SINGLE IMAGE SUPER RESOLUTION THROUGH SPARSE REPRESENTATION

Harsh Sinha (14265)

Department of Aerospace Engineering
IIT Kanpur

Satya Prakash Panuganti (14610)

Department of Electrical Engineering
IIT Kanpur

ABSTRACT

Image Super Resolution task has been widely studied in the realm of Image Processing. This task has been performed through various methods and one of those is the Dictionary Learning based method, which is used specifically for Single Image Super Resolution. In the following report we explore various techniques for Single Image Super Resolution with Dictionary Learning in the Wavelet Domain. We analyze the effect of the choice of different wavelet for this task and also preform a comparison between different methods for dictionary learning. We then also propose a simple CNN based architecture to see if wavelet domain might perform well in the Deep Learning setting as well.

Index Terms— Single Image Super Resolution, Wavelets

1. INTRODUCTION

The aim of Super Resolution(SR) is to recover a high resolution(HR) image from one or several low resolution(LR) images. SR was approached a decade ago primarily by two methods : (i) Classical Multi-Image SR, in which a set of low-resolution images of the same scene are used as input. [1, 2, 3] (ii) Example-Based SR, in which correspondences between LR and HR image patches are learned from a database of LR and HR images, was introduced by [4, 5, 6].

In Classical SR, the HR images recovered are guaranteed to be the true HR images provided enough LR images are available of the scene. In Example-Based SR, also termed as Single Image Super-Resolution(SISR), the HR images recovered are 'hallucinated' and are not guaranteed to be the true images.

Through various works it has been proved that sparse learned dictionaries preform well for many image processing tasks. The reason for this stems form the fact that in learning a sparse representation, one also ends up learning (and extracting) salient features of the image. In many recent works it was shown that the sparse dictionary learning methods also perform well on the Super Resolution tasks. Wavelets have also been thoroughly studied and explored in the past and have been applied to the field of image super resolution recently.

In this report we explore application of sparse learning in wavelet domain for singe image super resolution. We begin with a hypothesis that the choice of wavelet is important for effective application and proceed to show this experimentally and present an in-depth comparison of different wavelets for SISR. We then also explore the application of wavelets in Deep Learning with a simple and novel method based on Convolutional Neural Networks.

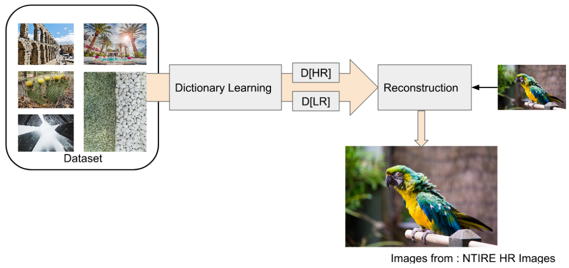

Fig. 1. General framework of SISR

2. RELATED WORK

Aharon et al, developed K-SVD, an algorithm to obtain a dictionary best representng a set of training signals under sparsity constraints in 2006. K-SVD has become a standard for dictionary learning tasks where it is used in conjunction with any pursuit method such as OMP [7] and FOCUSS[8].

Ophir et al. proposed a method in which multi-scale dictionaries were learned using wavelets in [9]. Using wavelets is shown to better capture the image features. In [10], Nazzal et al. were the first to use to wavelet domain dictionary learning for SISR using K-SVD.

Sparse approach to SISR was proposed in [11]. Yang et al. introduced the coupled-dictionary learning paradigm in [12].

In [13], Ahmed et al. use coupled dictionary learning building upon the work of [10], to take advantage of the directionality of wavelet sub-bands and the persistence of wavelet coefficients across scale.

Techniques such as semi-coupled dictionary learning [14], have also been used for SISR. Dictionary learning in wavelet domain for SISR is relatively a new field.

