Image Super-Resolution by Supervised Adaption of Patchwise Self-similarity from High-Resolution Image

Guorong Wu[1]([✉]), Xiaofeng Zhu[1], Qian Wang[2], and Dinggang Shen[1]

[1] Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, Chapel Hill, USA
guorong_wu@med.unc.edu
[2] Med-X Research Institute of Shanghai Jiao Tong University, Shanghai, China

Abstract. Image super-resolution is of great interest in medical imaging field. However, different from natural images studied in computer vision field, the low-resolution (LR) medical imaging data is often a stack of high-resolution (HR) 2D slices with large slice thickness. Consequently, the goal of super-resolution for medical imaging data is to reconstruct the missing slice(s) between any two consecutive slices. Since some modalities (e.g., T1-weighted MR image) are often acquired with high-resolution (HR) image, it is intuitive to harness the prior self-similarity information in the HR image for guiding the super-resolution of LR image (e.g., T2-weighted MR image). The conventional way is to find the profile of patchwise self-similarity in the HR image and then use it to reconstruct the missing information at the same location of LR image. However, the local morphological patterns could vary significantly across the LR and HR images, due to the use of different imaging protocols. Therefore, such direct (un-supervised) adaption of self-similarity profile from HR image is often not effective in revealing the actual information in the LR image. To this end, we propose to employ the existing image information in the LR image to supervise the estimation of self-similarity profile by requiring it *not only* being optimal in representing patches in the HR image, *but also* producing less reconstruction errors for the existing image information in the LR image. Moreover, to make the anatomical structures spatially consistent in the reconstructed image, we simultaneously estimate the self-similarity profiles for a stack of patches across consecutive slices by solving a group sparse patch representation problem. We have evaluated our proposed super-resolution method on both simulated brain MR images and real patient images with multiple sclerosis lesion, achieving promising results with more anatomical details and sharpness.

1 Introduction

In most imaging-based studies and clinical diagnosis, image resolution is important to reveal disease-specific imaging markers and quantify the structure/functional difference across individual subjects. However, a high-resolution 3D image is not always available due to consideration of radiation dose or scanning time, thus leading to a compromised low-resolution image (i.e., a stack of high in-plane resolution 2D slices with

large slice thickness). Hence, the key to resolution enhancement in medical imaging is to reconstruct the missing slices between any adjacent HR 2D slices.

So far, various image super-resolution (SR) methods have been proposed by using *either* single LR image *or* multiple LR images. In contrast, Rousseau [1] attempted to use both LR and HR images of the same subject for SR in magnetic resonance (MR) imaging. As an HR 3D T1-weighted MR image is often scanned, it is very attractive to enhance an LR image (e.g., T2-weighted MR image) by learning some prior from its corresponding HR T1 image. A possible choice of such prior is the patchwise self-similarity profile, which describes the representation of an image patch by a set of surrounding patches in the HR image. The underlying assumption is that: if the LR (T2) image and its HR (T1) image lie in the same space, their corresponding points should have a similar self-similarity profile. Thus, the LR (T2) image point can borrow the self-similarity profile computed from the HR (T1) image to recover the missing intensity value by a weighted average of existing intensity values in the LR (T2) image.

Since different imaging modalities measure different tissue properties, the above assumption could fail. For example, the imaging pattern of multiple sclerosis (MS) lesion is clearly visible in the T2-weighted image, but not in the T1-weightd image. Hence, Rousseau [1] proposed to transfer the self-similarity profiles from the HR image to the LR image according to the correlation between the self-similarity profiles separately estimated in the HR and LR images, where the latter has to be estimated from certain initialization. Due to the lack of clear guidance in adapting the self-similarity profiles from HR image, it is still limited in revealing the actual appearances in the LR image. In addition, most of the existing super-resolution methods estimate self-similarity profile at each image point independently, thus leading to inconsistency along the structure boundaries due to the independent estimations.

To overcome all the above limitations, we propose to use the existing image information in the LR image to guide the adaption of self-similarity profile from the HR image. Specifically, we require the estimated patchwise self-similarity profile should satisfy the following conditions: (1) it should best represent each image patch by its nearby patches in the HR image; (2) it should produce the lowest reconstruction error between the existing image appearance in the LR image and the predicted image appearance by adapting the self-similarity profile from HR image; (3) it should be consistent across neighboring slices; and (4) it should be free of initialization. We accordingly propose a novel SR method for jointly recovering all missing intensities for a stack of patches across consecutive slices by solving a group sparse patch representation problem.

