boldsymbol{A}_{l,s} = \arg\min_{\boldsymbol{A}_{l,s}}\left\{\|\boldsymbol{Y}_{l,s} - \boldsymbol{D}_{l,s}\boldsymbol{A}_{l,s}\|_F^2 + \beta_1\|\boldsymbol{A}_{l,s}\|_1\right\}, \tag{31}$$

and

$$\boldsymbol{A}_{l,l} = \arg\min_{\boldsymbol{A}_{l,l}}\left\{\|\boldsymbol{Y}_{l,l} - \boldsymbol{D}_{l,l}\boldsymbol{A}_{l,l}\|_F^2 + \beta_2\|\boldsymbol{A}_{l,l}\|_{2,1}\right\}, \tag{32}$$

Problem (31) is l_1-minimization problem, which can be easily solved by TwIST. For Eq. (32), the objective function is characterized by Frobenius norm and $l_{2,1}$-norm, and it can be solved by the method proposed in [61]. Finally, we summarize the above low-resolution image decomposition in Algorithm 2.

Algorithm 2 Algorithm for image decomposition

Input : Learned dictionaries $\boldsymbol{D}_{l,s}, \boldsymbol{D}_{l,l}, \boldsymbol{D}_0, \boldsymbol{H}$, initial $\boldsymbol{Y}_{l,s}, \boldsymbol{Y}_{l,l}, \boldsymbol{A}_{l,s}, \boldsymbol{A}_{l,l}$, parameters $\beta_i (i = 1, 2 \ldots, 4)$, maximum number of iterations $\mathbb{M}$.

while $\mathbb{M}$ not reached **do**

(a) Fix all other variables, update $\boldsymbol{Y}_{l,s}$ via Eq. (29).

(b) Fix other variables, update $\boldsymbol{Y}_{l,l}$ via Eq. (30).

(c) Fix other variables, update $\boldsymbol{A}_{l,s}$ and $\boldsymbol{A}_{l,l}$ respectively via Eqs. (31) and (32).

end while if the maximum number of iterations has been reached.

Output: $\boldsymbol{A}_{l,s}, \boldsymbol{A}_{l,l}$.

Fig. 3 shows the effects of our learned dictionaries and image decomposition algorithm. Fig. 3(a) and (b) are the low-rank and sparse components constructed from $\boldsymbol{D}_{h,l}\boldsymbol{Z}_{h,l}$ and $\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s}$, where $\boldsymbol{Z}_{h,l}$ and $\boldsymbol{Z}_{h,s}$ are obtained by solving $\min_z\|\boldsymbol{Z}\|_0, s.t.\|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{Z}\|_F^2 < 0.01$ via OMP algorithm [62], where $\boldsymbol{Z} = [\boldsymbol{Z}_{h,l}; \boldsymbol{Z}_{h,s}]$, and $\boldsymbol{D} = [\boldsymbol{D}_{h,l}, \boldsymbol{D}_{h,s}]$. Fig. 3(d) and (e) are the low-rank and sparse components separated by Algorithm 2. As can be known from these results, $\boldsymbol{D}_{h,l}$ and $\boldsymbol{D}_{h,s}$ are endowed with strong discriminative ability by our dictionary learning Algorithm 1, and most of the low-rank and sparse components can be separated from the input image. But from the magnified area in Fig. 3(a) and (b), we can see that some of the information is not separated from the source image. The separated results in Fig. 3(d) and (e) demonstrates that the developed image decomposition algorithm can solve this problem. Under the same reconstruction error (0.01), PSNR of the reconstruction result (see Fig. 3(f)) with coding coefficients generated by Algorithm 2 reaches 68.8462, and it is 5.17 higher than that of the result (see Fig. 3(c)) whose coding coefficients are generated by solving $\min_z\|\boldsymbol{Z}\|_0, s.t.\|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{Z}\|_F^2 < 0.01$.

3.4. Fusion and advantage embedding scheme

3.4.1. Simultaneous fusion and super-resolution

With the updated $\boldsymbol{A}_{l,s}, \boldsymbol{A}_{l,l}$ and $\boldsymbol{H}$, we employ the popular "max-absolute" rule to construct the coding coefficient of high-resolution fused image. Let $\boldsymbol{A}_{l,s}^j = [\boldsymbol{a}_{l,s,1}^j, \ldots, \boldsymbol{a}_{l,s,L}^j]$, and $\boldsymbol{A}_{l,l}^j = [\boldsymbol{a}_{l,l,1}^j, \ldots, \boldsymbol{a}_{l,l,L}^j]$. Here, $\boldsymbol{a}_{l,s,k}^j$ and $\boldsymbol{a}_{l,l,k}^j (k = 1, 2 \ldots, L)$ are the k-th columns of $\boldsymbol{A}_{l,s}^j$ and $\boldsymbol{A}_{l,l}^j$, respectively; and j denotes the j-th low-resolution image; and L denotes the number of patches of each input image. The fused HR sparse coefficients $\boldsymbol{A}_{h,l}^F$ and $\boldsymbol{A}_{h,s}^F$ of low-rank and sparse components can be constructed by the following scheme:

$$\begin{aligned}\boldsymbol{a}_{h,l,k}^F &= \boldsymbol{H}\boldsymbol{a}_{l,l,k}^{j^*}, \quad j^* \\ &= \arg\min_{j\in(1,2,\ldots,M)}\left\{\|\boldsymbol{a}_{l,l,k}^1\|_1, \|\boldsymbol{a}_{l,l,k}^2\|_1, \ldots, \|\boldsymbol{a}_{l,l,k}^j\|_1, \ldots, \|\boldsymbol{a}_{l,l,k}^M\|_1\right\}\boldsymbol{a}_{h,s,k}^F \\ &= \boldsymbol{H}\boldsymbol{a}_{l,s,k}^{j^*}, \quad j^* \\ &= \arg\min_{j\in(1,2,\ldots,M)}\left\{\|\boldsymbol{a}_{l,s,k}^1\|_1, \|\boldsymbol{a}_{l,s,k}^2\|_1, \ldots, \|\boldsymbol{a}_{l,s,k}^j\|_1, \ldots, \|\boldsymbol{a}_{l,s,k}^M\|_1\right\},\end{aligned} \tag{33}$$

Fig. 3. The effects of our learned dictionaries and image decomposition algorithm. (a) and (b) Low-rank and sparse components of "barbara" image generated only by the learned dictionaries; (c) the result of reconstruction by adding (a) and (b) (with PSNR = 68.8462); (d) and (e) low-rank and sparse components of "barbara" image generated by Algorithm 2; (f) the result of reconstruction by adding (d) and (e) (with PSNR = 74.0158).

where M is the number of input images. Let $\boldsymbol{A}_{l,l}^F = [\boldsymbol{a}_{h,l,1}^F, \ldots, \boldsymbol{a}_{h,l,L}^F]$, and $\boldsymbol{A}_{l,s}^F = [\boldsymbol{a}_{h,s,1}^F, \ldots, \boldsymbol{a}_{h,s,L}^F]$, then we can obtain the fused image patch with high-resolution by:

$$\boldsymbol{Y}_h^F = \boldsymbol{Y}_{h,s}^F + \boldsymbol{Y}_{h,l}^F. \tag{34}$$

where $\boldsymbol{Y}_{h,s}^F = \mathrm{Rec}(\boldsymbol{D}_{h,s}\boldsymbol{A}_{h,s}^F), \boldsymbol{Y}_{h,l}^F = \mathrm{Rec}(\boldsymbol{D}_{h,l}\boldsymbol{A}_{h,l}^F)$, and Rec is a reconstruct operation used to convert image patches into one image.

3.4.2. Advantage embedding scheme

In order to embed the advantages of existing fusion methods in preserving the visual effect into our results, we propose to utilize a deconvolution operation to compensate the advantages from one image to another. Let $\boldsymbol{X}$ be an ideal high-resolution fused image, the produced result by an existing fusion method with excellent performance is represented by $\boldsymbol{Y}$. Assuming that $\boldsymbol{Y}$ is obtained by downsampling $\boldsymbol{X}$, i.e. $\boldsymbol{Y} = \downarrow \boldsymbol{X}$. To endow $\boldsymbol{Y}_h^F$ with the advantage of $\boldsymbol{Y}$, inspired by [41], we propose to utilize the following objective function to refine $\boldsymbol{Y}_h^F$:

$$\boldsymbol{Y}^* = \arg\min_{\boldsymbol{Y}_h^F} \|\boldsymbol{Y} - \downarrow \boldsymbol{Y}_h^F\|_F^2 s.t. \downarrow \boldsymbol{X} = \boldsymbol{Y}, \tag{35}$$

This minimization problem can be solved by the method mentioned in Refs. [63,64], and the refining process is formulated as

$$\boldsymbol{Y}_{h,t+1}^F = \boldsymbol{Y}_{h,t}^F + ((\downarrow \boldsymbol{Y}_h^F - \boldsymbol{Y}) \uparrow) \otimes \boldsymbol{P}, \tag{36}$$

where $\uparrow$ is an upsampling operator, $\boldsymbol{P}$ is a back projection filter, $\otimes$ is a convolution operation, and t denotes the t-th iteration. The total number of iterations is the same as that in [63,64].

In order to highlight the edge details of the reconstructed image, we use an information compensation dictionary to enhance visualization and prevent the detail information from losing. In this process, we assume that the detail components that need to be supplemented share the same coding coefficients with the sparse components of the fused high-resolution image in their respective dictionaries. So the final fused high-resolution image $\boldsymbol{Y}^F$ can be formulated as:

$$\boldsymbol{Y}^F = \tilde{\boldsymbol{Y}}_h^F + \mathrm{Rec}(\boldsymbol{D}_0\boldsymbol{H}\boldsymbol{A}_{l,s}^F). \tag{37}$$

where $\mathrm{Rec}(\boldsymbol{D}_0\boldsymbol{H}\boldsymbol{A}_{l,s}^F)$ denotes the structure compensation information $\boldsymbol{C}_0$, and $\tilde{\boldsymbol{Y}}_h^F$ is the final refined result via Eq. (36).

4. Experiments and analysis

4.1. Experiment settings

4.1.1. Training and test images

In the section of dictionaries learning, we use the same 8 training samples as Ref. [8] to train high-resolution discriminative dictionary pairs, and the corresponding LR training images are generated by downsampling and then Bicubic upsampling back to the same size with HR images. For test images, we utilize two pairs of HR infrared and visible images (as shown in the first row of Fig. 4), two pairs of HR medical images (as shown in the second row of Fig. 4), and two pairs of HR multi-focus images(as shown in the last row of Fig. 4). All the source images have been registered, and can be obtained fromhttp://www.med.harvard.edu/AANLIB/home.html andhttp://www.imagefusion.org/. The low-resolution versions of these source images are shown in Fig. 5. All the fusion experiments in the present paper are finished in MATLAB 2016b on an i7-7500U 2.9 GHz machine with 16 GB RAM.

4.1.2. Baseline methods

To verify the superiority of our method, we compare it with several state-of-the-art image fusion and super-resolution methods. First, Li's method presented in Ref. [31] is taken as one of the competing methods since it can achieve image fusion and super-resolution simultaneously. However, as such kind of methods are quite limited, we perform super-resolution on the fusion results

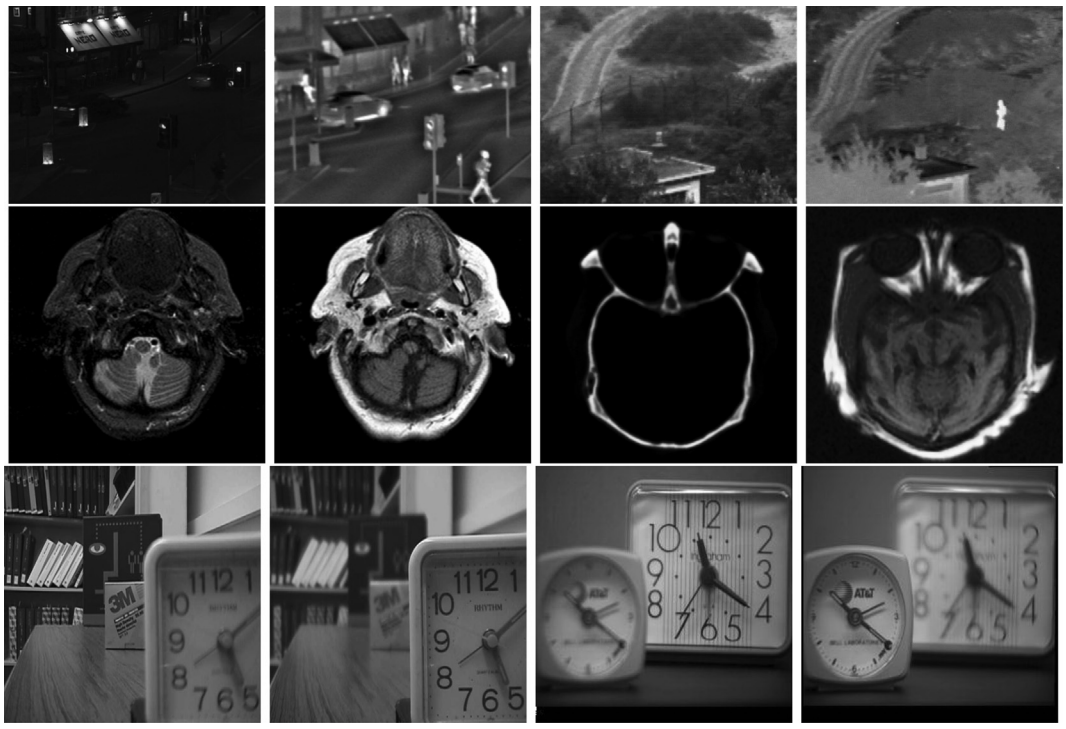

Fig. 4. The HR source images. From top to bottom: infrared and visible image pairs with size of 240 × 320, medical image pairs with size of 256 × 256, multi-focus image pairs with sizes of 240 × 320 and 256 × 256.

