U-FRESH: AN FRI-BASED SINGLE IMAGE SUPER RESOLUTION ALGORITHM AND AN APPLICATION IN IMAGE COMPRESSION

Xin Deng, Junjie Huang, Mengying Liu, and Pier Luigi Dragotti

Department of Electrical and Electronic Engineering, Imperial College London.

ABSTRACT

Learning based single image super resolution (SISR) methods have achieved notable results, however, they require large datasets for training, and may struggle when there is a mismatch between the testing and training data. To overcome these drawbacks, we propose an approach, named U-FRESH, which only requires a small dataset but can achieve state-of-the-art performance also in the presence of training and testing mismatches. We accomplish this by leveraging a method called FRESH, which enhances the image resolution using FRI theory. We start upscaling from the FRESH generated low resolution image. To minimize the reconstruction error, we propose a new regression selection technique to make the mapping more reliable and robust, and a wavelet based back projection technique to improve the quality of the reconstructed image. Based on U-FRESH, we also propose a new framework based on JPEG 2000 for image compression. Numerical results show that our U-FRESH method achieves state-of-the-art performance in SISR and provides better compression results than JPEG 2000.

***Index Terms*—** single image super resolution, image compression, local linear regression.

1. INTRODUCTION

Single image super resolution (SISR) aims to recover a high resolution (HR) image from a single low resolution (LR) one. This inverse process is highly ill-posed, because one LR image may correspond to many different HR images. In order to reduce this ambiguity, many methods have been proposed, which can be typically classified into two broad categories: methods that use external dictionaries, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] and methods based on self-learning or some forms of constrained reconstruction that do not require external datasets, e.g., [14, 15, 16, 17, 18]. Algorithms that use external dictionaries are based on the idea that each patch in one image can find similar patches in other images. By retrieving the high frequency component from these similar patches, the current patch resolution can be enhanced. The most popular way to retrieve the high frequency is by learning the relationship between LR and HR patches. According to the learning strategies, SISR methods based on external dictionaries can be further divided into methods based on neighbour embedding [1, 2], sparse coding [3, 4], direct mapping [6, 7, 8, 9, 10], and deep learning [11, 12, 13].

Despite their efficiency, methods based on external dictionaries require a large dataset for training, which may not be available. Moreover, the SR performance is seriously affected by the similarity between the training and testing data. In contrast, we have recently seen the developments of methods which do not require external dictionaries, but which are approximately as efficient as those using external datasets, e.g., [15, 16, 17, 18]. In particular, FRESH [15] leverages the theory of sampling signals with finite rate of innovation (FRI) [19, 20] to solve the SISR problem and obtains competitive results. Specifically, FRESH treats the LR image as the low-pass version of a wavelet decomposition and uses FRI theory to infer the missing wavelet coefficients. Each image line is modeled as a piece-wise smooth signal, which can be split into a piece-wise polynomial and a global smooth functions. The former is reconstructed using FRI theory and the latter using linear reconstruction. Although FRESH obtains good results, the limitation is that the piece-wise smooth model assumption is not accurate. Therefore, FRESH performs well on edges but may struggle otherwise.

In this paper, we aim to get the best out of the two approaches, and we do so by introducing U-FRESH which combines FRESH with a new low-complexity learning approach. The key insight here is that learning from external datasets is helpful but if strong priors are available, the training complexity and the size of the external datasets can be significantly reduced. Compared to other methods, an important advantage of U-FRESH is that we can achieve better SISR performance with less training images, i.e., only a fraction of others [4, 5, 8, 9, 7, 10, 11, 12, 13], and we are more resilient to the mismatch (i.e., using different blurring kernels) between the training and testing data. For resolution enhancement, U-FRESH employs the local regression learning approach as in [7, 10] but with some important changes: 1) different from other methods which start upscaling from bicubic interpolated LR images, we start from a higher position, the FRESH generated LR images. This allows us to use a smaller dataset for training, because the LR images generated by FRESH already contain

Xin Deng is supported by Imperial CSC scholarship.

