This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

SRHRF+: Self-Example Enhanced Single Image Super-Resolution Using Hierarchical Random Forests

Jun-Jie Huang, Tianrui Liu, Pier Luigi Dragotti, and Tania Stathaki
Imperial College London, UK
{j.huang15, t.liu15, p.dragotti, t.stathaki}@imperial.ac.uk

Abstract

Example-based single image super-resolution (SISR) methods use external training datasets and have recently attracted a lot of interest. Self-example based SISR methods exploit redundant non-local self-similar patterns in natural images and because of that are more able to adapt to the image at hand to generate high quality super-resolved images. In this paper, we propose to combine the advantages of example-based SISR and self-example based SISR. A novel hierarchical random forests based super-resolution (SRHRF) method is proposed to learn statistical priors from external training images. Each layer of random forests reduce the estimation error due to variance by aggregating prediction models from multiple decision trees. The hierarchical structure further boosts the performance by pushing the estimation error due to bias towards zero. In order to further adaptively improve the super-resolved image, a self-example random forests (SERF) is learned from an image pyramid pair constructed from the down-sampled SRHRF generated result. Extensive numerical results show that the SRHRF method enhanced using SERF (SRHRF+) achieves the state-of-the-art performance on natural images and yields substantially superior performance for image with rich self-similar patterns.

1. Introduction

Image super-resolution (SR) is an interesting and classical problem in computer vision and signal processing. It has many applications ranging from face super-resolution to hyper-spectral imaging. The objective is to recover a sharp and natural-looking high resolution (HR) image from one or more blurred low resolution (LR) images. In particular, single image super-resolution (SISR) aims to restore the HR image from a single input LR image. This is an ill-posed problem as the size of LR observation is much smaller than the desired HR output. A variety of priors have been proposed to regularize the problem. Example-based SR methods [25, 21, 4] learn statistical priors to model the correspondences between LR patches and HR patches from an external training dataset. Self-example based approaches [7, 8, 24] make use of the LR image itself as training data. In order to take advantage of the non-local similarity property of natural images, these methods adaptively enhance the super-resolved image using affine or perspective transformed similar patches across scales.

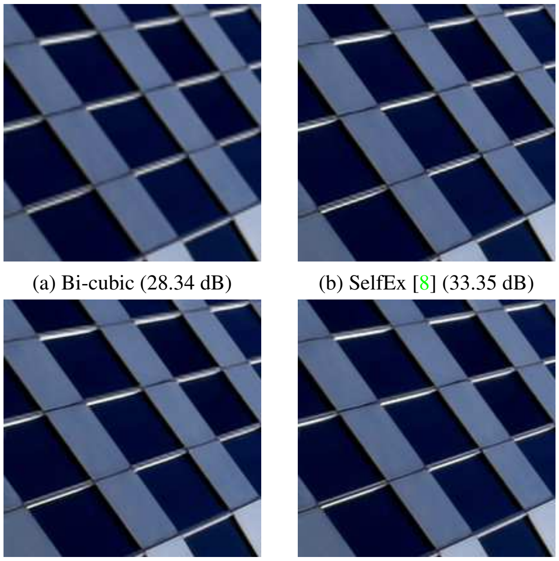

(a) Bi-cubic (28.34 dB) (b) SelfEx [8] (33.35 dB)
(c) SRHRF (32.92dB) (d) SRHRF+ (35.22dB)

Figure 1: Our proposed SR methods, SRHRF and SRHRF+, compares with the state-of-the-art self-example based method [8] (upscaling factor is 2).

In this paper, we propose to combine the merit of the example-based approaches with the adaptive capability of the self-example based approaches. A novel SR hierarchical random forests (SRHRF) method is first used to learn statistical priors from external training images. This is followed by a self-example algorithm which further exploit useful information from a pyramid of internal training images super-resolved by SRHRF. The proposed

SRHRF method combines the strength of the bagging (a.k.a **B**ootstrap **Agg**regat**ing**) strategy in random forests together with the boosting idea. In such way, SRHRF can optimize SR estimation error in terms of both variance and bias. More specifically, SRHRF can reduce the variance errors with random forests, and can meanwhile decrease the estimation bias via cascaded forests.