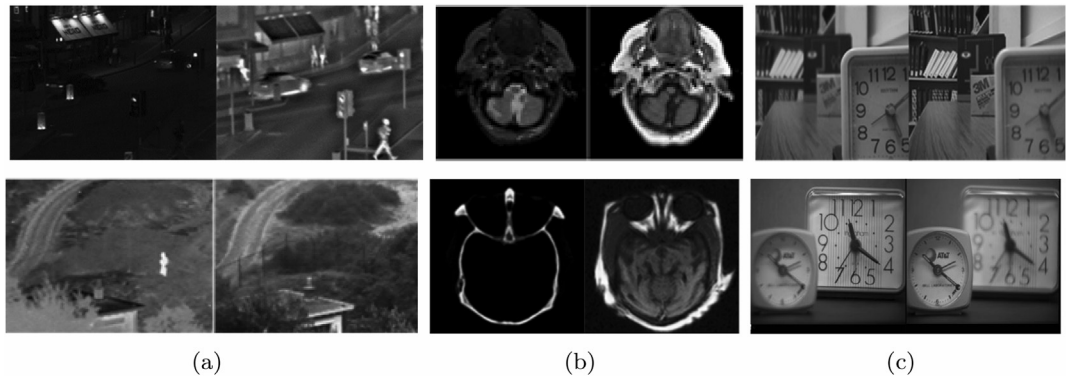

Fig. 5. The LR source images. (a) Infrared and visible image pairs with size of 120 × 160; (b) medical image pairs with size of 128 × 128; (c) multi-focus image pairs with sizes of 120 × 160 and 128 × 128.

generated by some excellent algorithms like Zhu's method [7], CNN-based method [24,12], GF-based method [65], CT-based method [9], and NSCT-based method [5]; and the low-resolution fusion images of these methods are shown in Fig. 6. In this paper, we compare the super-resolution results of these methods with those obtained by ours. For super-resolution, SLWSR [53] and the sparse representation based super-resolution (SRSR) [41] are used to construct the high-resolution results for Fig. 5. In order to incorporate the advantages of them in improving the visual effect of fusion results into the results of our fusion and super-resolution method, the fused results of the above four algorithms (see Fig. 6) are used as ideal low-resolution image to refine our fusion and super-resolution results.

4.1.3. Objective evaluation metrics

In this paper, four objective evaluation metrics are used to assess the perceptual quality of fusion and super-resolution results. These metrics are spatial frequency (SF) based metric Q_{SF} [66,67], the quality-aware clustering method Q_{AC} [68], the commonly used entropy (Q_{ENT}), and image gradient (Q_{GD}). Q_{SF} uses the ratio of SF error to measure the perceptual quality of the fused result. Specifically, if the value of Q_{SF} is below zero, it indicates that the information of the source image is lost during fusion, whereas if the value of Q_{SF} is greater than zero, and meanwhile no any artificial information and noise are introduced, it indicates that the details of the source image get enhanced. For the above metrics, the higher the objective evaluation score, the better the fusion quality.

4.2. Fusion and super-resolution of infrared and visible images

In the first experiment, we utilize two pairs of low-resolution infrared and visible images shown in Fig. 5 as the test images. As can be seen from these images, the warm objects, such as pedestrians, vehicles, and lights, are clearly visible against cool backgrounds in infrared images. While, only

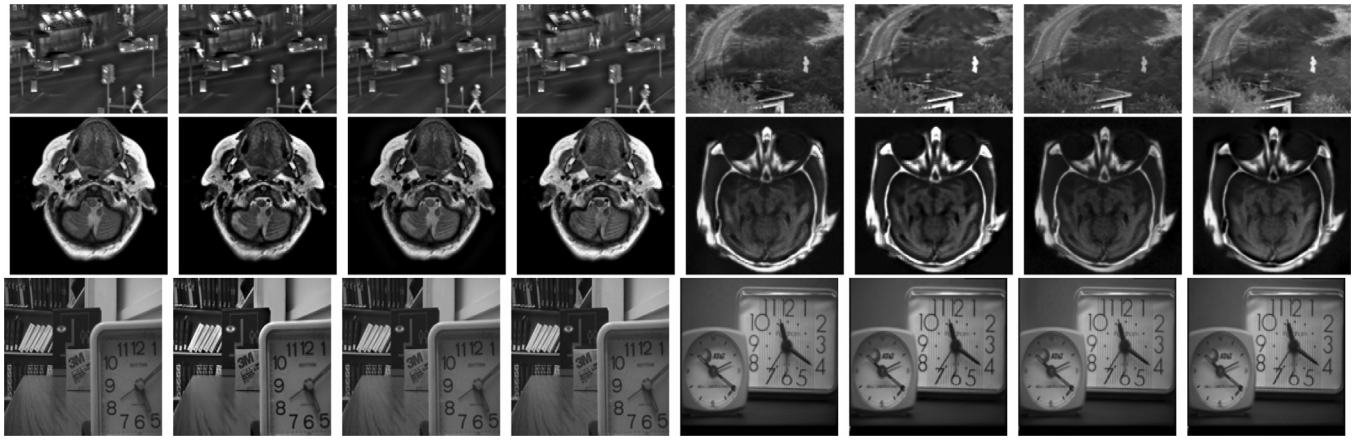

Fig. 6. Results of low-resolution image fusion with some representative methods. The images from left to right: the fusion results of CNN, CT, GF and NSCT respectively. the sizes of fusion results for all types of images are $120 \times 160, 120 \times 160$, $128 \times 128, 120 \times 160$, and 128×128, respectively.

the visible details of background, such as trees, lighted billboards, and fences, are clearly visible notable. The main purpose of our method is to integrate the complementary information of source images into the fused results, and improve its resolution at the same time. The fused results of different methods are shown in Figs. 7 and 8, in which "Ours-Nec" is our method but without advantage embedding and information compensation operations.

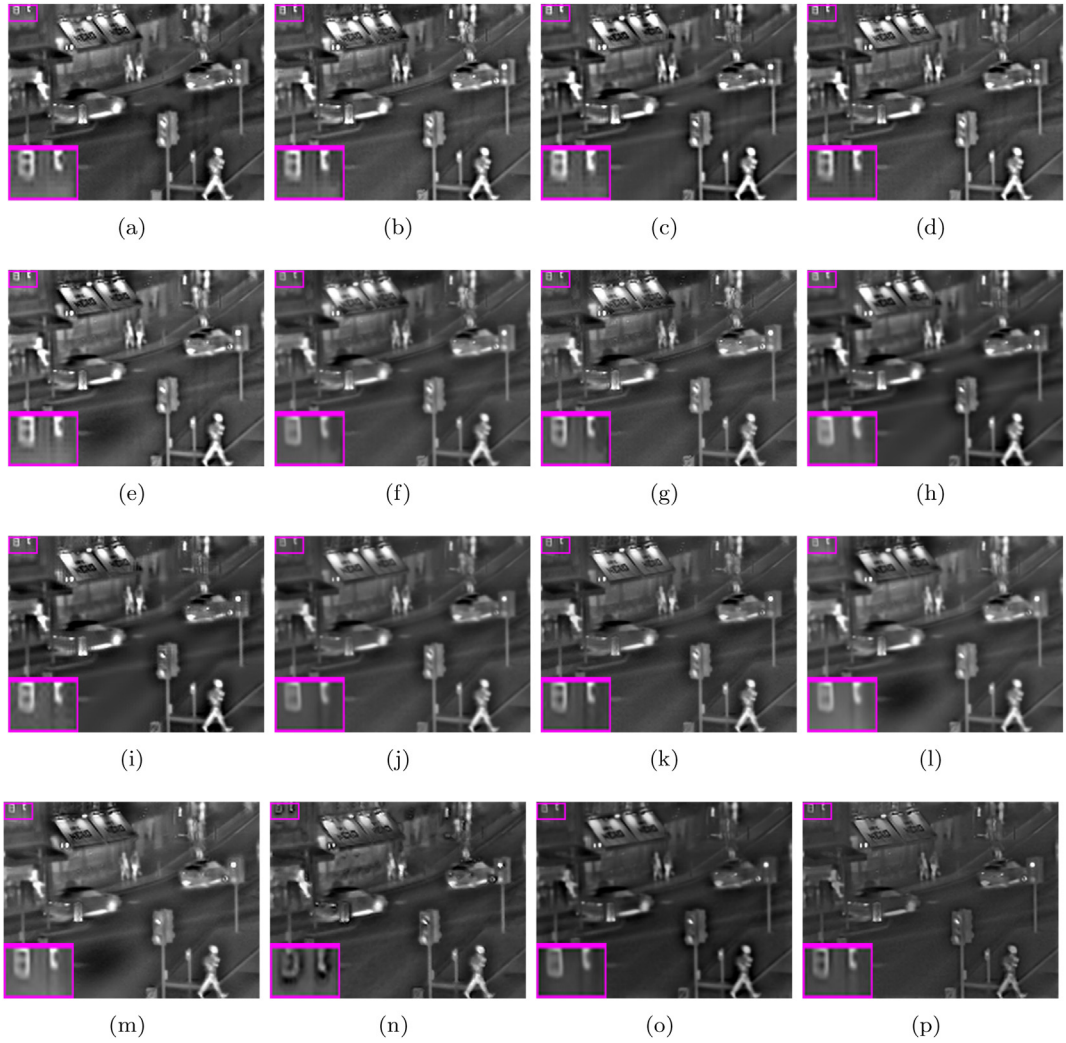

Fig. 7. Fusion and high-resolution results of the first infrared and visible image pair (each result with size of 240×320). (a) to (p) respectively represent: the fused and reconstructed images by Ours-Nec, Ours-CNN, Ours-CT, Ours-GF, Ours-NSCT, CNN(SLWSR), CNN(SRSR), CT(SLWSR), CT(SRSR), GF(SLWSR), GF(SRSR), NSCT(SLWSR), NSCT (SRSR), Li's, Zhu's(SLWSR), Zhu's(SRSR).