We have extensively evaluated our novel SR method on both simulated brain MR images (from Brainweb) and the clinical images with MS lesion. Promising results are achieved, with significant improvement on structural details and image sharpness.

2   Method

Super-Resolution Model in Medical Imaging Scenario. The SR model, widely used in computer vision, assumes that an LR image is a degraded version of the to-be-estimated HR image. Various regularization terms are used to constrain the space of solutions, which inevitably alters the existing intensity values in LR image. In medical imaging field, the task is slightly different since the LR image is usually a stack of HR 2D images. The goal of SR is mainly to increase the resolution across slices. Meanwhile, the estimated HR image is required to preserve the existing intensity values since they are critical in diagnosis and investigation.

In light of this, our proposed SR method is fully data-driven, aiming to achieve M-times SR by reconstructing the $M - 1$ missing slices between every two consecutive slices L_t^0 and L_{t+1}^0 in the LR image $\boldsymbol{L} = \left[L_1^0, \ldots, L_t^0, \ldots, L_N^0\right]$, which has N 2D slices. Hence, the enhanced HR image $\boldsymbol{H}$ is constructed as $\boldsymbol{H} = [[\boldsymbol{H}_1], \ldots, [\boldsymbol{H}_t] \ldots, [\boldsymbol{H}_N]]$, where each slice bundle $[\boldsymbol{H}_t] = \left[L_t^0, H_t^1, \ldots, H_t^{M-1}\right]$ starts with one original slice L_t^0, followed by $M - 1$ recovered slices $\left\{H_t^m | m = 1, \ldots, M-1\right\}$. In our method, there is an HR prior image $\boldsymbol{Y} = [[\boldsymbol{Y}_1], \ldots, [\boldsymbol{Y}_t], \ldots, [\boldsymbol{Y}_N]]$ that can be used to guide the SR, where $[\boldsymbol{Y}_t] = \left[Y_t^0, \ldots, Y_t^{M-1}\right]$ denotes a particular slice bundle of HR prior image. As the HR prior image is acquired from the same subject, it is not difficult to obtain good registration between $\boldsymbol{Y}$ and $\boldsymbol{L}$.

Unsupervised Self-similarity Adaption. Assume that the space of a 2D plane has been divided into the overlapped patches. For an arbitrary patch centered at $v \in \mathbb{R}^2$ in the 2D plane, we use $y_t^m(v)$, a column vector (green solid box in Fig. 1), to denote the intensity values within the patch from slice Y_t^m in the HR prior image. Next, we can estimate the self-similarity profile for $y_t^m(v)$ by using the image patches extracted from a search neighborhood $n(v)$ in slices Y_t^0 and Y_{t+1}^0, which form the HR patch dictionary $\boldsymbol{D} = \{y_{t+\varepsilon}^0(u) | u \in n(v), \varepsilon = 0, 1\}$. Note that we only collect the image patches from the HR prior image at slices Y_t^0 and Y_{t+1}^0, since we will reconstruct the missing intensity values by only using the existing imaging data in slices L_t^0 and L_{t+1}^0 of LR image. Similarly, we can construct another dictionary $\boldsymbol{E}$ by replacing each column in $\boldsymbol{D}$ with the corresponding patch from the LR image. Since the SR procedure is the same at every location v, we drop off the variable v in $y_t^m(v)$ for clarity. Non-local mean technique is used in [1] to determine the self-similarity of each atom in $\boldsymbol{D}$ w.r.t. the target y_t^m. In order to suppress the noisy patches, we go one step further to solve the self-similarity profile w_t^m (a column vector) with sparsity constraint [2, 3] as:

$$\hat{w}_t^m = \arg\min_{w_t^m} \left\|y_t^m - \boldsymbol{D}w_t^m\right\|_2^2 + \lambda\left\|w_t^m\right\|_1, \tag{1}$$