Promising SR results have been reported in the hierarchical decision trees (SRHDT) method [10] using boosted decision trees to reduce the estimation bias. At each stage in their work, a single decision tree is first trained and then expanded into a decision forest consisting of 8 decision trees. Apart from one trained decision tree, there are 7 other non-trained decision trees constructed by manipulating the regression models on each leaf node of the trained one. However, there is only around 0.1 dB gain through the aggregation of regression models in [10]. We conjecture that it is because the decision trees they learned are highly stable (with low diversity). Since diversity among decision trees is essential, we propose to use hierarchical random forests, rather than cascaded decision trees, to consistently refine the super-resolved image. By using the bagging strategy as well as limiting the minimum number of samples on leaf nodes, we can obtain random forests with highly diversified decision trees. In such way, the proposed SRHRF method can acquire significant gain through the fusion model idea which expands the number of decision trees in original learned random forests by a factor 8.

Our self-example enhanced algorithm (SRHRF+) makes further improvement on SR quality by exploiting image internal statistics (see Figure 1 for example). It takes a middle resolution (MR) image provided by SRHRF as input and applies an adaptive random forests, be refereed as self-example random forests (SERF), for further enhancement. The SERF is trained using LR and HR feature pairs extracted from an image pyramid pair which is constructed using the down-sampled versions of the MR image. Self-example SR is especially effective for the images with a large number of patches with similar patterns. Numerical results show that the proposed SRHRF already achieves state-of-the-art SISR performance, and SRHRF+ can further improve SRHRF up to 0.7 dB on average depending on the image content.

2. Related Work

2.1. Example-based Super-Resolution

Example-based SR mainly refers to methods where the learning is achieved using external datasets. Various models have been developed to capture the complex relationship between LR patches and HR patches. Chang *et al.* [3] proposed a neighbor embedding SR algorithm based on manifold learning. Sparse coding based algorithms [26, 25] assume that an LR patch and its corresponding HR patch share the same sparse code over a coupled LR and HR dictionaries. Kim and Kwon [14] proposed to learn a non-linear mapping using kernel ridge regression for SISR. The idea of "first-classification-then-regression" starts from the simple functions (SF) method proposed by Yang and Yang [23] and anchored neighbor regression (ANR) method proposed by Timofte *et al.* [20]. SF method [24] classifies the LR patch space using K-means clustering and learns a linear regression for each cluster to map LR patches belonging to this cluster to HR patch space. ANR method [20, 21] instead performs LR patch classification by treating each atom in a sparse dictionary as an anchoring point and finding a nearest neighbor atom for each input LR patch, and applies a ridge regression for HR patch estimation at each class. With a hierarchical search structure during testing, random forests [11, 9, 17, 16] have been applied for fast and high quality image upscaling. SR hierarchical decision trees (SRHDT) method [10] cascades multiple decision trees to boost the performance of a single layer decision tree or simulated random forests. Recently, deep learning based approaches [4, 5, 12, 13] utilize deep learning techniques to learn a neural network for resolution enhancement and achieve state-of-the-art performance.