3. COUPLED K-SVD IN WAVELET DOMAIN FOR SISR

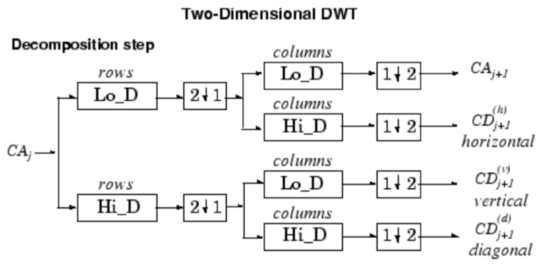

Fig. 2. 2D Discrete Wavelet Transform (Image obtained from Mathworks Website)

In sparse representation based SISR, a HR patch x_H is sparsely coded over a HR dictionary D_H and an LR patch x_L is sparsely coded over an LR dictionary D_L :

$$x_H \approx D_H \alpha_H \tag{1}$$

$$x_L \approx D_L \alpha_L \tag{2}$$

where α_H and α_L are the coefficient vectors for x_H and x_L respectively.

Representing the conversion of a HR image into an LR image by the operator, Ψ, and enforcing that the HR and LR dictionaries are coupled, we get :

$${}_H = \Psi D_L \tag{3}$$

$$x_L = \Psi x_H \approx \Psi D_H \alpha_H \approx D_L \alpha_H \tag{4}$$

$$\implies x_H \approx x_L \tag{5}$$

Hence, we get can reconstruct the HR patches by the following means :

$$x_H = D_H \alpha_H \approx D_H \alpha_L \tag{6}$$

The idea of using LR as the HR coefficients when the dictionaries are coupled together is used in [12, 15].

The overview of the entire algorithm is shown in 3 and 5.

3.1. Dictionary Training

A dataset containing HR images is taken as input. A two-level DWT (Discrete Wavelet Transform) decomposition is performed to generate the detail sub-bands. I_{Hh}, I_{Hv} and I_{Hd} are the horizontal, vertical and diagonal sub-bands of HR subbands and I_{Lh}, I_{Lv} and I_{Ld} are the horizontal, vertical and diagonal sub-bands of LR sub-bands. Wavelet interpolation is performed on the LR sub-bands to increase their size and match with the size of the HR image. The sub-bands obtained are termed as the Mid Resolution (MR) sub-bands and denoted by I_{Mh}, I_{Mv} and I_{Md}. For each of the detail sub-bands ($y = \{h, v, d\}$), samples of HR/LR pairs are obtained and vectorized to form the training data matrices W_{Hy} and W_{My}. The dictionaries are solved through the following optimization problem :

$$\underset{D_{Ly}, D_{Hy}}{\operatorname{argmin}} \Sigma_{i=1}^{N} ||W_{My}(i) - D_{Ly}\alpha_y(i)||_2^2 + ||W_{Hy}(i) - D_{Hy}\alpha_y(i)||_2^2$$
$$s.t.\ ||\alpha_y(i)||_0 \leq T_0, ||D_{Ly}(k)||_2 \leq 1, ||D_{Ly}(k)||_2 \leq 1,$$
$$k = 1, 2, ..., K$$

where K is the dictionary size, $W_{Hy}(i)$ and $W_{My}(i)$ are i^{th} training vector of the HR and LR dictionaries, $D_{Hy}(k)$ and $D_{Ly}(k)$ are the k^{th} atoms for the HR and LR dictionaries. T_0 is the level of sparsity. The entire dictionary learning algorithm is shown in 3.