where large value in the weighting vector $\hat{w}_t^m$ suggests high self-similarity between particular atom in $\boldsymbol{D}$ and y_t^m. Since the HR prior image $\boldsymbol{Y}$ is aligned with the LR image $\boldsymbol{L}$, the unsupervised way to recover the image patch h_t^m in LR image domain is to directly apply the self-similarity profile $\hat{w}_t^m$ to the dictionary $\boldsymbol{E}$ by $\hat{h}_t^m = \boldsymbol{E}\hat{w}_t^m$.
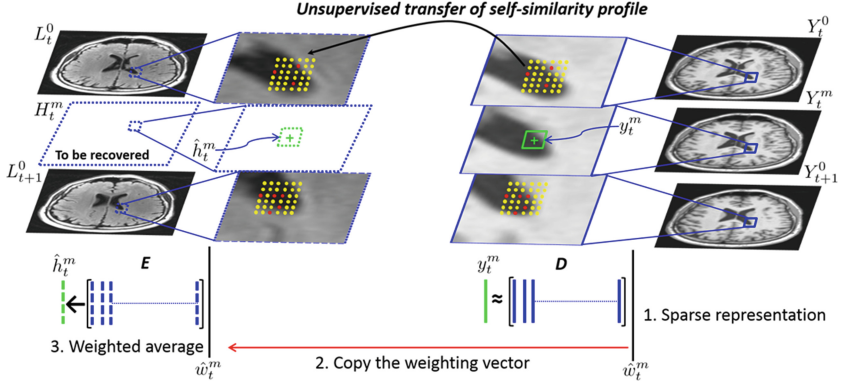
Fig. 1. Unsupervised adaptation of the self-similarity profile from HR image Y to LR image L (Color figure online).

As shown in Fig. 1, dots in the middle right and middle left are the centers of the atoms in D and E, respectively. Due to the sparsity constraint in (1), only a few atoms in D have the non-zero weights (i.e., those red dots), and the rest have zero weights (i.e., yellow dots). The unsupervised adaption of self-similarity profile is to simply copy and apply the weighting vector to E, as shown in the bottom of Fig. 1.

Supervised Self-similarity Adaption by Group Sparse Patch Representation. To overcome the limitations of unsupervised self-similarity adaption, we propose our new supervised approach with the following improvements.

A. Concurrent Self-similarity Profile Computation. Instead of computing the self-similarity profile for each patch independently, we simultaneously determine the self-similarity profiles for a stack of 2D image patches $\mathbf{\Psi} = \left[y_t^0, y_t^1 \ldots, y_t^{M-1}, y_{t+1}^0 \right]$, where they have the same centers in the 2D plane but in consecutive slices. The goal is to jointly estimate the stack of the missing patches $[h_t^m]_{m=1,\ldots M-1}$ between the two existing LR slices L_t^0 and L_{t+1}^0. It is apparent that the patches between two consecutive LR slices L_t^0 and L_{t+1}^0 have similar appearances, although the slice thickness might be large. Thus, it is reasonable to require their self-similarity profiles should be similar as well. Thus, by placing all self-similarity profiles into a weighting matrix $W = \left[w_t^0, w_t^1 \ldots, w_t^{M-1}, w_{t+1}^0 \right]$, the joint sparse patch representation can be formulated by introducing the $L_{2,1}$-norm [4] for requiring each weighting vector in W to have a similar sparsity pattern:

$$\hat{w} = \arg \min_W \|\mathbf{\Psi} - DW\|_2^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_{2,1} \tag{2}$$

where λ_1 and λ_2 are the coefficients balancing the strength of sparsity and the consistency of sparsity pattern.

B. Supervised Self-similarity Adaption. There are numerous possible self-similarity profiles w_t^m in the solution space of (2) if y_t^m is located in the white matter (WM) region of T1-weighted image. This is because most of the patches from WM region are very similar in T1-weighted image, regardless of having MS lesion or not. However, the MS lesions have unique and visible pattern in T2-weighted image. Therefore, clear supervision to adapt the self-similarity profile from HR prior image to the LR image domain is the key to obtain reasonable SR result.