2.2. Self-Example Based Super-Resolution

With the rapid development of example-based SR, learning LR-HR patch correspondences from external dataset seems to be the ideal solution to the SISR problem. However, the redundant non-local similar patches within the input LR image or enhanced MR image are valuable sources of further information to restore the missing high frequency details. Self-example SR based methods [7, 6, 1, 8, 19, 24, 22], which do not rely on external dataset, are mainly based on the fact that natural images contain a large number of repetitive structures within and across different scales. Glasner *et al.* [7] exploited the patch redundancy across scales and proposed a unified framework which combines the idea of classical multi-frame SR and example-based SR. Non-locally centralized sparse representation method [6] models the sparse coding noise as a Laplacian distribution and searches non-local similar patches to obtain a better estimation of the sparse coding. Bevilacqua *et al.* [1] built a double image pyramid using the LR input image through which a linear mapping is learned for each LR patch. Huang *et al.* [8] exploited transformed self-examplar patches and further expanded the internal patch search space. ANR based self-example method [19] applies the down-sampled versions of the input image and their rotated and flipped images to generate internal training data and shows better performance compared to the A+ (Adaptive ANR) method [21] which uses external training images. The finite rate of innovation based SISR (FRESH) method [22] enhances the image resolution line by line and further improves the quality using self-learning. The advantage of these self-example based approaches [7, 6, 1, 8, 19, 24, 22] is that they exploit repetitive structure and can adapt to the characteristics of the input image. However, they need longer processing time as on-line learning is required.

3. Super-Resolution via Hierarchical Random Forests

SR random forests (SRRF) [9, 17, 16] is a fast and effective SISR algorithm which learns from external training images. It exploits the property of random forests [2] to reduce error induced by estimation variance by averaging the results from multiple decorrelated decision trees.

From a mathematical point of view, SISR which aims to predict an HR image $\boldsymbol{I}_H$ from the given LR image $\boldsymbol{I}_L$ can be considered as a mapping $\boldsymbol{f}(\cdot)$ between $\boldsymbol{I}_L$ and $\boldsymbol{I}_H$ such that $\boldsymbol{I}_H = \boldsymbol{f}(\boldsymbol{I}_L) + \boldsymbol{\epsilon}$. Here $\boldsymbol{\epsilon}$ is the estimation error which mainly consists of two error components, i.e. variance error and bias error. Random forests can decrease the variance error by averaging the predictions from each decision tree, but retains the bias error. In order to jointly optimize the estimation in terms of both variance and bias error, we combine the bagging strategy together with the boosting strategy. The proposed SR via hierarchical random forests method consists of multiple layers of random forests. In each layer, the decision trees are aggregated to reduce the estimation error due to variance, and meanwhile the hierarchical structure consistently lowers down the estimation error due to bias.

3.1. Random Forests for SR

We begin with a brief review on the super-resolution random forests. For a detailed description, please refer to [11, 9, 17, 16]. Random forests $\mathcal{F}$ are an ensemble of multiple decision trees $\{\mathcal{T}_t\}_{t=1}^T$. A decision tree $\mathcal{T}_t$ has non-leaf nodes and leaf nodes. Each non-leaf node has two child nodes and a binary split parameter. A binary split function divides the input feature space into two non-overlapping parts. Through non-leaf nodes, the whole input feature space has been divided into small non-overlapping partitions which correspond to leaf nodes. As each leaf node only covers a small part of the input feature space, a simple model is able to well approximate the mapping between input feature space and output feature space. We control the minimum number of samples N_{min} to construct a leaf node and use the bagging strategy to increase the diversity among decision trees in a random forests.

For example-based SISR algorithms, the training data are LR and HR data pairs extracted from corresponding locations at LR and HR training images, respectively. Let us denote the N training LR and HR data pairs as $\{(\boldsymbol{l}_n, \boldsymbol{h}_n)\}_{n=1}^N$ with the LR feature $\boldsymbol{l}_n \in \mathbb{R}^{D_L}$ and the HR feature $\boldsymbol{h}_n \in \mathbb{R}^{D_H}$ where D_L and D_H is the dimension of the LR and HR feature, respectively.