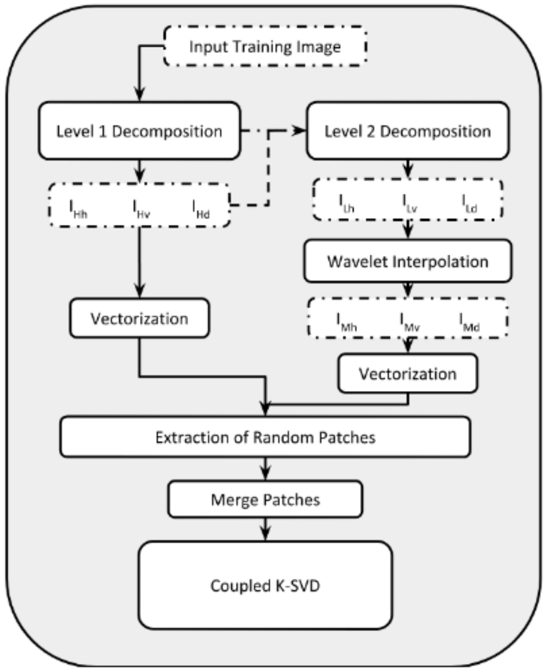

Fig. 3. Dictionary Training Algorithm

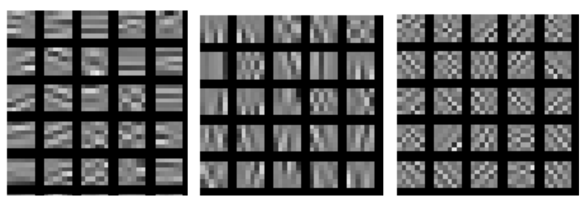

Fig. 4. The first few atoms for LR horizontal, vertical and diagonal dictionaries[From left to right]

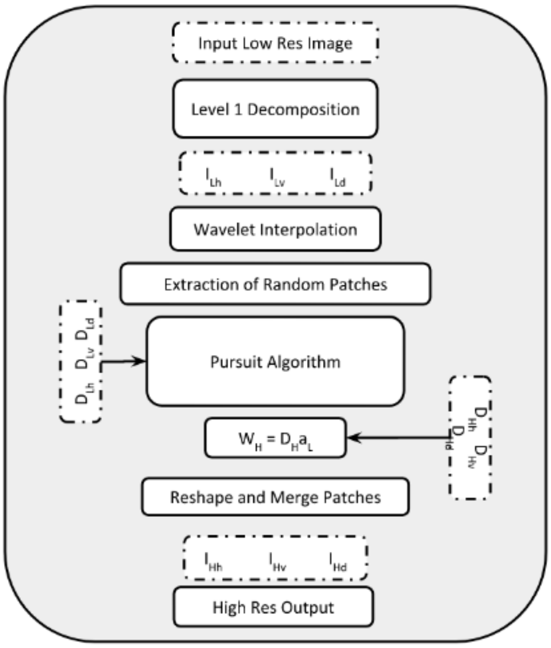

Fig. 5. Image Reconstruction Algorithm

3.2. Image Reconstruction

The algorithm reconstructs the HR image by estimating the HR detail wavelet sub-bands of the image. The LR image is assumed to be the approximation image of the wavelet transform of HR image. This assumption holds since the operation of blurring is similar to the low pass scaling filter of wavelet transform. A 1-level forward DWT of the LR image is obtained. Wavelet interpolation is performed for each detail wavelet detail sub-band to make them have the same size as that of the sub-bands of HR image which is to be reconstructed. The sparse coefficients of patches extracted from each sub-band is calculated by solving the optimization problem :

$$\underset{\alpha_{Ly}}{\operatorname{argmin}} ||W_{My} - D_{Ly}\alpha_{Ly}||_2 \; s.t. \; ||\alpha_{Ly}||_0 \leq T_0$$

The above basis pursuit problem can be solved by a number of ways. Orthogonal Matching Pursuit(OMP) [7] was chosen for the above task. Using the calculated sparse representation coefficients and the corresponding HR dictionaries the wavelet sub-bands of the HR images are then estimated :

$$W_{Hy} = D_{Hy}\alpha_{Hy} \approx D_{Hy}\alpha_{Ly} \tag{7}$$

.

The patches are reshaped into 2D patches and merged to construct the full wavelet detail sub-band images. An Inverse DWT is then performed using the recovered detail sub-bands and using the LR image as the approximation sub-band image to obtain the HR image. The entire dictionary learning algorithm is shown in 5.

4. THE ISSUE WITH *SYM29*

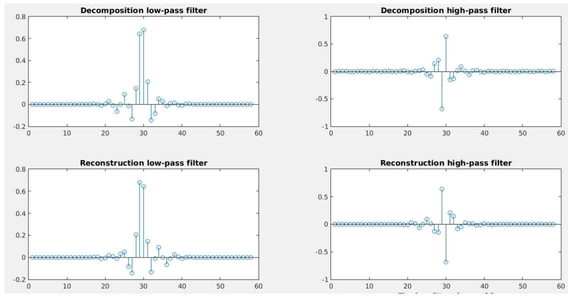

Fig. 6. The Decomposition and Reconstruction, low and high pass filters for 'sym29'

In [13], the authors have used symlet wavelet of order 29 and have treated the borders with periodic extensions. They have referred to Nazzal et. al's work [10] for the idea to use this specific wavelet. Now, Nazzal et al, do not give any specific reasons for the use of this wavelet and cite [16] for the inspiration.

