Combined Self-Learning Based Single-Image Super-Resolution and Dual-Tree Complex Wavelet Transform Denoising for Medical Images

Guang Yang *[a,b], Xujiong Ye [c], Greg Slabaugh [d],
Jennifer Keegan [a,b], Raad Mohiaddin [a,b], and David Firmin [a,b]
[a] Cardiovascular Biomedical Research Unit, Royal Brompton Hospital, SW3 6NP, London, UK;
[b] National Heart & Lung Institute, Imperial College London, SW7 2AZ, London, UK;
[c] School of Computer Science, University of Lincoln, LN6 7TS, Lincoln, UK;
[d] Department of Computer Science, City University London, EC1V 0HB, London, UK

ABSTRACT

In this paper, we propose a novel self-learning based single-image super-resolution (SR) method, which is coupled with dual-tree complex wavelet transform (DTCWT) based denoising to better recover high-resolution (HR) medical images. Unlike previous methods, this self-learning based SR approach enables us to reconstruct HR medical images from a single low-resolution (LR) image without extra training on HR image datasets in advance. The relationships between the given image and its scaled down versions are modeled using support vector regression with sparse coding and dictionary learning, without explicitly assuming reoccurrence or self-similarity across image scales. In addition, we perform DTCWT based denoising to initialize the HR images at each scale instead of simple bicubic interpolation. We evaluate our method on a variety of medical images. Both quantitative and qualitative results show that the proposed approach outperforms bicubic interpolation and state-of-the-art single-image SR methods while effectively removing noise.

Keywords: Self-learning, sparse representation, super-resolution, denoising, discrete wavelet transform, dual-tree complex wavelet transform, medical imaging analysis, image processing

1. INTRODUCTION

Images with high resolution are desirable in many real applications, e.g., medical image analysis, video surveillance and others [1]. However, image resolution and quality are often limited by the hardware of the imaging acquisition system, economic and health costs of the imaging procedure, and the acquisition time. In particular, medical imaging systems acquire useful clinical details about the anatomical, physiological, functional and metabolic information of patients using various imaging modalities. Despite recent advances in imaging hardware development and imaging protocol improvement, images are always obtained with degraded quality because of the inherent noise of the medical imaging system. Moreover, aforementioned health limitations (e.g., ionizing radiation dose) and acquisition time limitations (e.g., Specific Absorption Rate limits) restrict the image resolution, thus may reduce the visibility of vital pathological details and compromise the diagnostic accuracy and prognosis.

Instead of optimizing hardware settings and imaging sequences, image super-resolution (SR) provides an alternative solution to improve the perceptual quality of medical images in terms of the spatial resolution enhancement. Essentially, the goal of SR methods is to recover a high-resolution (HR) image from a single or multiple low-resolution (LR) images [2]. Existing SR algorithms can be broadly categorized into interpolation based, reconstruction based, and learning based methods [3]. Widely used interpolation based SR methods assume that images are spatially smooth; however, this assumption is usually inaccurate and results in overly smoothed edges with ringing and jagged artifacts [3,4]. Reconstruction based algorithms solve an inverse problem by recovering the HR image by fusing multiple LR images [3]. However, this approach is time-consuming and infeasible due to multiple LR images are required [4], and are numerically limited to an upsampling factor of two [5]. For learning based methods, the mapping function between LR and HR images (or their patches) is learned from a representative set of training image pairs. Once the mapping function is learned, it is applied to a single testing image to achieve SR. Despite numerous learning based methods [1,6,7] have claimed success for single-image SR, these methods hinge on certain amount of training data of LR and HR image pairs. Glasner et al. [2] proposed a self-similarity based single-image SR method using a training dataset that is directly established from the LR

* Asterisk indicates corresponding author. Send correspondence to G.Y., g.yang@imperial.ac.uk

input, by exploiting the patch redundancy among in-scale and cross-scale images in an image pyramid to enforce constraints for recovering the unknown HR image [3]. Obviously, the advantage of this method is that there is no need to construct an extrinsic large training dataset beforehand; however, the abundance of self-similar patches is questionable for medical images. Yang and Wang [8] circumvented this problem by applying support vector regression (SVR) that learned proper SR models from patches at different image scales instead of searching for similar image patches. In their study, sparse representation based feature extraction was applied to make their SR algorithm more computationally feasible for real applications that was originally proposed for solving SR problem by Yang et al. [9]. More recently, Singh and Ahuja [10] added sub-band energy constraints for the self-similarity based SR method, and Huang et al. [11] enriched the dataset of self-similar patches by looking at their transformed exemplars. Although these studies resulted in promising quantitative results, the SR methods were applied on relatively clean natural images. Therefore, the performances of these methods for noisy medical images are still questionable. Comprehensive reviews on various SR methods can be found elsewhere [12,13].