During training, a decision tree's root node (the first non-leaf node) is initiated with a subset training data $S_0 = \{(\boldsymbol{l}_n, \boldsymbol{h}_n)\}_{n=1}^{N_S}$ containing N_s pairs of data ($N_s < N$). At each node, a binary split function associated with binary split parameters $\boldsymbol{\theta} = \{p, q, \tau\}$ will compare two feature values on the LR feature $\boldsymbol{l}$, i.e. $\boldsymbol{l}(p)$ and $\boldsymbol{l}(q)$, with a threshold τ and will return either 0 or 1:

$$b(\boldsymbol{l}, \boldsymbol{\theta}) = \begin{cases} 1 & if\, \boldsymbol{l}(p) < \boldsymbol{l}(q) + \tau, \\ 0 & otherwise. \end{cases} \tag{1}$$

If the binary split function $b(\boldsymbol{l}, \boldsymbol{\theta})$ returns 1, the training data pair $(\boldsymbol{l}, \boldsymbol{h})$ at this node will be distributed to its left child node, otherwise to its right child node. Each non-leaf node tries to find the best binary split parameter from a pool of randomly generated binary split parameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_i\}_{i=1}^K$. At a node j, a linear regression model $\boldsymbol{C}$ is learned from the training data pairs:

$$\boldsymbol{C} = \arg\min_{\boldsymbol{C}} ||\boldsymbol{H} - \boldsymbol{C}\boldsymbol{L}||_2^2 + \lambda||\boldsymbol{C}||_2^2, \tag{2}$$

where $\boldsymbol{L}$ and $\boldsymbol{H}$ is respectively the matrix containing all LR and HR patch features at node j , and λ is the regularization parameter.

The ridge regression in Eqn. (2) has a closed form solution $\boldsymbol{C} = \boldsymbol{H}\boldsymbol{L}^T(\boldsymbol{L}\boldsymbol{L}^T + \lambda\boldsymbol{I})^{-1}$. The reconstruction error of training data S is defined as:

$$E(S) = \frac{1}{|S|}\sum_{n=1}^{|S|} ||\boldsymbol{h}_n - \boldsymbol{C}\boldsymbol{l}_n||_2^2, \tag{3}$$

where $|S|$ is the cardinality of S, and $\boldsymbol{C}$ is the learned regression model.

All the non-leaf nodes find the best binary split which achieves the highest information gain and divides the current training data S into two non-overlapping groups S^L and S^R. The information gain $I(S, \boldsymbol{\theta}_i)$ for training data S and binary split parameter $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}$ is defined as:

$$I(S, \boldsymbol{\theta}_i) = E(S) - \sum_{n\in\{L,R\}} \frac{|S^n|}{|S|} E(S^n). \tag{4}$$

When the number of training data pairs at a non-leaf node is smaller than N_{min} or the decision tree reaches the maximum depth ξ, this non-leaf node will become a leaf node and all of its training data will be applied to learn a linear regression model $\boldsymbol{C}$ as in Eqn.(2).

During testing, LR feature vectors are extracted from overlapping patches on the LR image and put into each decision tree. According to the result of the binary split function at each non-leaf node (0 or 1), the LR feature vectors will be mapped into one of the leaf node. The reconstructed HR

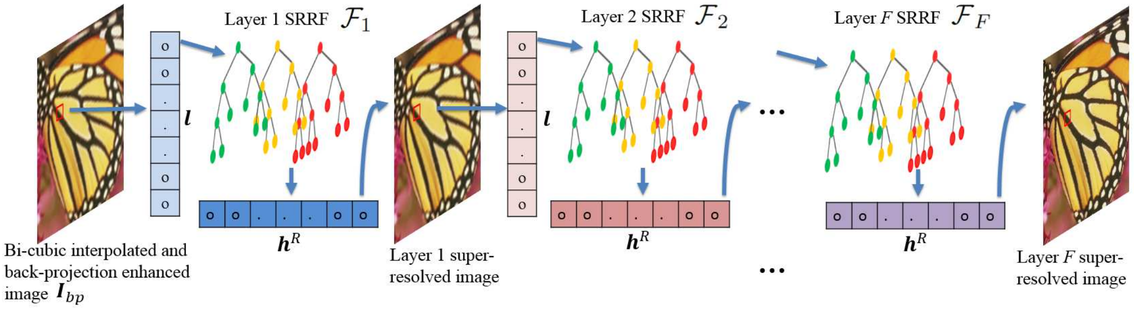

Layer 1 SRRF $\mathcal{F}_1$
Layer 2 SRRF $\mathcal{F}_2$
Layer F SRRF $\mathcal{F}_F$
$\boldsymbol{l}$
$\boldsymbol{l}$
...
$\boldsymbol{h}^R$
$\boldsymbol{h}^R$
$\boldsymbol{h}^R$
Bi-cubic interpolated and back-projection enhanced image $\boldsymbol{I}_{bp}$
Layer 1 super-resolved image
...
Layer F super-resolved image