Fortunately, by directly adapting w_t^0 and w_{t+1}^0 to the LR image domain, we can reconstruct the 2D patches $\widehat{h}_t^0 = \boldsymbol{E}w_t^0$ and $\widehat{h}_{t+1}^0 = \boldsymbol{E}w_{t+1}^0$. On the other hand, we have the observed image patches l_t^0 and l_{t+1}^0 at location v from the existing slices L_t^0 and L_{t+1}^0, respectively. Thus, $\widehat{h}_t^0$ and $\widehat{h}_{t+1}^0$ should be same as l_t^0 and l_{t+1}^0. With this guidance, we can supervise the adaption of self-similarity profile by minimizing the residuals, e.g., $\left\|l_t^0 - \boldsymbol{E}w_t^0\right\|_2^2$ and $\left\|l_{t+1}^0 - \boldsymbol{E}w_{t+1}^0\right\|_2^2$. Since we use the $L_{2,1}$-norm to enforce each column vector in $\boldsymbol{W}$ having the similar sparsity pattern, the adaption of w_t^0 and w_{t+1}^0 can be propagated to other self-similarity profiles w_t^m. Hence, we further extend the objective function of joint self-similarity profiles estimation in (2) to the supervised adaption as below:

$$\hat{w} = \arg\,\min_{\boldsymbol{W}}\|\boldsymbol{\Psi} - \boldsymbol{D}w\|_2^2 + \lambda_1\|\boldsymbol{W}\|_1 + \lambda_2\|\boldsymbol{W}\|_{2,1} + \lambda_3 \sum\nolimits_{\varepsilon=0}^1 \left\|l_{t+\varepsilon}^0 - \boldsymbol{E}w_{t+\varepsilon}^0\right\|_2^2, \quad (3)$$

where λ_3 is a coefficient to control the strength of supervision. To optimize (3), we use the $L_{2,1}$ regularized Euclidian projection method in [4, 5].

Figure 2 illustrates 2× super resolution scenario ($M = 2$) of our supervised self-similarity profile adaption. The pink solid boxes denote the patches from the HR prior image, while the pink dash boxes denote the corresponding patches in the LR image. We jointly optimize the self-similarity profiles for the image patches on the HR prior image (pink and green boxes in the right panel of Fig. 2) via the $L_{2,1}$-norm sparse patch representation. Meanwhile, we steer the adaption of the self-similarity profiles towards the LR image domain by minimizing the residuals between the existing and reconstructed image patches (black dash arrows in the bottom of Fig. 2). After we repeat the same procedure to all locations with missing intensity values, we can obtain the resolution enhanced image $\boldsymbol{H}$ from the LR image $\boldsymbol{L}$.

3   Experiments

We evaluate the performance of our new SR method on simulated MR brain image (from Brainweb) and real patient image with MS lesion, using PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) [6] as quantitative measures. We compare the SR performance with both cubic and B-spline interpolations. In addition, we evaluate the role of supervised self-similarity profile adaption by comparing our full method (with the supervision term in Eq. 3) with the degraded method (only using the group sparsity term in Eq. 2).
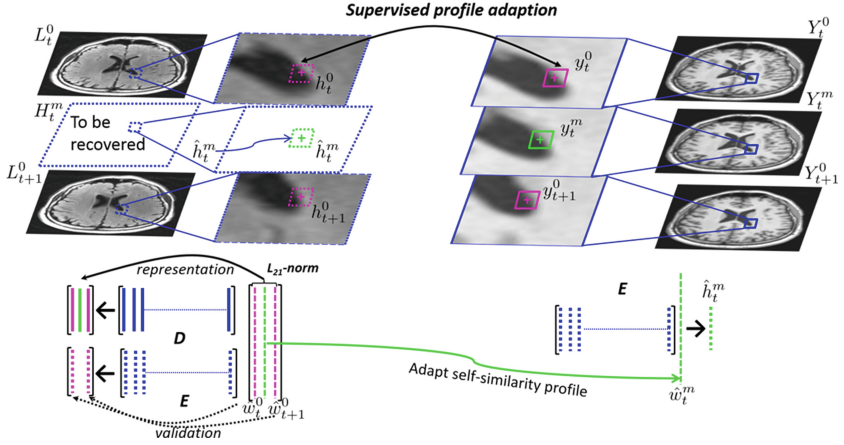
Fig. 2. Overview of the supervised self-similarity adaption (Color figure online).