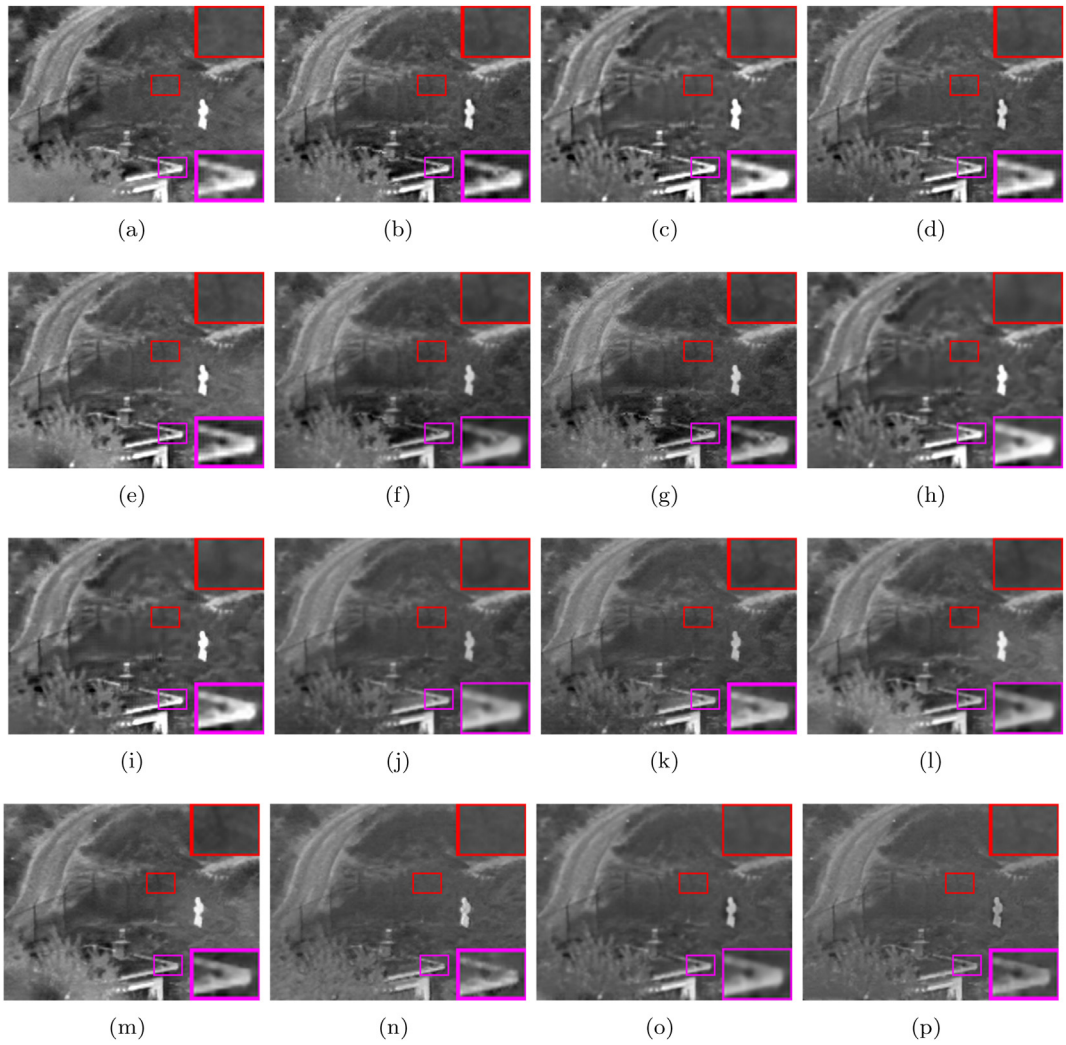

Fig. 8. Fusion and high-resolution results of the second infrared and visible image pair (each result with size of 240 × 320). (a) to (p) respectively represent: the fused and reconstructed images by Ours-Nec, Ours-CNN, Ours-CT, Ours-GF, Ours-NSCT, CNN(SLWSR), CNN(SRSR), CT(SLWSR), CT(SRSR), GF(SLWSR), GF(SRSR), NSCT(SLWSR), NSCT (SRSR), Li's, Zhu's(SLWSR), Zhu's(SRSR).

Table 1
Quantitative assessment of different fusion methods for fusion and super-resolution of the first infrared and visible image pair.

Methods	Q_{SF}	Q_{QAC}	Q_{ENT}	Q_{GD}
Ours-Nec	**0.0087**	**0.6952**	**6.9373**	**6.9904**
Ours-CNN	**0.0570**	**0.6922**	**6.7965**	**7.5185**
CNN(SLWSR)	–0.2629	0.6162	6.6782	5.0108
CNN(SRSR)	–0.1444	0.6713	6.6940	5.7411
Ours-CT	**0.0843**	**0.6919**	**6.8147**	**7.4449**
CT(SLWSR)	–0.1998	0.5449	6.7291	5.1674
CT(SRSR)	–0.0972	0.6232	6.7371	5.7219
Our-GF	**0.0070**	**0.6997**	**6.7275**	**7.0908**
GF(SLWSR)	–0.3288	0.5990	6.5563	4.4820
GF(SRSR)	–0.2329	0.6685	6.5718	5.0307
Our-NCST	**0.0385**	**0.6956**	**6.9097**	**7.3484**
NSCT(SLWSR)	–0.2777	0.5934	6.8280	4.8805
NSCT(SRSR)	–0.1714	0.6615	6.8432	5.4925
Li's	–0.1939	0.6636	6.5135	5.3025
Zhu's(SLWSR)	–0.3237	0.6276	6.3008	4.2576
Zhu's(SRSR)	–0.3463	0.7051	5.9441	4.0452

The optimal values are shown in bold.

Table 2
Quantitative assessment of different fusion methods for fusion and super-resolution of the second infrared and visible image pair.

Methods	Q_{SF}	Q_{QAC}	Q_{ENT}	Q_{GD}
Ours-Nec	**−0.0559**	**0.7379**	7.0537	**5.5135**
Ours-CNN	**0.0547**	**0.7127**	7.0444	**6.1798**
CNN(SLWSR)	−0.2869	0.6541	7.0603	4.1224
CNN(SRSR)	−0.1592	0.6918	**7.0775**	4.9370
Ours-CT	**0.1107**	**0.7061**	**6.7522**	**6.0184**
CT(SLWSR)	−0.1538	0.5734	6.7061	4.3425
CT(SRSR)	−0.0602	0.6440	6.7211	4.7768
Our-GF	**−0.0367**	**0.7287**	**6.6246**	**5.4533**
GF(SLWSR)	−0.3785	0.6319	6.5658	3.3471
GF(SRSR)	−0.2916	0.6959	6.5835	3.8423
Our-NCST	**0.0169**	**0.7180**	6.9164	**5.8202**
NSCT(SLWSR)	−0.3091	0.6252	6.9050	3.8632
NSCT(SRSR)	−0.2120	0.6800	**6.9194**	4.4260
Li's	−0.3961	0.7113	6.3704	3.3851
Zhu's(SLWSR)	−0.4140	0.6593	6.3253	3.1393
Zhu's(SRSR)	−0.4283	0.7288	6.0929	3.1458

The optimal values are shown in bold.

As we can be seen from the marked enlarged regions and the objective assessments in Table 1, the fused results of Ours-Nec outperform those of other methods in terms of visual perception and objective evaluation. But through careful observation we can find that Our-Nec has great performance in overall and local contrast enhancement, while the boundaries of details are not so clear (see Fig. 8(a)). These visual deficiencies have been improved in the results of Ours-CNN, Ours-CT, Ours-GF, and Ours-NSCT. It indicates that these methods can integrate the advantages of CNN, CT, GF, NSCT based methods into Ours-CNN, Ours-CT, Ours-GF, and Ours-NSCT methods. Thus, the edge details can be better preserved in our methods, if the methods based on CNN, CT, GF and NSCT have strong ability in preserving image edge details. In addition, by comparing the fusion results of our methods with those of competitive ones, it can also be found that the results obtained by the proposed method are superior to those obtained by direct super-resolution reconstruction on Fig. 6 in terms of visual effect and detail preservation, because the proposed method can integrate the advantages of other methods. Thus, conclusion can be drawn that the algorithm designed in this paper is very effective and reasonable.

4.3. Fusion and super-resolution of medical image

In the second experiment, two pairs of low-resolution medical images shown in Fig. 5(b) are fused. It can be known from these source images that the image information of the same part obtained by different devices is complementary. The goal of medical image fusion and super-resolution is to extract complementary information from low-resolution source images and inject it into the fused image, and meanwhile improve the resolution of the fused results. Moreover, the low-resolution source images shown in Fig. 5(b) are also fused and reconstructed by other competitive methods, and the factor of fusion and super-resolution results is twice that of their source images (see Fig. 4). (Table 2).

Figs. 9 and 10 show the fused and reconstructed results of different methods. As can be seen from Figs. 9(a), (f) and (p) and Figs. 10(a), (f) and (p), the proposed Our-Nec method is superior to Li's method and those step-by-step operation methods, including CNN (SLWSR), CNN (SRSR), CT (SLWSR), CT (SRSR), GF (SLWSR), GF (SRSR), NSCT (SLWSR), NSCT(SRSR), Zhu's (SLWSR) and Zhu's (SRSR) methods. Moreover, from the objective evaluation values listed in Tables 3 and 4, and the quality of visual perception of Figs. 9(b)–(m) and Figs. 10(b)–(m), it can be found that the results of our methods have the best visual quality with no obvious artifacts, that's because our framework can adjust the level of image visualization according to the input reference fused image, and also can preserve and enhance the edge details in a more efficient way due to the injection of compensation information. This proves that the developed joint framework outperforms the operation that image fusion and image super-resolution are performed separately.

4.4. Fusion and super-resolution of multi-focus image

In the third experiment, two set of multi-focus image pairs shown in the third row in Fig. 5(c) are utilized as the test samples. The low-resolution test images are created by downsampling two scale factors from their high-resolution images shown in the third row in Fig. 4. Both images in each high-resolution multi-focus image pair are captured by the same sensor modality but with different focus regions. Multi-focus image fusion is one of the most popular methods for obtaining an image with all objects focused. The fusion and super-resolution joint framework proposed in this paper is also suitable for fusion and super-resolution of multi-focus images. The fused results of different fusion methods are presented in Figs. 11 and 12.

By comparing the visual qualities of Fig. 11(a) and Figs. 11(f)–(p), Fig. 12(a) and Figs. 12(f)–(p), it can be known that our method Our-Nec outperforms other competitive methods in improving the contrast of the fused result as well as the visualization of fusion results. The only drawback of Our-Nec method is that some artifacts appear on the boundary of the object (see the enlarged region in Fig. 11(a)). But, in Figs. 11(b)–(e), this defect has been remedied because the images (see the last four figures in the third line of Fig. (6)) convoluted with Fig. 11(a) does not show such a defect. During the convolution, these defects are remedied, which makes the boundary of the object clearer. At the same time, the results after convolution and information compensation show better visual quality compared with others. The objective evaluation results in Tables 5 and 6 further verify the superiority of the proposed method over other methods.

4.5. Algorithm analysis

4.5.1. Discussions of parameter selection

There are five parameters, i.e. $\lambda_i (i = 1, 2, \ldots, 5)$ in discriminative dictionary learning model (5), one parameter, namely, λ_0, in

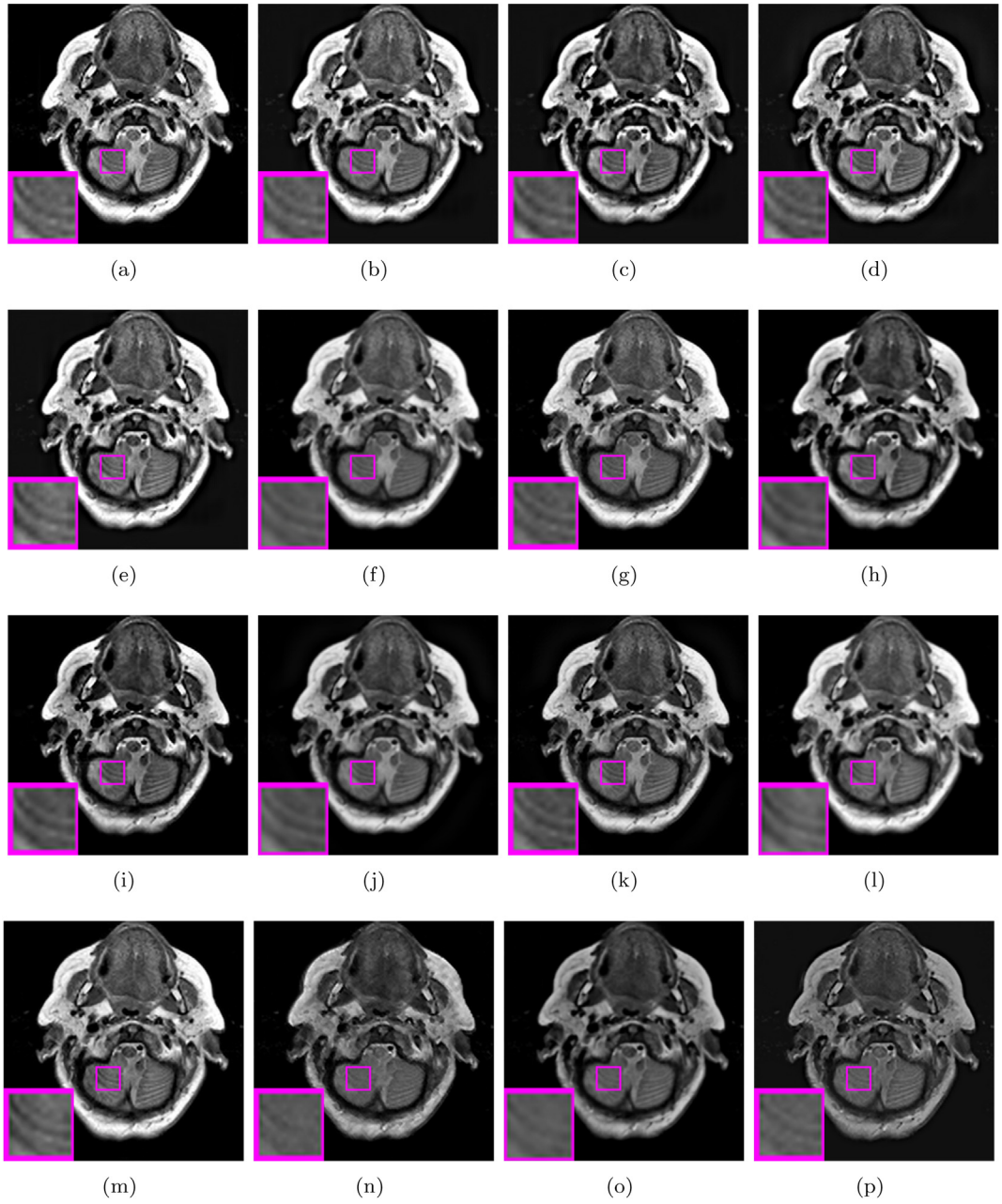

Fig. 9. Fusion and super-resolution results of the first medical image pair (each result with size of 256×256). (a) to (p) respectively represent the fused and reconstructed images by Ours-Nec, Ours-CNN, Ours-CT, Ours-GF, Ours-NSCT, CNN(SLWSR), CNN(SRSR), CT(SLWSR), CT(SRSR), GF(SLWSR), GF(SRSR), NSCT(SLWSR), NSCT(SRSR), Li's, Zhu's (SLWSR), Zhu's(SRSR).