Figure 2: Flowchart of the proposed image super-resolution via hierarchical random forests method.

feature $\boldsymbol{h}^R$ is obtained by multiplying the LR feature vector $\boldsymbol{l}$ with the averaged linear regression model retrieved from all decision trees:

$$\boldsymbol{h}^R = (\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{C}_t)\boldsymbol{l}. \tag{5}$$

In [10], regression models from 8 relevant leaf nodes of a decision tree are fused to further reduce the estimation error. By combining 8 regression models from each decision tree, we could gather 8 times more regression models for aggregation without extra training efforts. The additional testing complexity is small and comes from retrieving relevant leaf nodes and additions between regression models. We take the advantage of the fusion model idea and retrieve $8T$ regression models from random forests with T decision trees. We denote our method as SR random forests with fusion model (SRRFf). In SRRFf, a regression model from a leaf node is transformed into $\boldsymbol{C}^M$ as in [10], and are combined together to reconstruct the HR feature $\boldsymbol{h}^R$ as:

$$\boldsymbol{h}^R = (\frac{1}{8T}\sum_{t=1}^{8T}\boldsymbol{C}_t^M)\boldsymbol{l}. \tag{6}$$

As patches overlap with each other, an HR pixel value is the averaged pixel value from all the overlapping patches.

3.2. Hierarchical Random Forests for SR

When the random forests size T becomes large, increasing the number of decision trees will not lead to significant further improvement of the super-resolved images. It can be interpreted as the estimation error due to variance has already been minimized. In order to further reduce the bias in prediction, we propose the hierarchical random forests (SRHRF).

Flowchart of the proposed SRHRF method is shown in Figure 1. A hierarchical random forests $\mathcal{H}_{\mathcal{F}}$ consists of F layers of cascaded random forests $\{\mathcal{F}_f\}_{f=1}^F$. The $(n+1)^{th}$ layer random forests $\mathcal{F}_{n+1}$ make improvement on the basis of the super-resolved image through the hierarchical random forests $\{\mathcal{F}_f\}_{f=1}^n$. LR patch features are extracted from the image enhanced by the previous layer, and fed into the random forests of this layer. All $8T$ retrieved regression models from leaf nodes are averaged to form a more robust regression model. Multiplying the aggregated regression model with the LR feature vector gives the predicted HR patch feature. The estimated HR image is formed by patch overlapping.

The hierarchical random forests is trained layer by layer. The training LR images for the n^{th} layer of random forests are the output of the hierarchical random forests for image resolution enhancement up to level $n-1$. For the first layer of random forests, the training LR image is up-sampled by bi-cubic interpolation and enhanced by iterative back projection (BP).

4. Self-Example Random Forests

In this section, we further improve the SR quality by exploiting the internal statistics within the input image. Our self-example enhanced SRHRF method is named SRHRF+. SRHRF+ is effective especially for those images with a large number of similar patches. Once SRHRF has generated a high quality super-resolved image $\hat{\boldsymbol{I}}_H$ from the input LR image $\boldsymbol{I}_L$, we make use of a pyramid of internal training images to explore the non-local similarity property.