Since different wavelets have quite different properties, the effect of wavelet choice should be significant. With this hypothesis we will analyze the effect of different wavelets on the Super Resolution task in the same setting of [13].

5. EXPERIMENTATION

5.1. Comparison of different Wavelets

Three well-known prominent types of wavelets were chosen for comparison: (i) Daubechies (ii) Symplets (iii) Coiflets. All threes families of orthogonal wavelets were designed by Ingrid Daubechies. For the 3D surface plots, the x-axis is the sparsity chosen for the dictionary learning and reconstruction and y-axis is the wavelet number.

Symlet family of wavelets which are a modified version of Daubechies wavelets with increased symmetry appear to be the best for the task of SISR. The authors of this work are unable to find a theoretical explanation for the observed results.

The following observations can be made empirically:

- Sparsity has a weak effect. SNR and SSIM fall with increase in sparsity. One possible explanation why good results are being obtained at low sparsity is that the patches from the wavelet transformed images tend to be sparse resulting in fewer atoms being needed to represent the sampled patches.
- Choice of Wavelet is important to the task of SISR.
- Symlets with wave number in the form of 4N + 3 are a poor choice for SISR.

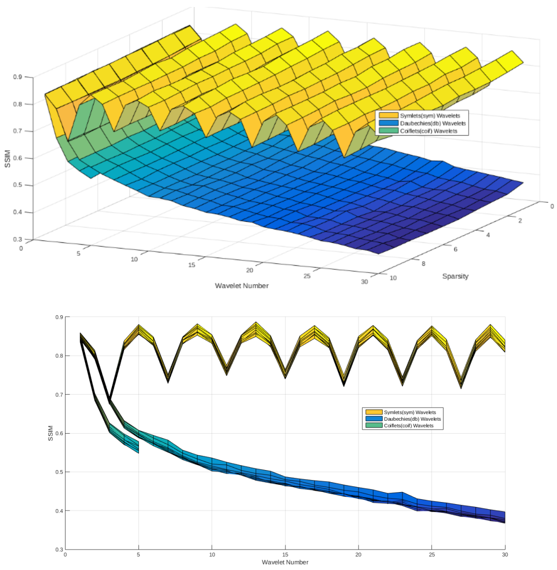

Fig. 7. The surface plot of average SSIM vs Choice of Wavelet

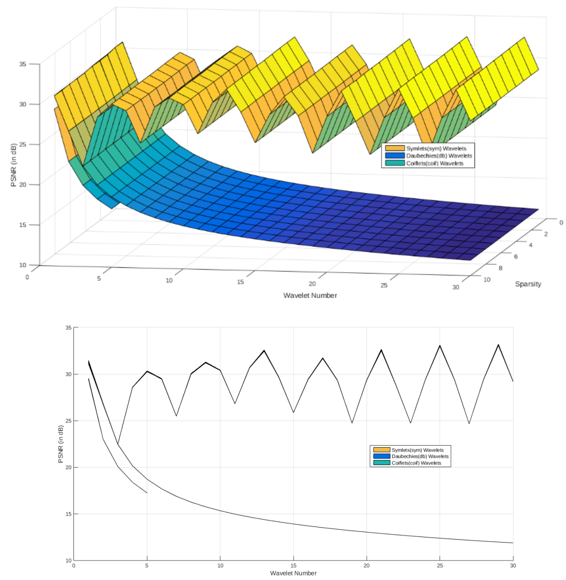

Fig. 9. The surface plot of average PSNR vs Choice of Wavelet

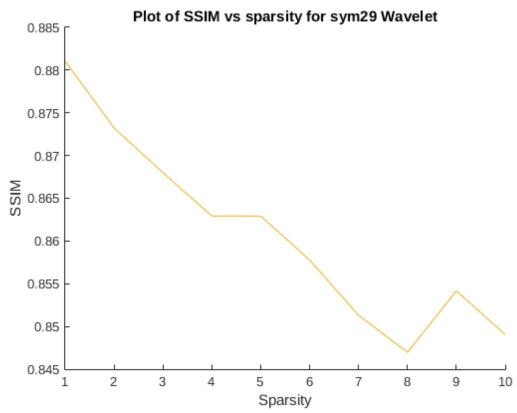

Fig. 8. SSIM for 'sym29' vs sparsity

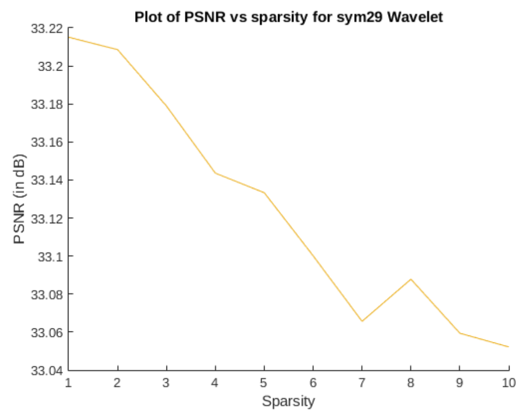

Fig. 10. PSNR for 'sym29' vs sparsity

- Symlets with wave number in the form of 4N + 1 are a good choice.

5.2. Convolutional Neural Networks in Wavelet Domain for SISR

Recently Deep Learning has become an extremely important part of the field of image processing. Convolutional Neural Networks specially have been used for a plethora of tasks including super resolution. As wavelets have not been explored much in conjunction with CNNs, we thus propose a simple CNN based architecture as shown in image 11