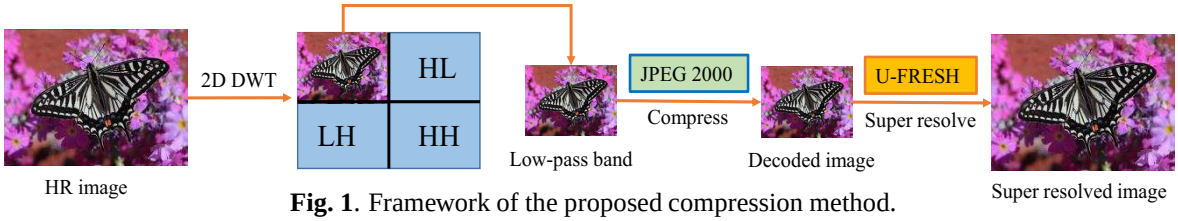


Fig. 1. Framework of the proposed compression method.

many high-frequency details. 2) a new regression selection algorithm is proposed to make the mapping more reliable and robust.

In addition to SISR, we also apply U-FRESH to image compression. Since U-FRESH is built around FRESH which treats LR images as low-pass versions of a wavelet decomposition, U-FRESH is a natural candidate to improve the performance of JPEG 2000 [21], which is a state-of-the-art image coding system based on wavelet technology. We focus on low bit-rate settings and in our approach, as shown in Fig. 1, instead of directly compressing the HR image, we perform discrete wavelet transform (DWT) on the HR image and only compress the low-pass subband using JPEG 2000. After that, we apply U-FRESH on the decoded low-pass image to upscale it to the original resolution. In this way, we can get good compression results even at low bit-rates.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed U-FRESH method and show how we tailor it for image compression. In Section 3, we show the performance of U-FRESH in both SISR and image compression scenarios. Section 4 concludes this paper.

2. PROPOSED APPROACH

2.1. Training

The training set contains only 20 images randomly selected from BSD300 dataset[1], which is much less than those used in other methods. Following [15], we first perform 2-D DWT on the HR images to extract the low-pass sub-bands. We then apply FRESH [15] on the low-pass sub-bands to obtain LR images, as Fig. 2 shows.

Patch pairs selection. After obtaining the LR and HR patch pairs, we perform feature extraction. The LR-HR patch pairs are normalized first, then as in [10], we use the mean-removed LR patch as the LR feature. The HR feature is the HR patch after removing the mean value of its corresponding LR patch. For effective training, we remove the smooth patches with small variance. We also discard the patch pairs which have low correlation coefficients (CC), since we notice that patch pairs with low CC values can degrade the training performance. The CC value between the LR and HR feature vectors $\mathbf{f}_l$ and $\mathbf{f}_h$ is given by

$$\boldsymbol{\rho}(\mathbf{f}_l, \mathbf{f}_h) = \frac{\langle \mathbf{f}_l, \mathbf{f}_h \rangle}{\|\mathbf{f}_l\|_2 \|\mathbf{f}_h\|_2}, \tag{1}$$

where $\langle \cdot \rangle$ indicates the inner product.

Linear regression. We use K-means clustering algorithm to split the training set in K groups with centroids $\{\boldsymbol{C}_i\}_{i=1}^K$.

[1]The training dataset used in this paper can be downloaded from https://drive.google.com/file/d/0Bzxdhi861FZacjlkeklDdFJCeDg/view?usp=sharing.

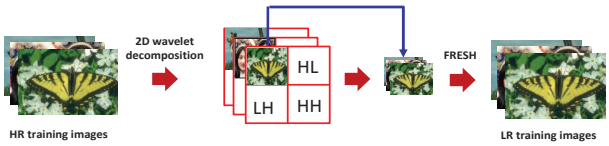


Fig. 2. The process of obtaining LR images from HR images.

For each centroid, we further select J nearest neighbours to form its cluster. We then calculate a linear regression $\{\boldsymbol{R}_i\}_{i=1}^K$ for each cluster, which maps the LR samples in the cluster to the corresponding HR samples with the minimum error:

$$\boldsymbol{R}_i = \underset{\boldsymbol{R}_i}{\operatorname{argmin}} \sum_{j=1}^{J} \|\boldsymbol{h}_i^j - \boldsymbol{R}_i \boldsymbol{\ell}_i^j\|_2^2 + \lambda \|\boldsymbol{R}_i\|_F^2, \tag{2}$$