3.1   Simulated MR Brain Image from Brainweb

Brainweb provides a simulated brain database, which is often used as gold standard for evaluating image enhancement performance. In this experiment, we use the HR T1-weighted image of $181 \times 217 \times 180$ voxels (voxel resolution $1 \times 1 \times 1\,\mathrm{mm}^3$) as the HR prior image. And, the HR T2-weigthed image ($181 \times 217 \times 180$ voxles, with voxel resolution $1 \times 1 \times 1\,\mathrm{mm}^3$) is used as the ground truth to compute the PSNR and SSIM measures. Table 1 shows the PSNR and SSIM scores on the noise free normal images, by cubic interpolation, B-spline interpolation, our degraded method (without supervised self-similarity profile adaption), and our full method. In order to specifically evaluate the advantage of *supervised self-similarity profile adaption* in our method, Table 2 shows the PSNR/SSIM scores computed only in the MS regions (i.e., the blue boxes shown in Fig. 3), instead of the entire brain. In both normal and MS cases, our supervised SR method achieves the highest PSNR and SSIM scores.

**Table 1.** PSNR/SSIM scores of SR methods on the noise free *normal* images.

Thickness	Cubic	B-spline	Degraded method	Our full method
2 mm	27.26/0.9380	27.35/0.9388	36.23/0.9682	**36.96/0.9802**
3 mm	22.21/0.8945	22.74/0.9015	30.18/0.9518	**32.83/0.9735**
5 mm	20.68/0.7692	20.93/0.7699	26.13/0.8126	**26.54/0.8163**
7 mm	18.31/0.7292	18.41/0.7305	24.36/0.7584	**24.89/0.7589**
9 mm	16.96/0.7024	17.11/0.7040	22.41/0.7306	**22.77/0.7353**

The reconstructed HR images by four methods are displayed in Fig. 3, with the normal image shown in the top panel while the MS images shown in the bottom panel. It is obvious that our method reveals more anatomical details than all other three methods, as indicated by the red arrows in normal image group and the blue boxes in

Table 2. PSNR/SSIM scores (only in MS region) of SR methods on the noise free *MS* images.

Thickness	Cubic	B-spline	Degraded method	Our full method
2 mm	26.74/0.8945	26.83/0.8949	35.97/0.9635	**36.25/0.9662**
3 mm	21.59/0.8831	27.36/0.9262	29.65/0.9438	**32.17/0.9637**
5 mm	20.05/0.7506	20.15/0.7510	25.43/0.8037	**25.83/0.8100**
7 mm	17.68/0.7161	17.83/0.7167	23.54/0.7416	**23.81/0.7422**
9 mm	16.16/0.7005	16.26/0.7011	21.86/0.7264	**22.05/0.7304**

MS lesion group. Since our full method provides supervision on adaption of self-similarity profile, the MS regions in the reconstructed T2-weighted image are much closer to gold standard than our degraded methods, as shown by the zoom-in views in the blue dash boxes in Fig. 3.

(a) Ground truth   (b) Cubic interp.   (c) B-spline interp.   (d) Our degraded method   (e) Our full method
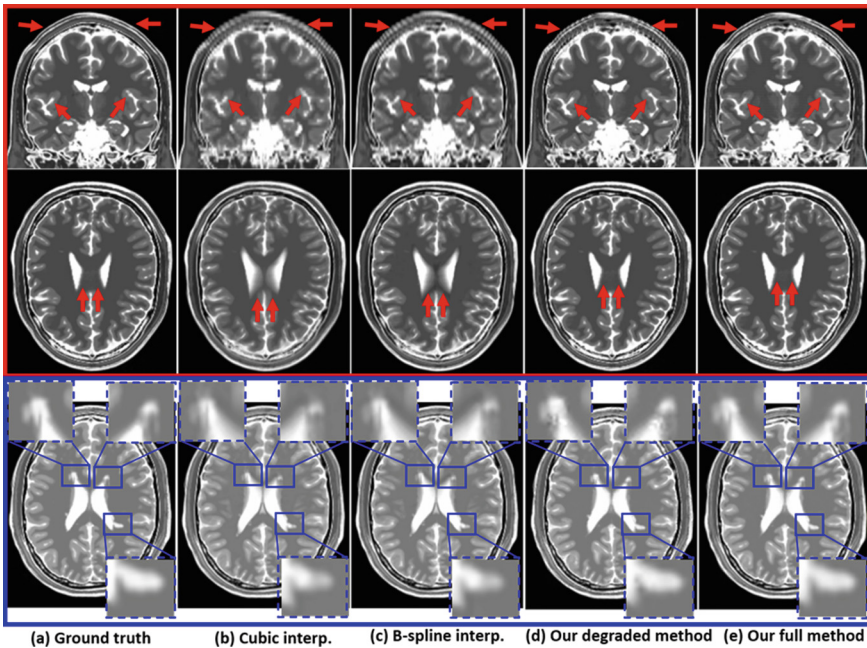
Fig. 3. The reconstructed HR T2-weighted images by four SR methods (Color figure online)