compensation dictionary learning model (6) and the number of iterations $\mathbb{K}$ in Algorithm 1 that need to be tuned. First, we find that the dictionary learning algorithm converges when $\mathbb{K}$ reaches 10. The parameters λ_1 and λ_2 are introduced to adjust the effects of $\|\boldsymbol{D}_{h,s}\boldsymbol{Z}_{h,s} - \boldsymbol{D}_{h,s}\boldsymbol{H}\boldsymbol{Z}_{l,s}\|_F^2$ and $\left\|\frac{1}{K}\boldsymbol{A}\boldsymbol{Z}_{h,l} - \boldsymbol{H}\boldsymbol{Z}_{l,l}\right\|_F^2$. A large number of experimental results show that λ_1 and λ_2 would not affect the learned results that much when $\lambda_1 \in [0.001, 10]$ and $\lambda_2 \in [0.0001, 1]$. So we set $\lambda_1 = 1$ and $\lambda_2 = 0.001$. λ_5 is introduced to adjust the effect $\|\boldsymbol{H}\|_F^2$. To investigate the effects of λ_5, λ_4 and λ_3, we fix $\lambda_1 = 1$ and $\lambda_2 = 0.001$.

The first and second rows in Fig. 13 show the sparse dictionary pairs under different parameter settings, while the second row shows the low-rank dictionary pair. High-resolution dictionaries are displayed on the right side of each pair, while low-resolution dictionaries are displayed on the left side. Figs. 13(a)–(c) show the effect of λ_5 when $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are fixed at $1, 0.001, 1.5$, and 0.01. As we can see from these results, $\lambda_5 = 0.00001$ is a good choice for producing a pleasing dictionary pair $\boldsymbol{D}_{h,s}$ and $\boldsymbol{D}_{l,s}$. Figs. 13 (d)–(f) illustrate the visual effect of the learned low-rank dictionary pair $\boldsymbol{D}_{h,l}$ and $\boldsymbol{D}_{l,l}$ when $\lambda_1, \lambda_2, \lambda_4$ and λ_5 are fixed at $1, 0.001, 0.01$, and 0.00001 and λ_3 varies. It can be found from these results that the satisfactory visualization results can be obtained when $\lambda_3 \geqslant 1$, so we set λ_3 to 1.5. Subsequently, we investigate the influence of λ_4 on Algorithm 1. Figs. 13(g)–(i) show the learned $\boldsymbol{D}_{h,s}$ and $\boldsymbol{D}_{l,s}$ under different values of λ_4. It is noted that Algorithm 1 can produces pleasing results when λ_4 varies within a reasonable range, and the best learned results are obtained when $\lambda_4 = 0.01$.

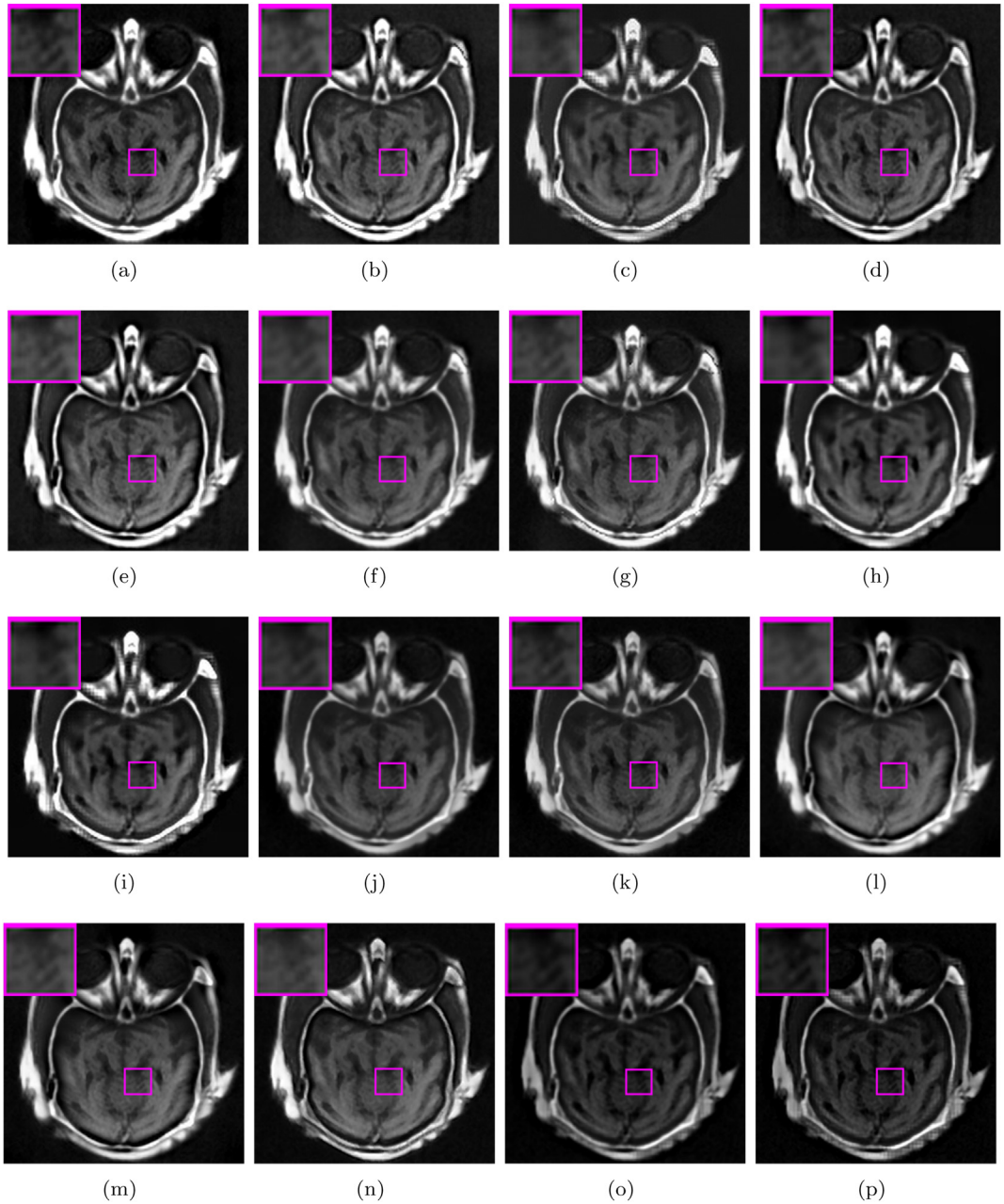

Fig. 10. Fusion and high-resolution results of the first medical image pair (each result with size of 256×256). (a) to (p) respectively represent the fused and reconstructed images by Ours-Nec, Ours-CNN, Ours-CT, Ours-GF, Ours-NSCT, CNN(SLWSR), CNN(SRSR), CT(SLWSR), CT(SRSR), GF(SLWSR), GF(SRSR), NSCT(SLWSR), NSCT(SRSR), Li's, Zhu's (SLWSR), Zhu's(SRSR).

The Algorithm 2 involves four parameters, i.e., $\beta_i (i = 1, 2, 3, 4)$. We use the same strategy as Algorithm 1 to determine the optimal values of these parameters. During this process, we employ the first medical image pair shown in Fig. 5(b) as the test samples to investigate the effects of $\beta_i (i = 1, 2, 3, 4)$. To analyze the effects of β_2, we assign the other three parameters β_1, β_3 and β_4 to $0.001, 0.001$, and 1.5, respectively, and search for the optimal value of β_2 within $[0.001, 1]$. As shown in Figs. 14(a)–(d), when β_2 varies within $[0.001, 1]$, the visual perception quality will not be affect significantly, so the value of β_2 is assigned to 0.75 throughout this paper. In addition, the optimal value of β_1 is searched for within $\{0.001, 0.01, 0, 02\}$. Figs. 14(c), (e) and (f) show that as β_1 increases, the visual perception quality declines gradually. Accordingly, it is set to 0.001 in this paper. Subsequently, to determine the value of β_3, we fix β_1, β_2 and β_4 at 0.001, 0.75, and 1.5 respectively. Figs. 14(g)–(i) show the fused results when β_3 equals 0.0001, 0.01, and 0.1 respectively. The results show that there is no significant change in visual quality when β_3 changes from 0.0001 to 0.1, so we set it to 0.001 in this paper. Figs. 14(j)–(l) illustrate the fused results when β_4 searches for the optimal value from a small set $\{1, 2.5, 3\}$, β_1, β_2 and β_3 are fixed at 0.001, 0.75, and 0.001 respectively. It can be known from these results that Algorithm 2 is sensitive to β_4; and as β_4 increases, the details of the image become clear, but in the event that its value is too large, the brightness of the image will be over enhanced, which is not conducive to the perception of human eye. Accordingly, β_4 is assigned to 2.5 in this paper.

In addition, the effect of the patch size and the dictionary size on fused results are also be examined to achieve good reconstruction quality. As confirmed in [6,35], image patch with a size of

Table 3
Quantitative assessment of different fusion methods for fusion and super-resolution of the first medical image pair.

Methods	Q_{SF}	Q_{QAC}	Q_{ENT}	Q_{GD}
Ours-Nec	**0.0629**	**0.5477**	5.1484	**9.6018**
Ours-CNN	**0.0165**	**0.7027**	**6.2841**	**9.6487**
CNN(SLWSR)	−0.2796	0.5147	5.2968	6.4145
CNN(SRSR)	−0.1608	0.5268	5.3332	7.4132
Ours-CT	**0.1009**	**0.7095**	**6.1764**	**10.1930**
CT(SLWSR)	−0.1764	0.4747	5.1443	7.1852
CT(SRSR)	−0.0371	0.4779	5.1673	8.2395
Our-GF	**0.0170**	**0.7141**	**6.5481**	**9.7152**
GF(SLWSR)	−0.2886	0.5568	6.0563	6.3772
GF(SRSR)	−0.1659	0.5870	6.0565	7.4200
Our-NCST	**0.0384**	**0.7048**	**6.2709**	**9.8728**
NSCT(SLWSR)	−0.2565	0.4971	5.2745	6.7046
NSCT(SRSR)	−0.1314	0.5149	5.3114	7.7636
Li's	−0.3078	0.5324	5.1790	6.1539
Zhu's(SLWSR)	−0.3705	0.4934	5.0227	5.4162
Zhu's(SRSR)	−0.3522	0.5515	5.4483	5.7657

The optimal values are shown in bold.

Table 4
Quantitative assessment of different fusion methods for fusion and super-resolution of the second medical image pair.

Methods	Q_{SF},	Q_{QAC}	Q_{ENT}	Q_{GD}
Ours-Nec	**0.1324**	**0.6864**	6.8639	**8.4374**
Ours-CNN	**0.2509**	**0.7109**	6.9635	**9.4897**
CNN(SLWSR)	−0.1335	0.6475	7.0456	6.4346
CNN(SRSR)	−0.0356	0.6895	**7.0543**	7.0652
Ours-CT	**0.3573**	**0.7157**	**6.7368**	**9.8230**
CT(SLWSR)	0.0365	0.5705	6.5997	7.0896
CT(SRSR)	0.1507	0.6232	6.6039	7.6955
Our-GF	**0.1602**	**0.7127**	**6.9670**	**9.0385**
GF(SLWSR)	−0.2921	0.6224	6.8149	5.3368
GF(SRSR)	−0.2230	0.6761	6.8327	5.7739
Our-NCST	**0.2270**	**0.7090**	**6.9848**	**9.2928**
NSCT(SLWSR)	−0.1720	0.6400	6.9313	6.1936
NSCT(SRSR)	−0.0955	0.6842	6.9413	6.6450
Li's	−0.0133	0.6900	6.7599	7.2269
Zhu's(SLWSR)	−0.2613	0.6613	6.3081	5.3814
Zhu's(SRSR)	−0.1783	0.6973	6.3197	5.9027

The optimal values are shown in bold.