In this paper, we propose an SVR based SR method that learns a sparse representation as robust and effective features across various down-sampled and denoised versions (using dual-tree complex wavelet transform, i.e., DTCWT) of a single input LR image. Once the SVR has been trained, we can apply the best model to predict the final SR image from a single LR image. Compared to the previous published learning based methods [1,6,7], our approach does not require the extra collection of HR training datasets. Compared to Glasner et al. [2] and more recent work by Singh and Ahuja [10], our method does not make assumption about self-similarity of image patches. In contrast to the SR methods [8,14], in which SVR was originally applied, we use DTCWT to suppress the noise, and we hypothesize that our method is more applicable for medical imaging applications, in which inherent noises are inevitable due to the limitations of the imaging systems and acquisition protocols. Quantitative and qualitative results show that our method outperforms bicubic interpolation and a self-learning based single-image SR method [8] while effectively removing noise.

2. METHODOLOGY

The overall framework of our method is summarized as shown in the flowchart in Figure 1. Detailed steps of our method are described below.

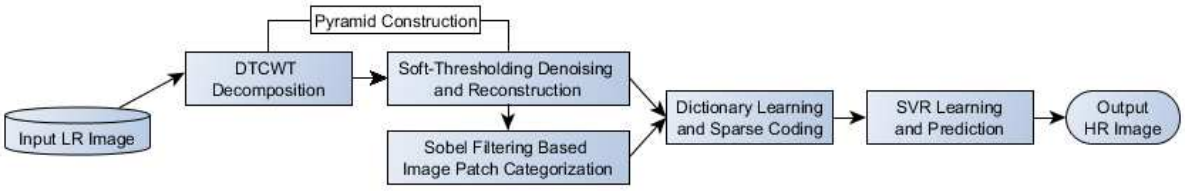

Figure 1: Flowchart of our approach including both learning and prediction (For DTCWT, 4 layers of decomposition was used).

2.1 Denoising and Pyramid Construction Using Dual-Tree Complex Wavelet Transform (DTCWT)

The widely applied discrete wavelet transform, which replaces the infinitely oscillating sinusoidal basis function of the Fourier transform with a set of locally oscillating basis functions, has some known limitations, e.g., shift variance and lack of directionality [15]. The dual-tree complex wavelet transform (DTCWT) overcomes shortcomings of the DWT such that it is nearly shift invariant and directionally selective in two and higher dimensions [15]. A detailed review of DTCWT can be found in [15]. Instead of using bicubic interpolated down-sampled versions of the original image to form the image pyramid [8], we decomposed the original image using DTCWT and each level was reconstructed using a soft-thresholding scheme [16] to accomplish the denoising and image pyramid construction, which we used to train the SVR.

2.2 Feature Extraction Using Image Sparse Representation

Instead of working directly with pixels, a sparse image representation [17] was learned, representing robust and effective features for SVR that can be formulated as an optimization problem,

$$\min \frac{1}{2}\|D\alpha - x\|_2^2 + \lambda\|\alpha\|_1 , \quad (1)$$

in which x is the image patch (in our study we used a patch size of 5 by 5), D is the over-complete dictionary to be learned, α is the corresponding sparse coefficient vector, and λ is the regularization term that balances the model sparsity

(l_1-norm term) and the l_2-norm based residual. According to Yang and Wang [8], better SR performance can be achieved by patch categorization, i.e., clustering patches into low and high spatial frequency ones and learning their dictionaries separately. In our study, this has been done using a Sobel filter, such that any patch including an edge derived from the Sobel filter is considered to have high spatial frequency.