where $\{\boldsymbol{h}_i^j\}_{j=1}^J$ and $\{\boldsymbol{\ell}_i^j\}_{j=1}^J$ indicate the HR and LR features in the i-th cluster $\mathbf{C}_i$. Here, λ is a regularization parameter. By denoting $\boldsymbol{H}_i = [\boldsymbol{h}_i^1, \boldsymbol{h}_i^2, ..., \boldsymbol{h}_i^J]$ and $\boldsymbol{L}_i = [\boldsymbol{\ell}_i^1, \boldsymbol{\ell}_i^2, ..., \boldsymbol{\ell}_i^J]$, (2) can be written as

$$\boldsymbol{R}_i = \underset{\boldsymbol{R}_i}{\operatorname{argmin}} \|\boldsymbol{H}_i - \boldsymbol{R}_i \boldsymbol{L}_i\|_2^2 + \lambda \|\boldsymbol{R}_i\|_F^2, \tag{3}$$

and ridge regression gives the following closed-form solution to (3),

$$\boldsymbol{R}_i = \boldsymbol{H}_i \boldsymbol{L}_i^T (\boldsymbol{L}_i \boldsymbol{L}_i^T + \lambda \boldsymbol{I})^{-1}, \tag{4}$$

where $\boldsymbol{I}$ is the identity matrix. The value of λ is selected via the validation set, through which we choose the value leading to the minimum regression error.

Here, we decided to use linear regression to learn the mappings, instead of other mapping algorithms such as sparse coding, because linear regression is more suited to reconstruct the smooth and textured regions where the FRESH algorithm, which our U-FRESH is based on, struggles.

2.2. Reliable Regression Selection

After training, we have obtained K regressions $\{\boldsymbol{R}_i\}_{i=1}^K$ for clusters with centroids $\{\boldsymbol{C}_i\}_{i=1}^K$. In the testing, given an input LR feature, the most important step is to select a reliable regression which can map the LR feature to HR feature with the smallest error. The traditional way is to look for the nearest centroid to the input LR feature and select its corresponding regression matrix to do the mapping [10]. However, this mapping can be unreliable, especially when the LR feature is located far away from the centroid. Dai *et.al* have realised this problem and they proposed to optimize the regression calculation by clustering samples based on the regression instead of sample values [9]. Different from [9], we propose a new technique to select the most

reliable regression for the input LR through estimating the reconstruction error, with the following three steps:

Step 1. Given an input LR feature, instead of searching for its nearest centroid, we use k-nearest neighbour (kNN) algorithm to search for its N nearest centroids. Typically, $N = 3$.

Step 2. From the samples in the N nearest clusters, we select M nearest samples to the LR input. These M LR samples form a subset $\mathbf{S}_M^l$, and their corresponding HR samples form a subset $\mathbf{S}_M^h$.

Step 3. From the N nearest clusters, we can have N regressions forming a regression set $\mathbf{R}_N$. For each regression, we calculate its reconstruction error on the M samples gathered in Step 2, and choose as the regression the one that leads to the smallest error, as formulated in Eq. (5):

$$\boldsymbol{R} = \underset{\boldsymbol{R}}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^{M} \|\boldsymbol{h}_m - \mathbf{R}\boldsymbol{\ell}_m\|_2^2, \tag{5}$$

where $\boldsymbol{R} \in \{\boldsymbol{R}_N\}$, $\boldsymbol{h}_m \in \{\boldsymbol{S}_M^h\}$, and $\boldsymbol{\ell}_m \in \{\mathbf{S}_M^l\}$.

To further enhance the super-resolved image quality, we use another two important steps: the wavelet based back projection (WBP) and ensemble learning based resolution enhancement. WBP is to ensure that the estimated HR image is consistent with the input LR image. Recall that we treat LR images as the low-pass sub-band of the HR images. Thus, given the super-resolved HR image, we do wavelet decomposition, and replace its low-pass band with the original LR input but keep its high-pass sub-bands unchanged. Finally, benefitting from the power of ensemble learning [22], we create different variations of LR images through affine transformation, i.e., four LR variations with rotations $\{0, 90, 180, 270\}$ and average their HR estimates to obtain the final HR image.

In this paper, the cascaded structure is employed to implement upscaling for scaling factors larger than 2, e.g., $4\times$. We first do upscaling by scaling factor 2, and then after error correction by WBP, we further do upscaling by 2. Note that the second U-FRESH regressions need to be re-trained.