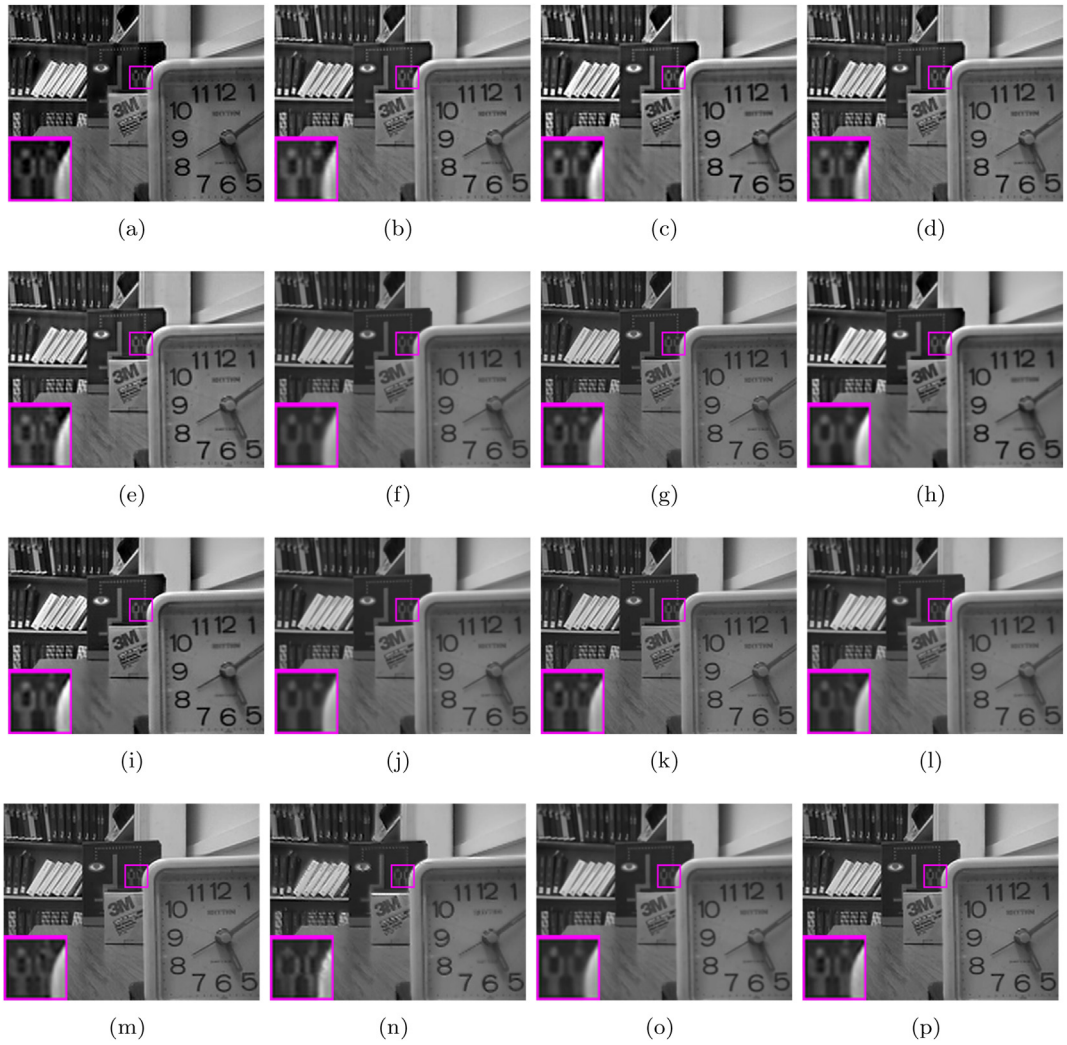

Fig. 11. Fusion and high-resolution results of the first multi-focus image pair (each result with size of 240 × 320). (a) to (p) respectively represent the fused and reconstructed images by Ours-Nec, Ours-CNN, Ours-CT, Ours-GF, Ours-NSCT, CNN(SLWSR), CNN(SRSR), CT(SLWSR), CT(SRSR), GF(SLWSR), GF(SRSR), NSCT(SLWSR), NSCT(SRSR), Li's, Zhu's (SLWSR), Zhu's(SRSR).

Fig. 12. Fusion and high-resolution results of the second multi-focus image pair (each result with size of 256×256). (a) to (p) respectively represent the fused and reconstructed images by Ours-Nec, Ours-CNN, Ours-CT, Ours-GF, Ours-NSCT, CNN(SLWSR), CNN(SRSR), CT(SLWSR), CT(SRSR), GF(SLWSR), GF(SRSR), NSCT(SLWSR), NSCT (SRSR), Li's, Zhu's(SLWSR), Zhu's(SRSR).

8×8 is very effective for image fusion. Furthermore, Figs. 15(a)–(d) shows the reconstructed results under different sizes of dictionary, in which case the numbers of atoms in these dictionaries are 128, 256, 512, and 1024, respectively. By comparison, it is found that there are no significant visual differences in these reconstructions. However, Fig. 15(e) shows that the value of PSNR increases along with the dictionary size, so does the time consumption of representation and reconstruction. Therefore, the number of atoms is set to 256 to achieve a trade-off between the reconstruction quality and time consumption.

4.5.2. Effects of key steps

The proposed image fusion and super-resolution joint framework involves three main steps, i.e., joint implementation of image fusion and super-resolution, advantage embedding and compensation of detail information. In this section, we first show the advantages of the joint image fusion and super-resolution framework over step-by-step processing. To this end, three different sets of source images as shown in Fig. 5 are selected as the test images, and the fused results are illustrated in Fig. 16, where "Ours-Nec $+C_0$" means the compensation component C_0 of detail information is added to the results of "Ours-Nec"; "Ours-Nec$\otimes$ CNN" means that the fused results of "Ours-Nec" convolutes with the results of CNN, while "Ours-Nec$\otimes$ CNN$+C_0$" means C_0 is added to the fused results of "Ours-Nec$\otimes$ CNN". Although the details of the images in Fig. 16(c) are clearer than those of others, the details of these results are over sharpened by "Ours-Nec$+C_0$", which is not conducive to the observation of human eye. The details are sharpened in "Ours-Nec$+C_0$", but from visual perception, it can be found that the results of"Ours-Nec$\otimes$ CNN$+C_0$" are superior to those of other

Table 5
Quantitative assessment of different fusion methods for fusion and super-resolution of the first multi-focus image pair.

Methods	Q_{SF}	Q_{QAC}	Q_{ENT}	Q_{GD}
Ours-Nec	**−0.0423**	**0.7003**	**7.4989**	**9.1477**
Ours-CNN	**−0.0612**	**0.7006**	**7.4281**	**9.1752**
CNN(SLWSR)	−0.3761	0.5985	7.2100	5.4329
CNN(SRSR)	−0.2326	0.6424	7.2554	6.7510
Ours-CT	**0.1102**	**0.6954**	**7.5426**	**10.4068**
CT(SLWSR)	−0.1726	0.5743	7.4470	6.9966
CT(SRSR)	−0.0002	0.6463	7.4844	8.4359
Our-GF	**−0.0586**	**0.7048**	**7.4302**	**9.1965**
GF(SLWSR)	−0.3735	0.5973	7.2144	5.4495
GF(SRSR)	−0.2291	0.6420	7.2611	6.7791
Ours-NCST	**−0.0460**	**0.7021**	**7.4438**	**9.3122**
NSCT(SLWSR)	−0.3596	0.5981	7.2542	5.5637
NSCT(SRSR)	−0.2112	0.6527	7.2958	6.9396
Li's	−0.3241	0.6226	7.2467	5.8701
Zhu's(SLWSR)	−0.4237	0.6116	7.2470	5.3555
Zhu's(SRSR)	−0.3713	0.6026	7.2377	5.4637

The optimal values are shown in bold.

Table 6
Quantitative assessment of different fusion methods for fusion and super-resolution of the second multi-focus image pair.

Methods	Q_{SF}	Q_{QAC}	Q_{ENT}	Q_{GD}
Ours-Nec	**−0.0503**	**0.6833**	7.3334	**7.6775**
Ours-CNN	**−0.0447**	**0.6578**	**7.4512**	**7.8960**
CNN(SLWSR)	−0.3553	0.5698	7.3690	4.9446
CNN(SRSR)	−0.2255	0.6258	7.3501	5.8996
Ours-CT	**−0.0134**	**0.6474**	**7.4632**	**8.0572**
CT(SLWSR)	−0.2583	0.5274	7.4210	5.1107
CT(SRSR)	−0.1899	0.5732	7.4089	6.1601
Our-GF	**−0.0420**	**0.6567**	**7.4514**	**7.9490**
GF(SLWSR)	−0.2978	0.5534	7.3754	4.8614
GF(SRSR)	−0.2233	0.6112	7.3935	5.9655
Ours-NCST	**−0.0196**	**0.6601**	**7.4736**	**8.1211**
NSCT(SLWSR)	−0.2702	0.5580	7.4130	5.0473
NSCT(SRSR)	−0.1965	0.6159	7.4071	6.1901
Li's	−0.3353	0.5812	7.3895	5.1484
Zhu's(SLWSR)	−0.2831	0.5635	7.3811	4.9497
Zhu's(SRSR)	−0.3397	0.6024	7.1953	5.0554

The optimal values are shown in bold.

methods. More importantly, some artifacts in the enlarged region in Fig. 16(a) are effectively suppressed in the final results of "Ours-Nec⊗ CNN+C_0".

In order to verify the effectiveness and superiority of the proposed joint image fusion and super-resolution framework over step-by-step method, the joint framework is divided into two independent approaches, i.e., fusion and super-resolution. Through step-by-step operations, the fusion and super-resolution of low-resolution image can be achieved. The fused and reconstructed results of the joint and step-by-step implementation are shown in Fig. 17, where "Fus-Sur" indicates that the low-resolution image is fused by the strategy proposed in this paper, and then the super-resolution image is reconstructed by our reconstruction method, while "SuR-Fus" indicates that the low-resolution source images are first reconstructed for super-resolution by employing the method developed in this paper, and then the reconstructed results are fused by the simple fusion method proposed in this paper. By comparing the visual effects of Figure.17, it is clear that the joint

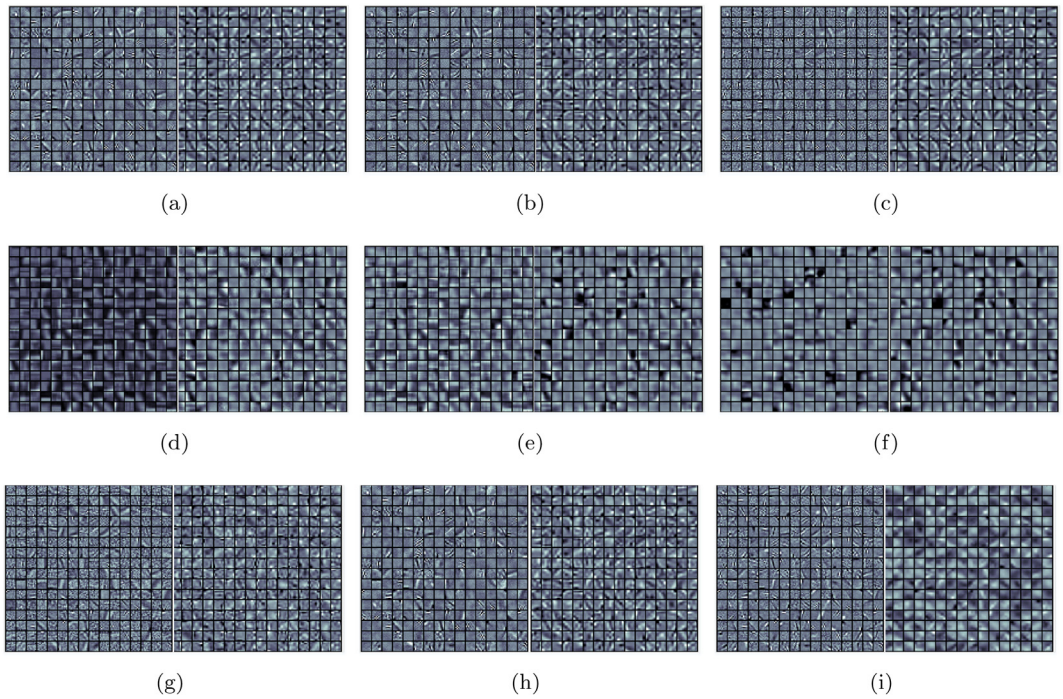

Fig. 13. The effects of different parameters on learned dictionaries. (a) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 1.5, \lambda_4 = 0.01$, and $\lambda_5 = 0.00001$ (b) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 1.5, \lambda_4 = 0.01$, and $\lambda_5 = 0.1$ (c) Produced at $\lambda_1 = 1; \lambda_2 = 0.001, \lambda_3 = 1.5, \lambda_4 = 0.01$, and $\lambda_5 = 1.5$ (d) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 0.001, \lambda_4 = 0.01$, and $\lambda_5 = 0.00001$ (e) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 1, \lambda_4 = 0.01$, and $\lambda_5 = 0.00001$ (f) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 3; \lambda_4 = 0.01$, and $\lambda_5 = 0.00001$ (g) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 1.5, \lambda_4 = 0.001$, and $\lambda_5 = 0.00001$ (h) Produced at $\lambda_1 = 1, \lambda_2 = 0.001, \lambda_3 = 1.5, \lambda_4 = 0.01$, and $\lambda_5 = 0.00001$ (i) Produced at $\lambda_1 = 1; \lambda_2 = 0.001, \lambda_3 = 1.5, \lambda_4 = 0.1$, and $\lambda_5 = 0.00001$.