3.2   Real Patient MR Brain Image with MS Lesion

We repeat the experiment on 12 T2-weighted MR images of MS patients with slice thickness 3 mm. The T1-wighted prior image for each patient reaches 1 mm slice thickness. To quantitatively evaluate the SR performance, we first generate the low-resolution images from 3 mm to 6 mm thickness by discarding the odd number slices. Then we apply 2× super resolution using four different SR methods to reconstruct back to 3 mm thickness. Table 3 shows the averaged PSNR and SSIM scores

Table 3. Performance (PSNR/SSIM) of SR methods on MS patients.

6 mm → 3 mm	Cubic	B-spline	Degraded method	Our full method
Whole brain	15.25/0.6532	15.49/0.6539	19.35/0.7116	**20.77/0.7206**
MS region only	17.43/0.7069	17.53/0.7077	22.16/0.7328	**23.26/0.7416**

(in whole brain and MS regions, respectively) by cubic interpolation, B-spline interpolation, our degraded method, and our full method.

Second, we enhance the original T2-weighted image from 3 mm to 1 mm. The reconstructed HR images by four SR methods are shown in Fig. 4. Through visual inspection, the reconstructed images by our full SR method have better image quality (i.e., anatomical details and sharpness) than other three methods.

(a) Cubic interpolation   (b) B-spline interpolation   (c) Our degraded method   (d) Our full method
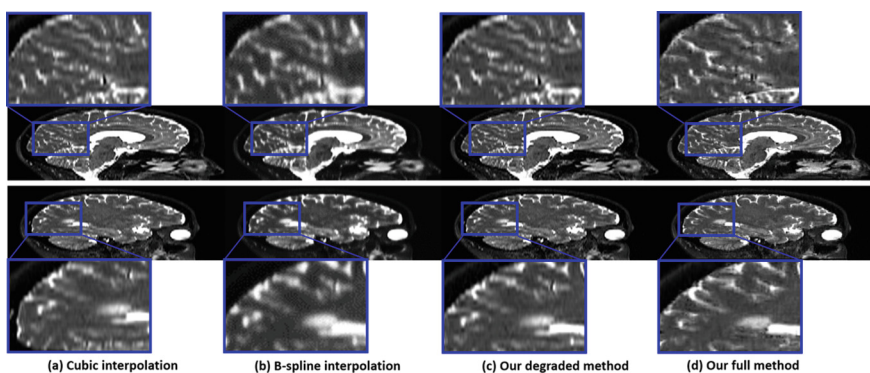
Fig. 4. Visualization of resolution enhancement on MS patient data by four SR methods.

4   Conclusion

We have developed a novel super-resolution method that can reconstruct the missing slices of a given image for resolution enhancement. This is achieved by adapting the self-similarity profiles from the HR prior image, using the LR image to supervise the adaption procedure via group sparse patch representation. Promising super resolution results have been achieved on both simulated and real patient data, which demonstrate its wide possible applications in various clinical studies.

References

1. Rousseau, F.: A non-local approach for image super-resolution using intermodality priors. Med. Image Anal. **14**, 594–605 (2010)
2. Tong, T., Wolz, R., Coupé, P., Hajnal, J., Rueckert, D.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage **76**, 11–23 (2013)

3. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B **58**(1), 267–288 (1996)
4. Liu, J., Ji, S., Ye, J.: Multi-task feature learning via efficient L2,1-norm minimization. In: Proceeding of the 25th Conference on Uncertainty in Artificial Intelligence, Montreal, Canada (2012)
5. Liu, J., Ji, S., Ye, J.: SLEP: sparse learning with efficient projections. Arizona State University (2009)
6. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)