Our proposed network takes as input a stack of all four of the outputs of a discreet wavelet transform with some of the wavelets which we found were performing better on SISR in the course of experimentation in previous section, for instance

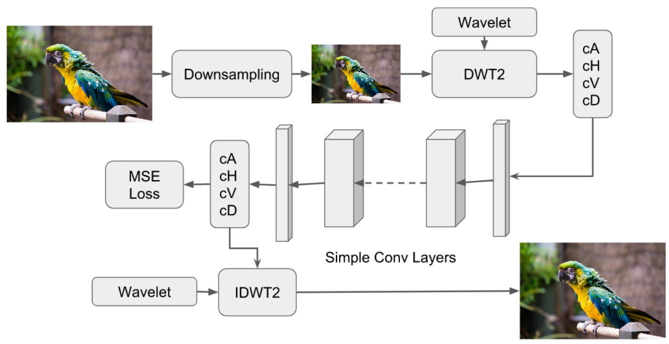

Fig. 11. Simple CNN for SISR in Wavelet Domain

sym9. In the middle we have 5-10 hidden convolutional layers each of which is activated with a ReLU. At the output we calculate a simple Mean Squared Error (MSE) Loss between the 4 layer stack formed from the Discrete wavelet transform using same parameters as before and the output of the CNNs. We minimize this loss with a gradually decreasing learning rate and at testing time we reconstruct the image from the 4 layer output of the network using an Inverse Discrete Wavelet Transform(IDWT2).

We train the network on 800 high resolution images from the NTIRE dataset. We then conduct test for this using 100 images from the same dataset. We use Peak Signal to Noise Ratios (PSNR) scores as the metric for the results. While conducting these test we have observed values of PSNRs which are as high as 44 dBs and an average of 34 dB, which is very close the highest performing methods on the NTIRE dataset.

Fig. 12. Results with the image of a flower. Image taken from NTIRE dataset. (Top Left : WaveNN, Top Right : Bicubic, Bottom : Original)

5.3. Comparison of Methods

A comparison of the algorithms of [13, 14, 12] has been performed. 'sym29', a sparsity of 3 and dictionary size of 256 has been used for [13]. The K-SVD library, KSVDBox v12, implemented by Ron Rubinstein `http://www.cs.technion.ac.il/~ronrubin/software.html` is used to perform the dictionary learning. The code for [14] was obtained from `http://www4.comp.polyu.edu.hk/~cslzhang/SCDL.htm`. The code for [12] was obtained from `http://www.ifp.illinois.edu/~jyang29/codes/ScSR.rar`

The training data used in all 3 algorithms consisted of 70 images made available by J. Yang as part of the code for their implementation [11]. The testing data consisted of 5 images.