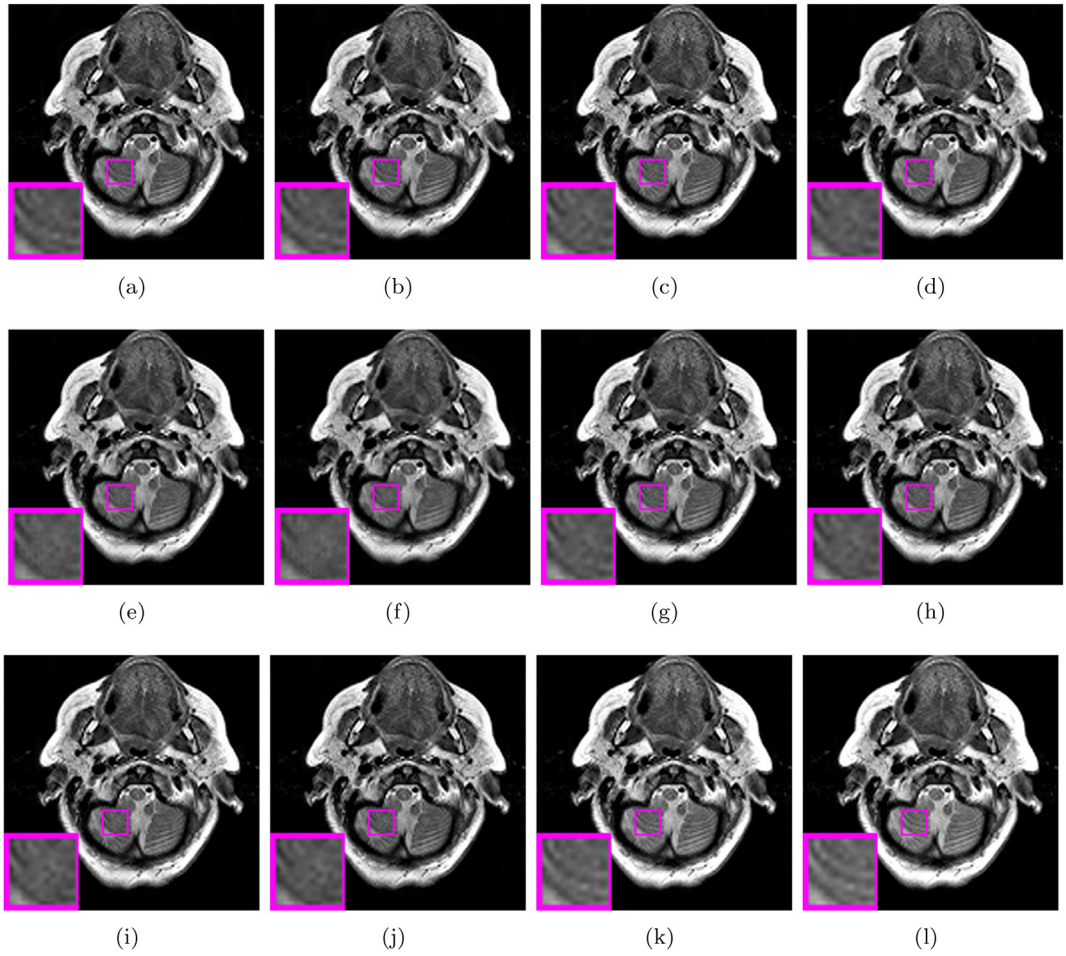

Fig. 14. The effect of different parameters on fused results. (a) Produced at $\beta_1 = 0.001, \beta_2 = 0.001, \beta_3 = 0.001$, and $\beta_4 = 1.5$ (b) Produced at $\beta_1 = 0.001, \beta_2 = 0.01, \beta_3 = 0.001$, and $\beta_4 = 1.5$ (c) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.001$, and $\beta_4 = 1.5$ (d) Produced at $\beta_1 = 0.001, \beta_2 = 1, \beta_3 = 0.001$, and $\beta_4 = 1.5$ (e) Produced at $\beta_1 = 0.01, \beta_2 = 0.75, \beta_3 = 0.001$, and $\beta_4 = 1.5$ (f) Produced at $\beta_1 = 0.02, \beta_2 = 0.75, \beta_3 = 0.001$, and $\beta_4 = 1.5$ (g) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.0001$, and $\beta_4 = 1.5$ (h) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.01$, and $\beta_4 = 1.5$ (i) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.1$, and $\beta_4 = 1.5$ (j) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.001$, and $\beta_4 = 1.0$ (k) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.001$, and $\beta_4 = 2.5$ (l) Produced at $\beta_1 = 0.001, \beta_2 = 0.75, \beta_3 = 0.001$, and $\beta_4 = 3$.

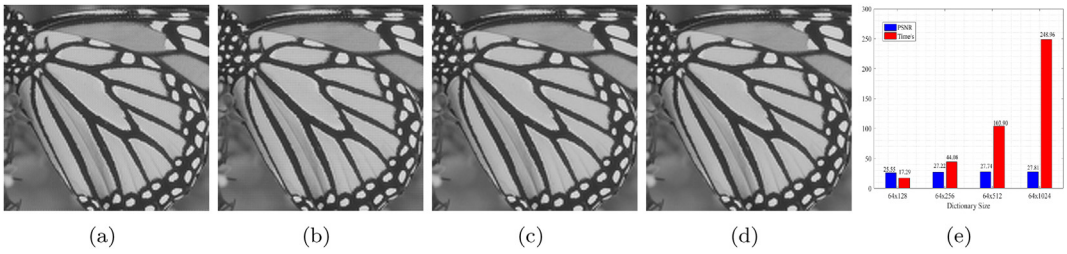

Fig. 15. Effects of dictionary size on the reconstruction of the butterfly image. From left to right: (a) dictionary size 64×128, (b) dictionary size 64×256, (c) dictionary size 64×512, (d) dictionary size 64×1024, (e) PSNR and time consumption of these images.

processing strategy outperforms the step-by-step method in preserving the edge details of the source images and improving the visual perception quality of final results. That's because the step-by-step operation needs a dictionary to represent the source image and the fused image (or super-resolution image), whereas the joint implementation only needs to represent the source images, and multiple image representation operations are more likely to lead to the loss of image information. In addition, as can be seen from the enlarged area of Fig. 17, step-by-step operation can easily introduce artifacts. If these artifacts are introduced in the first step, they may be further magnified in the latter operation. The quantitative comparison of the results of Fig. 17 is shown in Fig. 18, and it can be seen that most of the metrics assign larger values to the results of "Ours-Nec", which further verifies the superiority of joint operation.

4.6. Convergence analysis

Our model in Eq. (5) is convex for $\boldsymbol{D}_0$ due to other variables are fixed, thus it convergence can be ensured. The model in Eq. (4) is

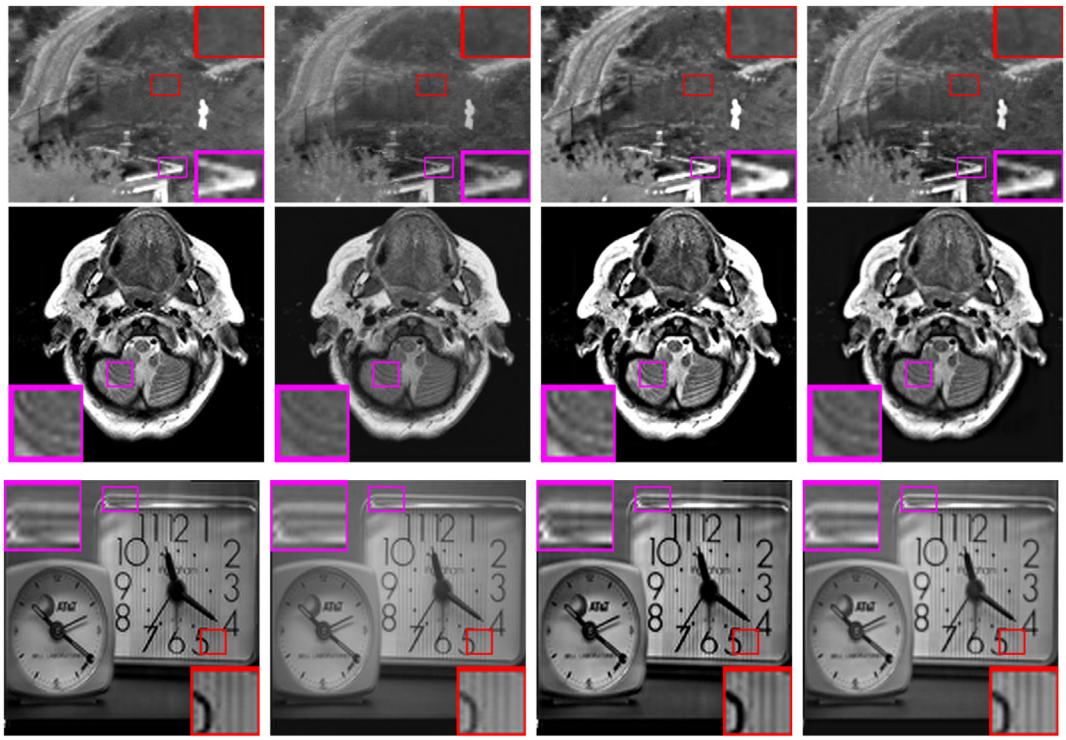

Fig. 16. Effectiveness of different components: (a) Fused results of "Ours-Nec", (b) fused results of "Ours-Nec⊗ CNN", (c) results of "Ours-Nec+C_0", (d) results of "Ours-Nec⊗ CNN+C_0".

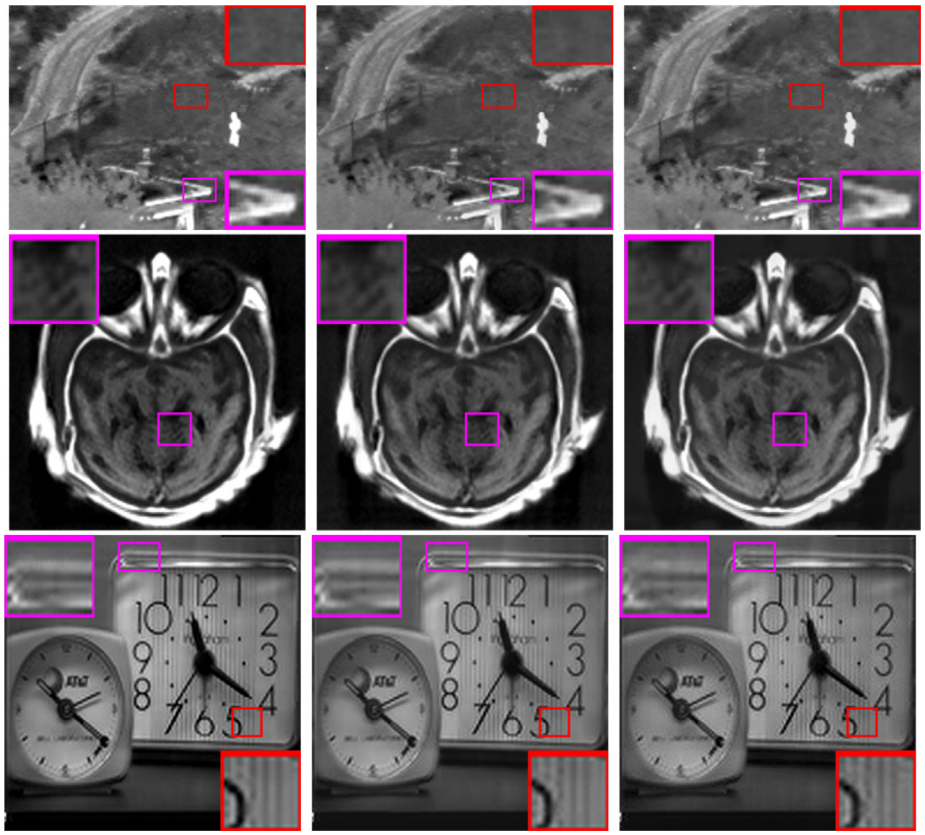

Fig. 17. Superiority of joint image fusion and super-resolution framework over step-by-step method. The fusion results obtained by "Ours-Nec", "SuR-Fus", and "Fus-SuR" are presented from left to right.