2.3. Application to image compression

To apply U-FRESH to image compression, we need to establish a training dataset different from that of SISR. In the context of image compression, as shown in Fig.1, the U-FRESH algorithm is used to map the decoded low-pass images to the HR images. Thus, to prepare the LR training images, we first do DWT on the HR image to get the low-pass images and then apply JPEG 2000 to compress them. After that, we perform FRESH on the decoded low-pass images to obtain the LR training images. The mappings between the LR and HR pairs are re-trained for image compression.

3. EXPERIMENTAL RESULTS

In the experiments, two datasets were used for testing, including 5 images from Set5 [23] and 14 images from Set14 [4]. All the test pictures were changed from RGB to YCbCr format and only the luminance channel was used for testing.

Table 1. Performance of U-FRESH and other methods with *bior4.4* as the blurring kernel, with the best scores bold and the second bests underlined.

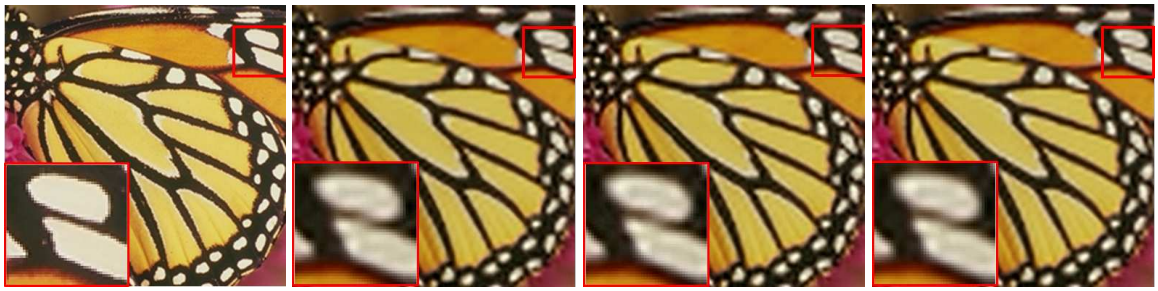
Scaling factor	2×				4×			
Dataset	Set5		Set14		Set5		Set14	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ScSR[3]	33.14	0.9323	30.12	0.8804	27.56	0.8072	25.58	0.7064
Zeyde[4]	34.83	0.9459	31.31	0.8989	28.68	0.8360	26.37	0.7336
GR [5]	33.96	0.9394	30.71	0.8940	28.15	0.8170	25.95	0.7217
ANR [5]	34.72	0.9453	31.18	0.8987	28.59	0.8320	26.26	0.7310
NE+LLE [1]	34.68	0.9444	31.15	0.8976	28.52	0.8311	26.23	0.7300
A+ [8]	35.41	0.9491	31.65	0.9022	29.16	0.8504	26.70	0.7440
SelfEx[18]	34.66	0.9439	31.16	0.8956	28.30	0.8296	26.09	0.7310
SRCNN [11]	35.47	0.9493	31.68	0.9024	29.17	0.8504	26.74	0.7447
FRESH [15]	35.38	0.9487	31.59	0.9012	29.43	0.8551	26.77	0.7457
Ours	**35.65**	**0.9496**	**31.93**	**0.9035**	**29.65**	**0.8594**	**26.95**	**0.7494**

Table 2. Performance of U-FRESH and other methods for $4\times$ upscaling, with arbitrary blurring kernels.