Fig. 13. Results with the image of a bird. Image taken from NTIRE dataset. (Top Left : Orignal, Top Right : [13], Bottom Left : Bicubic, Bottom Right : [12])

Fig. 14. Results with Lena. (Top Left : Orignal, Top Right : [13], Bottom Left : Bicubic, Bottom Right : [12])

Time analysis of algorithms during Reconstruction (in seconds)

Image	[13] (Our Implementation)	[12]	[14]
Lena	13.53	651.93	1724.5
Bird	7.37	417.02	1066.65
PCB	12.42	1728.58	1653.50
Crab	7.27	393.64	1047.79
Butterfly	63.74	3399.46	5152.26
Sum	104.33	6590.63	10644.7

SSIM of algorithms

Image	[13] (Our Implementation)	[12]	[14]
Lena	0.9331	0.8881	0.8752
Bird	0.9710	0.8527	0.8398
PCB	0.9365	0.8743	0.9264
Crab	0.8894	0.8743	0.9148
Butterfly	0.9615	0.8857	0.9414
Mean	0.9383	0.8750	0.8995

PSNR of algorithms

Image	[13] (Our Implementation)	[12]	[14]
Lena	36.33	34.09	36.80
Bird	30.33	29.22	31.56
PCB	30.71	28.73	29.43
Crab	36.60	35.38	37.76
Butterfly	32.14	32.9	35.03
Mean	33.22	32.08	34.12

6. FUTURE WORK

In this project we established a few heuristics for the selection of the wavelets which are to be used for SISR. As this was done experimentally one of the best expansion to this project would be the addition of theoretical basis for the results. In the deep learning architecture, we can introduce a patch based method with overlap. This method should increase the PSNR values as one patch would be seen by the network multiple times and in different contexts.

7. REFERENCES

[1] David Capel, "Image mosaicing," in *Image Mosaicing and Super-resolution*, pp. 47–79. Springer, 2004.

[2] Michal Irani and Shmuel Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[3] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar, "Fast and robust multiframe super resolution," *IEEE transactions on image processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[4] Simon Baker and Takeo Kanade, "Hallucinating faces," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 83–88.

[5] William T Freeman, Egon C Pasztor, and Owen T Carmichael, "Learning low-level vision," *International journal of computer vision*, vol. 40, no. 1, pp. 25–47, 2000.

[6] William T Freeman, Thouis R Jones, and Egon C Pasztor, "Example-based super-resolution," *IEEE Computer graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[7] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[8] Irina F Gorodnitsky and Bhaskar D Rao, "Sparse signal reconstruction from limited data using focuss: A reweighted minimum norm algorithm," *IEEE Transactions on signal processing*, vol. 45, no. 3, pp. 600–616, 1997.

[9] Boaz Ophir, Michael Lustig, and Michael Elad, "Multiscale dictionary learning using wavelets," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1014–1024, 2011.

[10] Mahmoud Nazzal and H Ozkaramanli, "Wavelet domain dictionary learning-based single image superresolution," *Signal, Image and Video Processing*, vol. 9, no. 7, pp. 1491–1501, 2015.

[11] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[12] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang, "Coupled dictionary training for image super-resolution," *IEEE transactions on image processing*, vol. 21, no. 8, pp. 3467–3478, 2012.

[13] Junaid Ahmed, Gulsher Lund Baloch, and Huseyin Ozkaramanli, "Coupled k-svd dictionary learning algorithm in wavelet domain for single image super-resolution," in *Imaging Systems and Techniques (IST), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–5.

[14] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2216–2223.

[15] Jian Xu, Chun Qi, and Zhiguo Chang, "Coupled k-svd dictionary training for super-resolution," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3910–3914.

[16] Ingrid Daubechies, *Ten lectures on wavelets*, vol. 61, Siam, 1992.