2.3 Support Vector Regression (SVR)

Finally, SVR [18], which can fit the data in a high dimensional feature space without assuming the data distribution, was applied to model the relationship between the input sparse coefficient vector α and the associated SR pixel value. In training, our SVR solves the following problem

$$\min \frac{1}{2}\omega^T\omega + \rho\left(\frac{1}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*)\right) \quad (2)$$

$$\text{subject to } y_i - \langle\omega, \phi(\alpha_i)\rangle - b \le \epsilon + \zeta_i,$$
$$\langle\omega, \phi(\alpha_i)\rangle + b - y_i \le \epsilon + \zeta_i^*,$$
$$\zeta_i, \zeta_i^* \ge 0, \epsilon \ge 0, i = 1, \dots, n,$$

in which y_i denotes the associated target variable, i.e., pixel value at the center of the patch considered in the HR image. $\phi(\alpha_i)$ is the sparse image patch features in the transformed space. And ω is the norm vector of the nonlinear mapping function to be learned, and ρ is the tradeoff between the generalization and the upper and lower bounds of training errors ζ_i and ζ_i^* subject to a threshold of ϵ. Gaussian kernels were used, and parameters were estimated via cross-validation. SVR model selection was achieved according to a minimum-error-rate classification rule based on Bayesian decision theory [8], and the trained best SVR model was applied to predict the final SR output for a test LR input.

3. EXPERIMENTS AND RESULTS

Experiments have been performed on relatively clean HR ankle and knee MRI images, from which LR inputs were synthesized with additive Gaussian noises ($\sigma = 10, 15, 20, \text{and } 25$) for the sake of simplicity. For these datasets, we have ground truth of the HR images without noise. In addition, we also tested our approach on noise corrupted LR clinical datasets, i.e., LR cardiac and brain MRI images (Table 1), but for which no ground truth was available. Evaluations have been done qualitatively by visual inspection and quantitatively using peak signal to noise ratio (PSNR) and mean squared error (MSE) that are defined as

$$\text{PSNR} = 20\ \log_{10}\frac{255}{\sqrt{\text{MSE}}} \quad (3)$$

where MSE is calculated by

$$\text{MSE} = \frac{1}{\text{MN}}\sum_{n=1}^{\text{M}}\sum_{m=1}^{\text{N}}[\widehat{\text{X}}(n,m) - \text{X}(n,m)]^2. \quad (4)$$

The smaller the MSE, the closer the super-resolved result ($\widehat{\text{X}}$) is to the ground truth (X).

The performance of our method was compared with a conventional Gaussian filter + bicubic interpolation and the self-learning based SR approach (SLSR) proposed in [8]. Example SR results are shown in Figure 2–5. PSNR and MSE results of ankle and knee MRI images with HR ground truth, and various additive noises [($\sigma = 10, 15, 20, \text{and } 25$) and run 10 times for each noise level] are shown in Figure 6.

Table 1: Experimental data.

Data	Original Resolution	Input Resolution	Output Resolution	Ground Truth
Ankle MRI	400x400	200x200	400x400	Yes
Knee MRI	400x400	200x200	400x400	Yes
Cardiac MRI	256x256	256x256	512x512	No
Brain MRI	256x256	256x256	512x512	No

4. CONCLUSION

In this study, we proposed a self-learning based SR approach, which coupled DTCWT based denoising and SVR on image sparse representation based features. Both qualitative visual appearances and quantitative results have

demonstrated that our method can achieve promising single-image SR compared to conventional bicubic interpolation and recently developed SLSR methods.