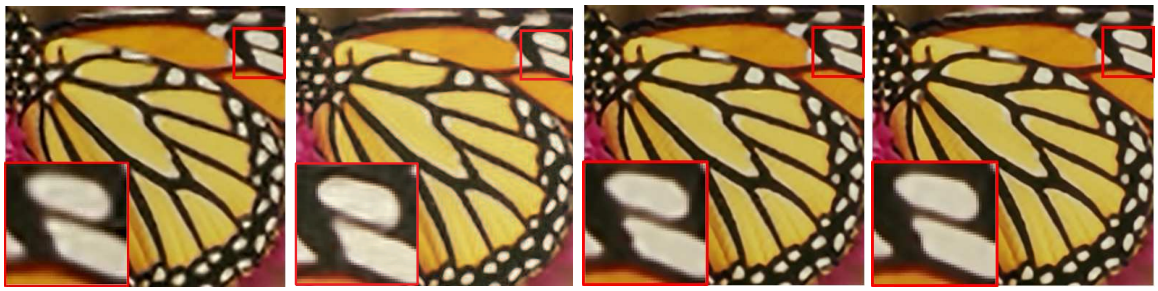
Blurring kernel	*bior2.4*		*bior6.8*		*rbio2.8*		*linear spline*	
	Set5	Set14	Set5	Set14	Set5	Set14	Set5	Set14
CSCN[12]	24.44	23.12	28.23	26.27	28.09	26.30	24.96	23.42
VDSR[13]	24.30	22.73	28.32	25.93	28.83	26.55	24.78	22.95
FRESH [15]	29.33	26.74	29.54	26.89	29.63	26.96	29.10	26.54
Ours	**29.54**	**26.87**	**29.68**	**26.96**	**29.71**	**27.00**	**29.40**	**26.77**

Parameters setting. In the training, the threshold for removing smooth patches is 0.5 and the threshold of CC value is 0.65. In the regression step, λ is 0.01. In reliable regression selection, we use K=2048, N=3 and M=128. The patch size is 5×5 for upscaling by 2 and 9×9 for upscaling by 4.

SISR performance. Since the blurring kernel in practical scenario usually cannot be predicted, we evaluated the SISR performance of our method in two cases: 1) the training and testing processes use the same blurring kernel; 2) the testing process uses arbitrary blurring kernels different from that used during the training. Table 1 shows the PSNR and SSIM [24] results of our and other methods in the first case. For fair comparison, the blurring kernel in all methods was modified to be *bior4.4* as [15], and all the dictionaries and network were retrained using this kernel. We can see that our method consistently outperforms all the other methods. Specifically, for $2\times$ upscaling, we improve PSNR by 0.18 dB in Set5 and 0.25 dB in Set14, compared with the second best SRCNN [11]. For $4\times$ upscaling, we achieve 0.22 dB PSNR improvement in Set5 and 0.18 dB improvement in Set 14, compared with the second best FRESH[15]. Note that we only use 20 training images while others [1, 3, 4, 5, 8, 11] use 91 images. We can use less training images and still achieve better SR performance, because the FRESH generated LR images already include some high frequency details which do not need to be recovered again from training images. For visual comparisons, Fig. 3 shows the $4\times$ upscaling results on *Butterfly*. We can see that our U-FRESH method can generate sharper edges with less ringing artifacts.



(a) Ground truth (b) ScSR (22.3/0.7543) (c) ANR (23.6/0.7959) (d) NE+LLE (23.5/0.7970)



(e) A+ (24.5/0.8469) (f) SRCNN (25.0/0.8590) (g) FRESH (25.6/0.8738) (h) Ours (**26.3**/**0.8874**)

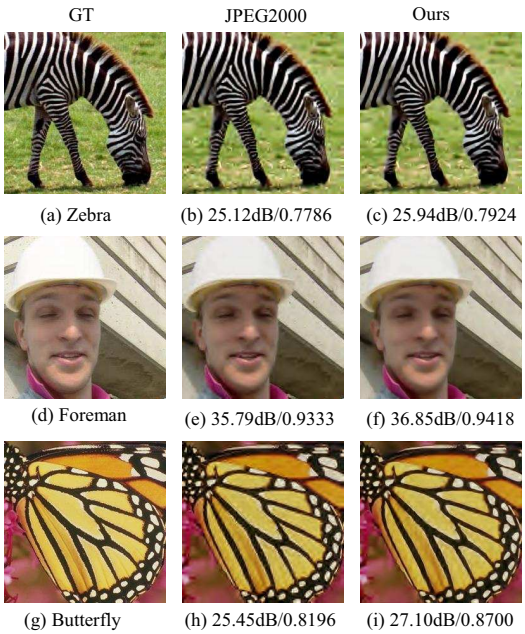
Fig. 3. SISR results of *Butterfly* by our and other methods with 4× upscaling. The values in the bracket are PSNR/SSIM.