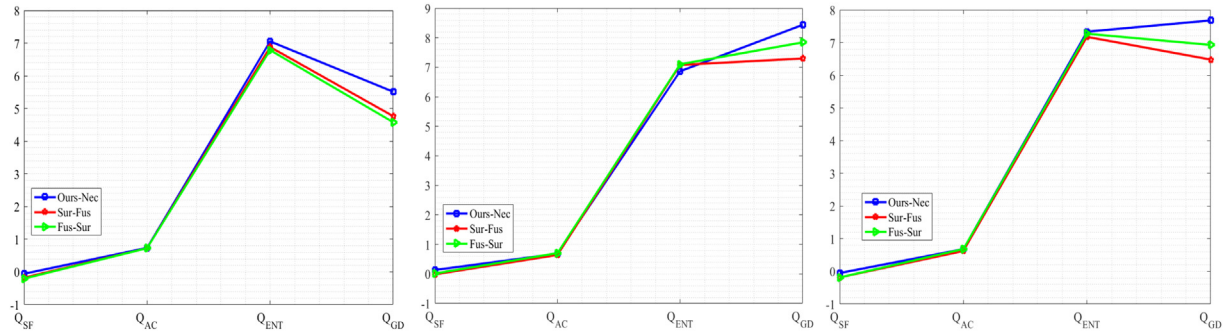

Fig. 18. Quantitative comparison of the results of Fig. 17. (a) Quantitative comparison of the first row in Fig. 17, (b) quantitative comparison of the second row in Fig. 17, (c) quantitative comparison of the third row in Fig. 17.

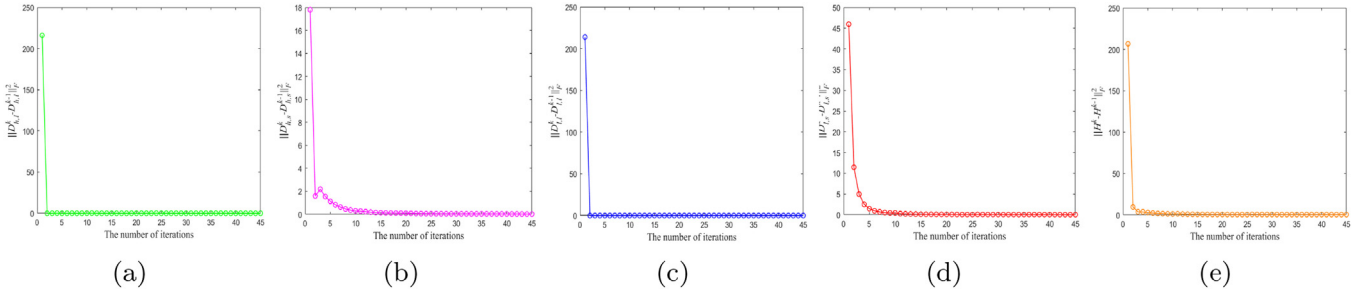

Fig. 19. Convergence analysis of Algorithm 1. (a) Convergence curve of $\boldsymbol{D}_{h,l}$ (b) Convergence curve of $\boldsymbol{D}_{h,s}$, (c) Convergence curve of $\boldsymbol{D}_{l,l}$, (d) Convergence curve of $\boldsymbol{D}_{l,s}$, (e) Convergence curve of $\boldsymbol{H}$.

non-convex jointly with matrix variables $\boldsymbol{D}_{h,l}, \boldsymbol{D}_{h,s}, \boldsymbol{D}_{l,l}, \boldsymbol{D}_{l,s}$ and $\boldsymbol{H}$. However, when we update one variable and keep the rest fixed, the model is convex, and we adopt the standard alternate iterative learning procedure to obtain the optimal solution of each variable. Fig. 19 shows the convergence curves of Algorithm 1 on the training data of dictionary. From these results we can see that Algorithm 1 converges to a stable value for each variable when the iterations reaches 10. Thus in Algorithm 1 we set $\mathbb{K}$ to 10.

4.7. Super-resolution with triple factor

Finally, the performance of our method in reconstructing the super-resolution fused image with magnification factor 3 is tested. As discussed above, in addition to our methods, reconstruction methods based on SRSR have better performances than the "Bicubic"-based methods. Thus, in this section only "CNN (SRSR)", "CT (SRSR)", "GF (SRSR)", and "NSCT (SRSR)" are selected as the competitive methods. The fused and reconstructed results of different methods are shown in Fig. 20, and the corresponding low-resolution source images are given in Fig. 5. From the local enlarged region of the fusion results of infrared visible images, it can be known that the proposed method is superior to other methods in improving the contrast and sharpening the boundary of edge structure. For medical images, the results obtained by our method and others have similar visual quality. However, the objective evaluation results illustrated in Fig. 21 indicate that our method is superior to other algorithms in quality. Similar to fusion of infrared visible images, our method performs better in preserving image edge detail and contrast in fusing multi-focus images; Furthermore, the quantitative evaluation shown in Fig. 21 also demonstrates that our method outperforms the others. That's mainly because we develop a more efficient image decomposition, as well as information compensation method, which play a positive role in preserving the image edge detail and contrast.

5. Conclusion

In this paper, a novel semi-coupled discriminative dictionary learning and advantage embedding approach is developed for enhanced visualization of joint image fusion and super-resolution. In our method, the input source images are separated into low-rank sparse components. To achieve the image fusion and super-resolution simultaneously, we first establish the potential relationship between low-resolution images and high-resolution images through the relation transformation dictionary, and then two pairs of low-rank sparse dictionaries and a conversion dictionary are jointly learned. To improve the visualization of the fused and super-resolution reconstructed result, the structure information compensation dictionary is used to compensate for the lost details. Furthermore, to integrate the advantages of excellent image fusion methods, such as CNN-based method, CT-based method, GF-based method, and NSCT-based method, into the fusion and super-resolution results, we propose a deconvolution-based advantage embedding scheme to refine the fusion and reconstruction results to make it more suitable for human eye perception. Experimental results demonstrate the effectiveness of the proposed method. However, our method fails to achieve image fusion and deblurring simultaneously. Thus, further studies will be carried out on the joint implementation of image fusion and deblurring based on the latest progress of deep learning.

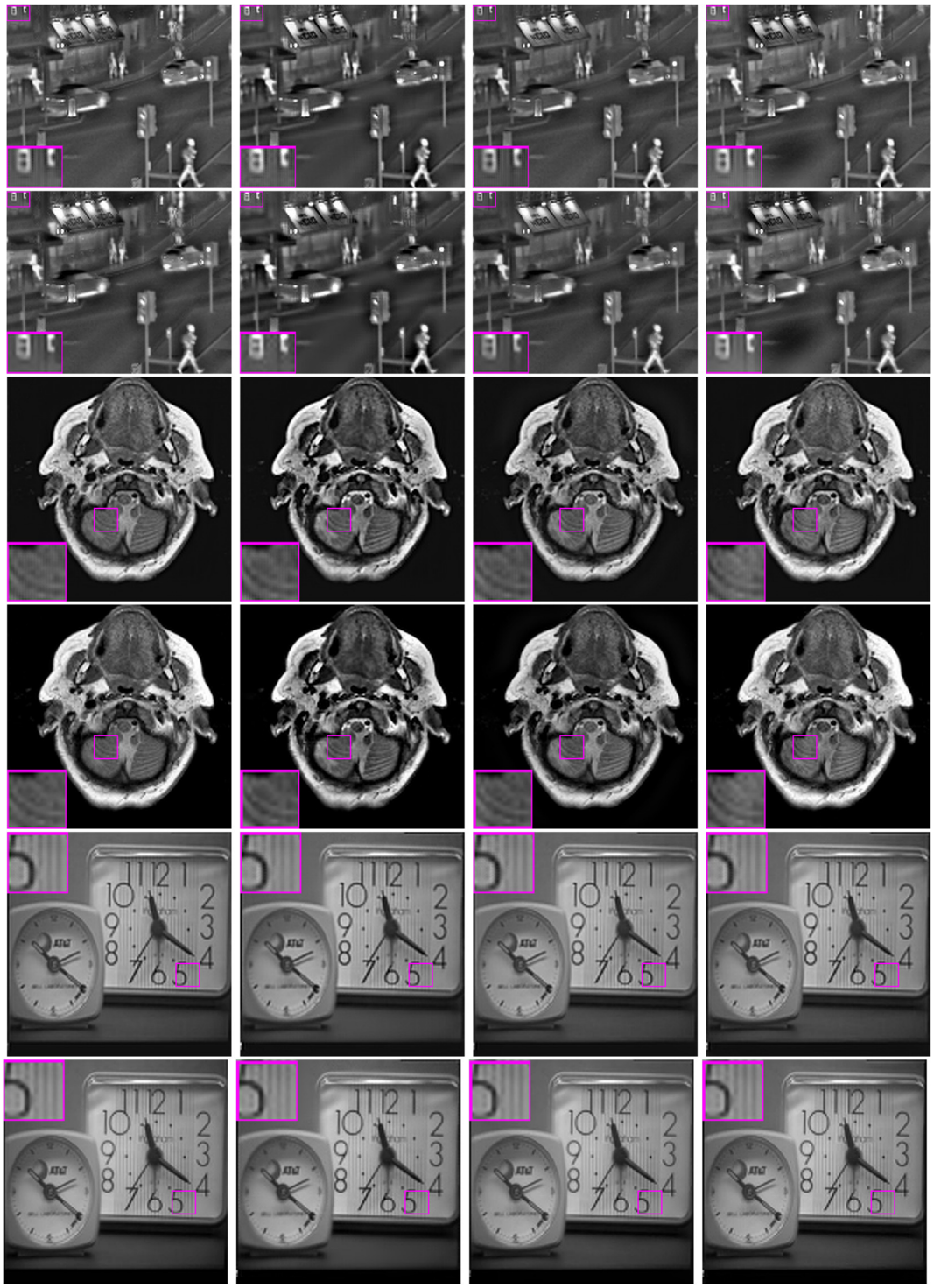

Fig. 20. Results with magnification fact 3. From left to right and top to bottom: reconstruction results generated by "Our-CNN", "Our-CT", "Our-GF", "Our-NSCT", "CNN (SRSR)", "CT(SRSR)", "GF(SRSR)", and "NSCT(SRSR)".

CRediT authorship contribution statement

Huafeng Li: Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Moyuan Yang:** Software, Data curation, Validation, Writing - original draft. **Zhengtao Yu:** Investigation, Resources, Writing - review & editing.

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.

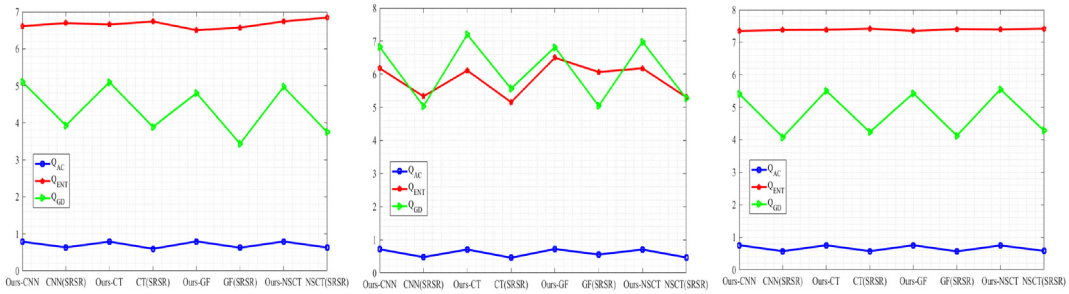

Fig. 21. Quantitative comparison of the results of Fig. 20. (a) quantitative comparison of the first row of Fig. 20 (b) quantitative comparison of the second row of Fig. 20, (c) quantitative comparison of the third row of Fig. 20.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (61562053, 61302041, 61563025, 61572486), the National Key Research and Development Plan Project (Nos.2018YFC0830105, 2018YFC0830100), Yunnan Science and Technology Project(2016FD039, 2016FB105, 2017FB094, 2016FB109) and Kunming University of Science and Technology to introduce talent fund (KKSY201403116).