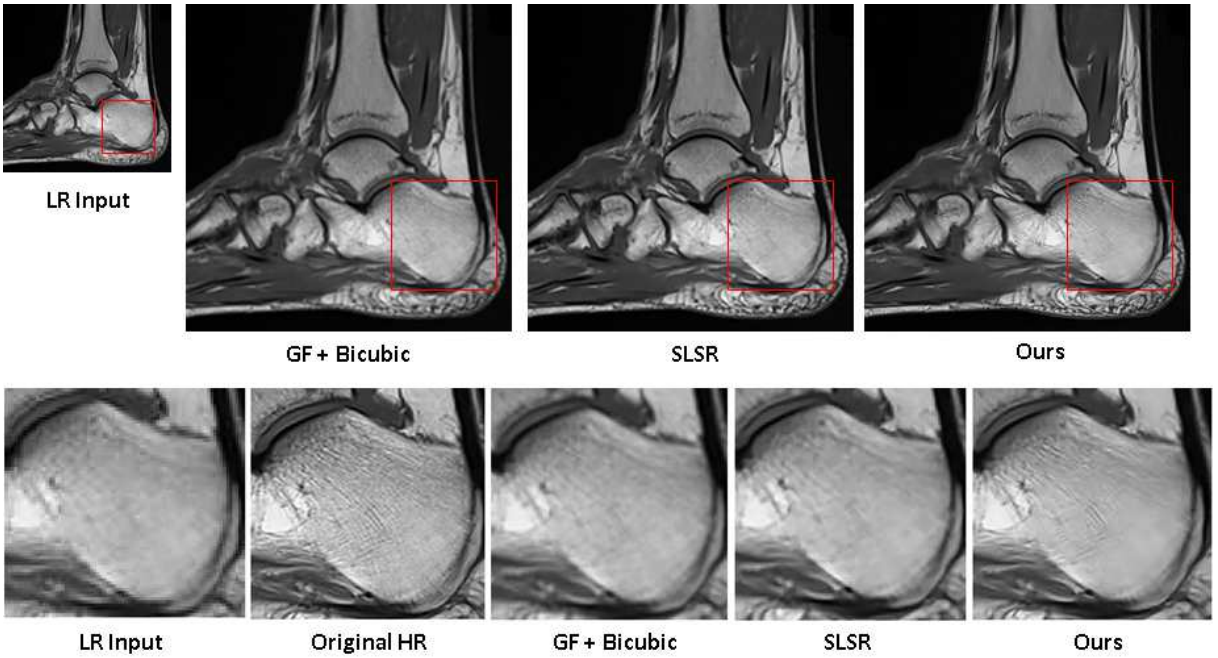

Figure 2: First row from left to right: LR input, Gaussian filtering (GF) and bicubic interpolation result, SLSR result, and result of our proposed method for the ankle MRI image. Second row from left to right: Zoomed in version (red square region) of the LR input (nearest neighbor interpolated for displaying purpose), original HR ground truth, GF + bicubic interpolation result, SLSR result, and result of our proposed method.

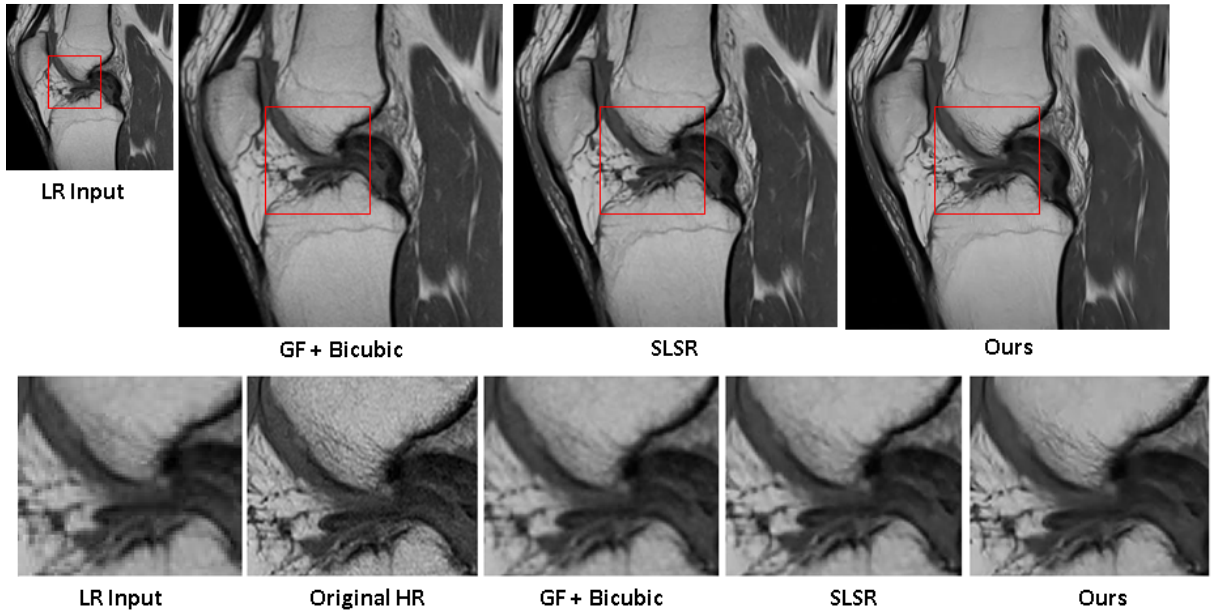

Figure 3: Similar to Figure 2 but applied to the knee MRI image.