Table 2 presents the PSNR results of our method in the second case, i.e., the testing process uses different kernels from the training which uses *bior4.4*. We compare the results with other two state-of-the-art deep learning based methods, CSCN[12] and VDSR[13], and we can see that our method is more resilient to the mismatches in the blurring kernels while [12] and [13] can be seriously affected[2], e.g., *bior2.4* kernel.

Image compression performance. We compare our compression results with JPEG 2000 at different bit rates, with the improvements shown in Table 3. As can be seen, our compression method improves significantly over JPEG 2000 at small bit rates. Specifically, at 0.2 bit per pixel (bpp), we achieve nearly 0.4 dB improvement over JPEG 2000 in Set 14. Fig. 4 compares visually the compressed images and we can see that our method achieves higher reconstruction quality than JPEG 2000.

Table 3. The improvement of our method over JPEG 2000.

Bitrate	Set5		Set14	
(bpp)	PSNR	SSIM	PSNR	SSIM
0.08	0.0987	0.0071	0.2451	0.0045
0.1	0.1807	0.0080	0.3423	0.0090
0.2	0.3337	0.0101	0.3985	0.0000
0.3	0.3635	0.0086	0.1786	-0.0039



(a) Zebra (b) 25.12dB/0.7786 (c) 25.94dB/0.7924

(d) Foreman (e) 35.79dB/0.9333 (f) 36.85dB/0.9418

(g) Butterfly (h) 25.45dB/0.8196 (i) 27.10dB/0.8700

Fig. 4. Results of images compressed by JPEG 2000 and our method. The compression bitrate is 0.3 bpp. The values shown are PSNR/SSIM. Better seen in enlarged version.

4. CONCLUSION

In this paper, we propose an U-FRESH method for SISR and image compression. Benefitting from FRESH, our method only needs a small dataset for training. We develop a reliable regression selection technique to make mapping robust and use wavelet based back projection to eliminate reconstruction errors. Our method combines the merits of learning based and signal processing based methods, achieving nearly 0.3 dB PSNR improvement over other state-of-the-art methods in SISR and 0.4 dB improvement over JPEG 2000 in low bit-rate image compression regime.

[2]Here, we simply use the trained network of [12] and [13] which use *bicubic* as training kernel. In our method, the training kernel is *bior4.4*, which is also different from the testing kernels.

5. REFERENCES

[1] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 275–282, IEEE, 2004.

[2] T.-M. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognition Letters*, vol. 30, no. 5, pp. 494–502, 2009.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[4] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*, pp. 711–730, Springer, 2010.

[5] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1920–1927, 2013.

[6] J.-J. Huang, W.-C. Siu, and T.-R. Liu, "Fast image interpolation via random forests," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3232–3245, 2015.

[7] A. Choudhury and P. van Beek, "Boosting performance and speed of single-image super-resolution based on partitioned linear regression," in *Image Processing (ICIP), 2016 IEEE International Conference on*, pp. 1419–1423, IEEE, 2016.

[8] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian Conference on Computer Vision (ACCV)*, pp. 111–126, Springer, 2014.

[9] D. Dai, R. Timofte, and L. Van Gool, "Jointly optimized regressors for image super-resolution," in *Computer Graphics Forum*, vol. 34, pp. 95–104, Wiley Online Library, 2015.

[10] C.-Y. Yang and M.-H. Yang, "Fast direct super-resolution by simple functions," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 561–568, 2013.

[11] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*, pp. 184–199, Springer, 2014.

[12] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 370–378, 2015.

[13] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*, June 2016.

[14] X. Wei and P. L. Dragotti, "Sampling piecewise smooth signals and its application to image up-sampling," in *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 4293–4297, IEEE, 2015.

[15] X. Wei and P.-L. Dragotti, "FRESH-FRI-based single-image super-resolution algorithm," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3723–3735, 2016.

[16] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 349–356, IEEE, 2009.

[17] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5334–5347, 2014.

[18] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206, 2015.

[19] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[20] J. A. Urigüen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, 2013.

[21] D. S. Taubman and M. W. Marcellin, "JPEG2000: Standard for interactive imaging," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1336–1357, 2002.

[22] Z.-H. Zhou, "Ensemble learning," *Encyclopedia of Biometrics*, pp. 411–416, 2015.

[23] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in BMVC, 2012.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.