References

[1] R. Singh, A. Khare, Fusion of multimodal medical images using daubechies complex wavelet transform–a multiresolution approach, Information Fusion 19 (2014) 49–60.
[2] Y. Yang, Multimodal medical image fusion through a new dwt based technique, in: International Conference on Bioinformatics and Biomedical Engineering (iCBBE), IEEE, 2010, pp. 1–4.
[3] Z. Zhu, M. Zheng, G. Qi, D. Wang, Y. Xiang, A phase congruency and local laplacian energy based multi-modality medical image fusion method in nsct domain, IEEE Access 7 (2019) 2169–3536.
[4] H. Li, X. Liu, Z. Yu, Y. Zhang, Performance improvement scheme of multifocus image fusion derived by difference images, Signal Processing 128 (2016) 474–493.
[5] H. Li, H. Qiu, Z. Yu, Y. Zhang, Infrared and visible image fusion scheme based on nsct and low-level visual features, Infrared Physics & Technology 76 (2016) 174–184.
[6] H. Li, X. He, D. Tao, Y. Tang, R. Wang, Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning, Pattern Recognition 79 (2018) 130–146.
[7] Z. Zhu, Y. Chai, H. Yin, A. Li, A novel dictionary learning approach for multi-modality medical image fusion, Neurocomputing 214 (2016) 471–482.
[8] H. Li, Y. Wang, Z. Yang, R. Wang, X. Li, D. Tao, Discriminative dictionary learning-based multiple component decomposition for detail-preserving noisy image fusion, IEEE Transactions on Instrumentation and Measurement 69 (4) (2020) 1082–1102.
[9] Z. Zhu, H. Yin, Y. Chai, Y. Li, G. Qi, A novel multi- modality image fusion method based on image decomposition and sparse representation, Information Sciences 432 (2018) 516–529.
[10] H. Li, X. He, Z. Yu, J. Luo, Noise-robust image fusion with low-rank sparse decomposition guided by external patch prior, Information Sciences 523 (2020) 14–37.
[11] Y. Liu, X. Chen, Z. Wang, Z. Wang, R. Ward, X. Wang, Deep learning for pixel-level image fusion: Recent advances and future prospects, Information Fusion 42 (2018) 158–173.
[12] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Information Fusion 36 (2017) 191–207.
[13] H. Jung, Y. Kim, H. Jang, N. Ha, K. Sohn, Unsupervised deep image fusion with structure tensor representations, IEEE Transactions on Image Processing 29 (2020) 3845–3858.
[14] J. Ma, H. Xu, J. Jiang, X. Mei, X. Zhang, Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion, IEEE Transactions on Image Processing 29 (2020) 4980–4995.
[15] M. Wang, X. Liu, H. Jin, A generative image fusion approach based on supervised deep convolution network driven by weighted gradient flow, Image and Vision Computing 86 (2019) 1–16.
[16] S. Ioannidou, V. Karathanassi, Investigation of the dual-tree complex and shift-invariant discrete wavelet transforms on quickbird image fusion, IEEE Geoscience and Remote Sensing Letters 4 (1) (2007) 166–170.
[17] S. Aymaz, C. Köse, A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion, Information Fusion 45 (2019) 113–127.
[18] X. Yuan, C. Pun, C. Chen, Robust mel-frequency cepstral coefficients feature detection and dual-tree complex wavelet transform for digital audio watermarking, Information Sciences 298 (2015) 159–179.
[19] W. Lim, The discrete shearlet transform: A new directional transform and compactly supported shearlet frames, IEEE Transactions on Image Processing 19 (2010) 1166–1180.
[20] K. Thakur, O. Damodare, A. Sapkal, Hybrid method for medical image denoising using shearlet transform and bilateral filter, in: in: 2015 International Conference on Information Processing (ICIP), IEEE, 2015, pp. 220–224.
[21] E. Candes, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Modeling and Simulation 5 (3) (2006) 861–899.
[22] M. Do, M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, IEEE Transactions on Image Processing 14 (12) (2005) 2091–2106.
[23] A. Cunha, J. Zhou, M. Do, The nonsubsampled contourlet transform: Theory, design and applications, IEEE Transactions on Image Processing 15 (10) (2006) 3089–3101.
[24] Y. Liu, X. Chen, J. Cheng, H. Peng, Z. Wang, Infrared and visible image fusion with convolutional neural networks, International Journal of Wavelets, Multiresolution and Information Processing 16 (3) (2018) 1850018.
[25] K. Ma, Z. Duanmu, H. Zhu, Y. Fang, Z. Wang, Deep guided learning for fast multi-exposure image fusion, IEEE Transactions on Image Processing 29 (2020) 2808–2819.
[26] J. Li, X. Guo, G. Lu, B. Zhang, Y. Xu, F. Wu, D. Zhang, Drpl: Deep regression pair learning for multi-focus image fusion, IEEE Transactions on Image Processing 29 (2020) 4816–4831.
[27] S. Singh, R.S. Anand, Multimodal medical image fusion using hybrid layer decomposition with cnn-based feature mapping and structural clustering, IEEE Transactions on Instrumentation and Measurement 69 (6) (2020) 3855–3865.
[28] H. Yin, S. Li, L. Fang, Simultaneous image fusion and super-resolution using sparse representation, Information Fusion 14 (2013) 229–240.
[29] M. Iqbal, J. Chen, Unification of image fusion and super-resolution using jointly trained dictionaries and local information contents, IET Image Processing 6 (9) (2012) 1299–1310.
[30] S. Wang, L. Zhang, Y. Liang, Q. Pan, Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2012) 2216–2223.
[31] H. Li, Z. Yu, C. Mao, Fractional differential and variational method for image fusion and super-resolution, Neurocomputing 171 (2016) 138–148.
[32] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Transactions on Image Processing 22 (4) (2013) 1620–1630.
[33] F. Wu, X. Jing, X. You, D. Yue, Hu Ruimin, J. Yang, Multi-view low-rank dictionary learning for image classification, Pattern Recognition 50 (2016) 143–154.
[34] J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution, IEEE Transactions on Image Processing 21 (8) (2012) 3467–3478.
[35] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Transactions on Instrumentation and Measurement 59 (4) (2010) 884–892.
[36] Y. Zhang, M. Yang, N. Li, Z. Yu, Analysis-synthesis dictionary pair learning and patch saliency measure for image fusion, Signal Processing 167 (2020) 107327.
[37] M. Xie, Z. Zhou, Y. Zhang, Joint framework for image fusion and super-resolution via multicomponent analysis and residual compensation, IEEE Access 7 (2019) 174092–174107.
[38] M. Kim, D. Han, H. Ko, Joint patch clustering-based dictionary learning for multimodal image fusion, Information Fusion 27 (2016) 198–214.
[39] S. Li, H. Yin, L.F. and, Group-sparse representation with dictionary learning for medical image denoising and fusion, IEEE Transactions on Biomedical Engineering 59 (12) (2012) 3450–3459.
[40] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Information Fusion 24 (2015) 147–164.

[41] T.H.J. Yang, J. Wright, Y. Ma, Image super-resolution as sparse representation of raw image patches, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2008, pp. 1–8.
[42] H. Li, J. Xu, Z. Yu, J. Luo, Jointly learning commonality and specificity dictionaries for person re-identification, IEEE Transactions on Image Processing 29 (2020) 7345–7358.
[43] H. Li, S. Yan, Z. Yu, D. Tao, Attribute-identity embedding and self-supervised learning for scalable person re-identification, IEEE Transactions on Circuits and Systems for Video Technology (2020), https://doi.org/10.1109/TCSVT.2019.2952550.
[44] J. Song, X. Xie, G. Shi, W. Dong, Multi-layer discriminative dictionary learning with locality constraint for image classification, Pattern Recognition 91 (2019) 135–146.
[45] Y. Rong, S. Xiong, Y. Gao, Low-rank double dictionary learning from corrupted data for robust image classification, Pattern Recognition 72 (2017) 419–432.
[46] Z. Zhu, G. Qi, Y. Chai, P. Li, A geometric dictionary learning based approach for fluorescence spectroscopy image fusion, Applied Sciences 7 (2) (2017) 161.
[47] S. Singh, R.S. Anand, Multimodal medical image sensor fusion model using sparse k-svd dictionary learning in nonsubsampled shearlet domain, IEEE Transactions on Instrumentation and Measurement 69 (2) (2020) 593–607.
[48] S. Singh, R.S. Anand, D. Gupta, Ct and mr image information fusion scheme using a cascaded framework in ripplet and nsst domain, IET Image Process 12 (5) (2018) 696–707.
[49] Y. Jiang, M. Wang, Image fusion with morphological component analysis, Information Fusion 18 (2014) 108–118.
[50] W. Liu, S. Li, Multi-morphology image super-resolution via sparse representation, Neurocomputing 120 (2013) 645–654.
[51] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, The European Conference on Computer Vision (ECCV) (2018) 286–301.
[52] K. Zhang, W. Zuo, L. Zhang, Deep plug-and-play super-resolution for arbitrary blur kernels, in: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1671–1681.
[53] B. Li, J. Liu, B. Wang, Z. Qi, Y. Shi, s-lwsr: Super lightweight super-resolution network (2019) arXiv:1909.10774.
[54] K. Nazeri, H. Thasarathan, M. Ebrahimi, Edge-informed single image super-resolution, in: The IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
[55] Z. Ding, M. Shao, Y. Fu, Low-rank embedded ensemble semantic dictionary for zero-shot learning, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 2050–2058.
[56] J. Cai, E. Candes, Z. Shen, A singular value thresholding algorithm for matrix completion, Siam Journal on Optimization 20 (4) (2008) 1956–1982.
[57] I. Daubechies, M. Defriese, C. DeMol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on Pure & Applied Mathematics 57 (11) (2004) 1413–1457.
[58] A. Beck, M. Teboulle, A fast iterative shrinkage thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (1) (2009) 183–202.
[59] J. Bioucas-Dias, M. Figueiredo, A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration, IEEE Transactions on Image Processing 16 (12) (2007) 2992–3004.
[60] H. Lee, A. Battle, R. Raina, A. Ng, Efficient sparse coding algorithms, in: Conference and Workshop on Neural Information Processing Systems(NIPS), vol. 19, 2006, pp. 801–808.
[61] F. Nie, H. Huang, X. Cai, C. Ding, Efficient and robust feature selection via joint l_2,1)norms minimization, in: Annual Conference on Neural Information Processing Systems(NIPS), vol. 2, 2010, pp. 1813–1821.
[62] T. Cai, L. Wang, Orthogonal matching pursuit for sparse signal recovery with noise, IEEE Transactions on Image Processing 57 (7) (2011) 4680–4688.
[63] D. Capel, Image mosaicing and super-resolution, Ph.d. thesis, University of Oxford (3 2001).
[64] M. Irani, S. Peleg, Motion analysis for image enhancement: Resolution, occlusion, and transparency, Journal of Visual Communication and Image Representation 4 (4) (1993) 324–335.
[65] S. Li, X. Kuang, J. Hu, Image fusion with guided filtering, IEEE Transactions on Image Processing 22 (7) (2013) 2864–2875.
[66] Y. Zheng, E. Essock, B. Hansen, A. Haun, A new metric based on extended spatial frequency and its application to dwt based fusion algorithm, Information Fusion 8 (2) (2007) 177–192.
[67] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganière, Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision:a comparative study, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (1) (2012) 94–109.
[68] W. Xue, L. Zhang, X. Mou, Learning without human scores for blind image quality assessment, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 995–1002.

Huafeng Li received the M.S. degree in applied mathematics major from Chongqing University in 2009 and obtained his Ph.D. degree in control theory and control engineering major from Chongqing University in 2012. He is currently a professor at the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, China. His research interests include image processing, computer vision, and information fusion.

Moyuan Yang received his M.E. degree in electronics and communications engineering from Kunming University of Science and Technology in 2020. Her research interests include image processing and computer vision.

Zhengtao Yu received his Ph.D degree in computer application technology from Beijing Institute of Technology, Beijing, China, in 2005. He is currently a professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, China. His main research interests include natural language process, image processing and machine learning.