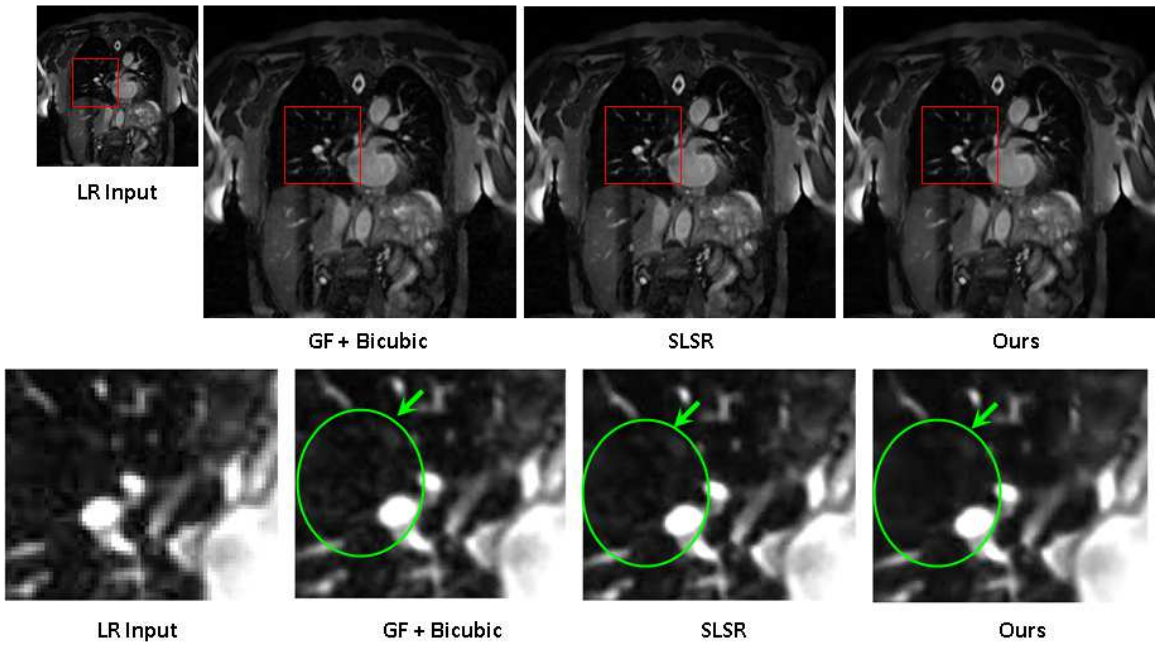

Figure 4: Similar to Figure 2 but applied to the cardiac MRI image (note that there is no HR ground truth for this case). Contrast has been altered for the zoomed-in results for better presentation. Compared to GF + bicubic interpolation and SLSR results, our method obtained much less confounded structures that may attribute to the partial volume effects from the slices above and below (green arrow pointed circle region).

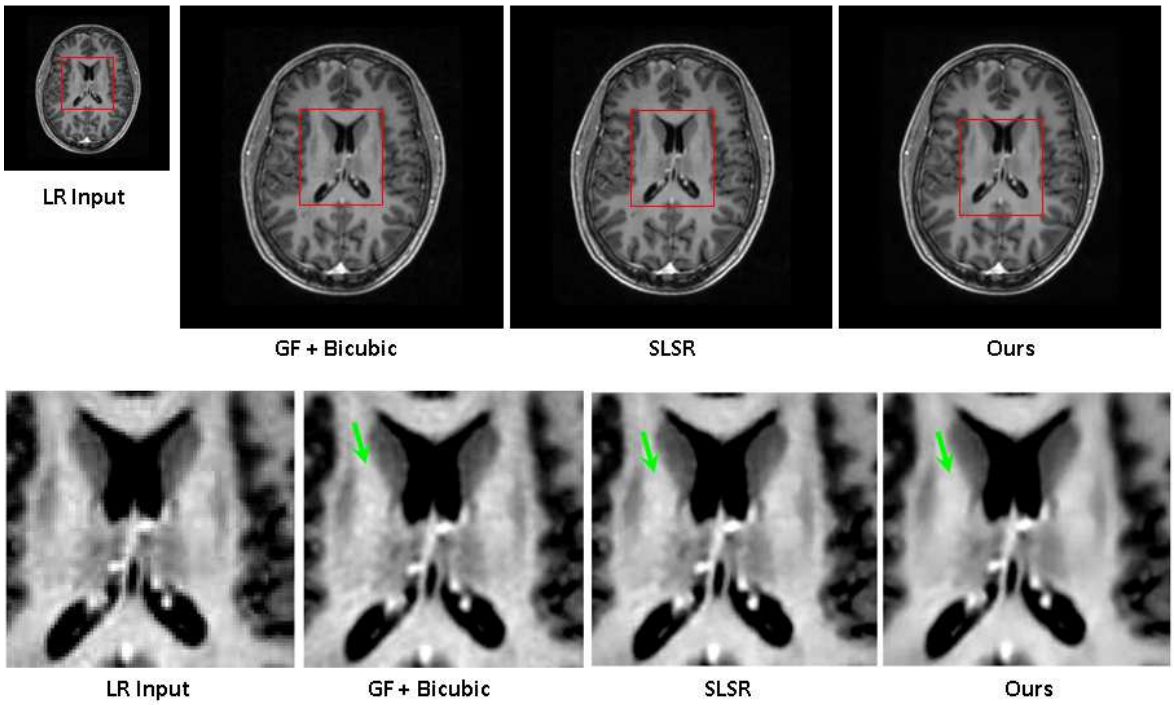

Figure 5: Similar to Figure 2 but applied to the brain MRI image (note that there is no HR ground truth for this case). Contrast has been altered for the zoomed-in results for better presentation. Compared to GF + bicubic interpolation and SLSR results, our method obtained a more homogenous result, e.g., green arrow pointed white matter region.

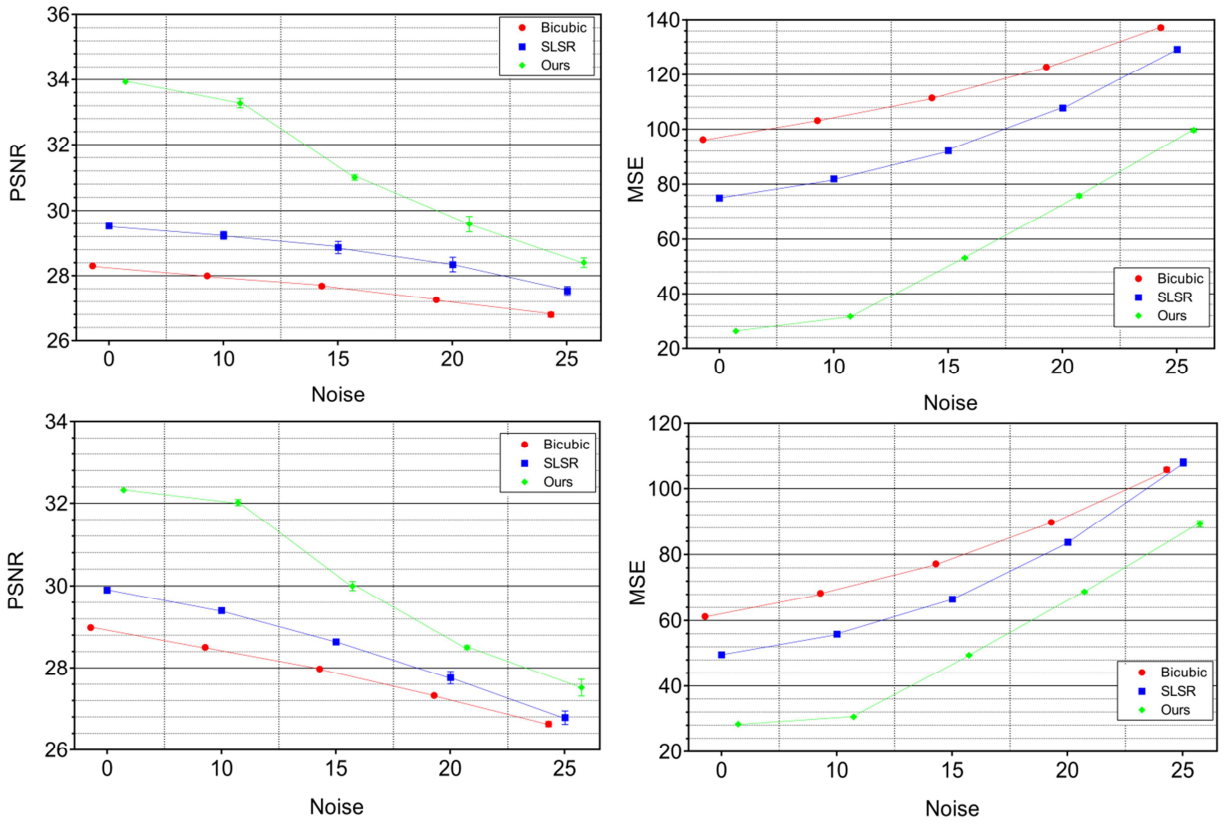

Figure 6: PSNR (dB) and MSE for the ankle (first row) and knee (second row) MRI images with HR ground truth, and the results with additive noises [($\sigma = 10, 15, 20,$ and 25) and run 10 times for each noise level to obtain the error bars].

REFERENCES

[1] Trinh, D. H., Luong, M., Dibos, F., Rocchisani, J. M., Pham, C. D.., Nguyen, T. Q., "Novel example-based method for super-resolution and denoising of medical images," IEEE Trans. Image Process. 23(4), 1882–1895 (2014).

[2] Glasner, D., Bagon, S.., Irani, M., "Super-resolution from a single image," Proc. IEEE Int. Conf. Comput. Vis., 349–356 (2009).

[3] Yang, C. Y., Huang, J. Bin., Yang, M. H., "Exploiting self-similarities for single frame super-resolution," Comput. Vis. – ACCV 2010 6494 LNCS(PART 3), 497–510 (2011).

[4] Kang, L., Zhuang, B., Hsu, C., Lin, C.., Yeh, C., "Self-Learning-Based Low-Quality Single Image Super-Resolution," IEEE MMSP 1(i) (2013).

[5] Lin, Z.C.., Shum, H.Y., "Fundamental limits of reconstruction-based superresolution algorithms under local translation," IEEE Trans. Pattern Anal. Mach. Intell. 26(1), 83–97 (2004).

[6] Chang, H., Yeung, D.-Y.., Xiong, Y., "Super-resolution through neighbor embedding," Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, 2004. CVPR 2004. 1 (2004).

[7] Yang, J., Wang, Z., Lin, Z., Cohen, S.., Huang, T., "Coupled dictionary training for image super-resolution," IEEE Trans. Image Process. 21(8), 3467–3478 (2012).

[8] Yang, M. C.., Wang, Y. C. F., "A Self-Learning Approach to Single Image Super-Resolution," Multimedia, IEEE Trans. 15(3), 498–508 (2013).

[9] Yang, J., Wright, J., Huang, T. S.., Ma, Y., “Image Super-Resolution Via Sparse Representation,” IEEE Trans. Image Process. 19(11), 2861–2873 (2010).
[10] Singh, A.., Ahuja, N., “Sub-band Energy Constraints for Self-Similarity Based Super-resolution,” 2014 22nd Int. Conf. Pattern Recognit., 4447–4452 (2014).
[11] Huang, J., Singh, A.., Ahuja, N., “Single Image Super-resolution from Transformed Self-Exemplars,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 5197–5206 (2015).
[12] Nasrollahi, K.., Moeslund, T. B., “Super-resolution: A comprehensive survey,” Mach. Vis. Appl. 25(6), 1423–1468 (2014).
[13] Yang, C.-Y., Ma, C.., Yang, M., “Single-Image Super-Resolution: A Benchmark,” Comput. Vis. – ECCV 2014, 372–386 (2014).
[14] Ni, K. S.., Nguyen, T. Q., “Image superresolution using support vector regression.,” IEEE Trans. Image Process. 16(6), 1596–1610 (2007).
[15] Selesnick, I. W., Baraniuk, R. G.., Kingsbury, N. G., “The Dual-Tree Complex Wavelet Transform,” Signal Process. Mag. IEEE 22(6), 123–151 (2005).
[16] Cohen, R., “Signal Denoising Using Wavelets,” Technical Report, Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa (2012).
[17] Mairal, J., Bach, F., Ponce, J.., Sapiro, G., “Online Learning for Matrix Factorization and Sparse Coding,” J. Mach. Learn. Res. 11, 19–60 (2010).
[18] Chang, C.., Lin, C., “LIBSVM : A Library for Support Vector Machines,” ACM Trans. Intell. Syst. Technol. 2(5), 1–39 (2011).