The internal statistics can be exploited from an image pyramid pair $\mathcal{P}_H = \{\hat{\boldsymbol{I}}_{H_s}^i\}_{i=1}^{N_d}$ and $\mathcal{P}_L = \{\hat{\boldsymbol{I}}_{L_s}^i\}_{i=1}^{N_d}$ which simulates the ground truth HR images and the SRHRF super-resolved images. The image pyramid $\mathcal{P}_H$ corresponds to N_d simulated HR images with decreasing resolutions by a scale factor 0.98 (for example, $\hat{\boldsymbol{I}}_{H_s}^2$ is obtained by down-sampling $\hat{\boldsymbol{I}}_{H_s}^1$ by a factor 0.98). For SR with upscaling factor s, the images in $\mathcal{P}_L$ are obtained by down-sampling images in $\mathcal{P}_H$ by a factor s and then enhancing them by SRHRF. To well model the correspondences between the desired HR image $\boldsymbol{I}_H$ and $\hat{\boldsymbol{I}}_H$, the simulated HR images should be as close to $\boldsymbol{I}_H$ as possible. Furthermore, the image pyramids should be able to provide sufficient number of reference patch pairs. For most self-example based methods [24, 1, 8, 19], the simulated HR images are generated

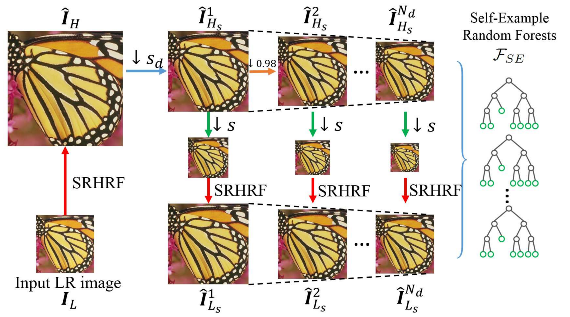


Figure 3: Self-example random forests learning from the image pyramid pair constructed from the SRHRF super-resolved image.

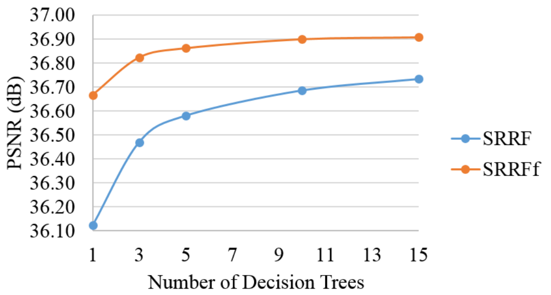


Figure 4: Comparison between SRRF and SRRFf with different number of decision trees on *Set5* with upscaling factor 2.

by down-sampling the input LR image $\boldsymbol{I}_L$. However, these simulated HR images are too small to have faithful similar structures and be able to generate enough number of training samples, especially for large upscaling factors.

As shown in Figure 3, with an initial image $\hat{\boldsymbol{I}}_H$ super-resolved by SRHRF, the largest simulated HR image $\hat{\boldsymbol{I}}^1_{H_s}$ is obtained by down-sampling $\hat{\boldsymbol{I}}_H$ by a factor s_d. Here s_d ranges from $[1/s, 1)$ for SR with upscaling factor s, e.g. for ×2 SR, s_d can range from $[1/2, 1)$. Intuitively, the higher the quality of $\hat{\boldsymbol{I}}_H$ we can get from a SR algorithm, the larger the s_d can be used to generate a natural-looking down-sampled image. When the resolution of the images in $\mathcal{P}_H$ and $\mathcal{P}_L$ is closer to that of $\hat{\boldsymbol{I}}_H$, a patch in $\hat{\boldsymbol{I}}_H$ could have more similar patches recurring in the image pyramids. Empirically, we select the value of s_d via cross-validation which will be discussed in Section 5.4.

A self-example random forests (SERF) $\mathcal{F}_{SE}$ is then learned using the self-example training data which are extracted from HR and LR patch pairs in the image pyramids $\mathcal{P}_H$ and $\mathcal{P}_L$. The training method follows that in Section 3.1. During testing, fusion model idea is not applied as $\mathcal{F}_{SE}$ is a data dependent model and model fusion does not provide significant improvement.

5. Numerical Results

In this section, we report the numerical results of our proposed methods and compare them to other commonly used SISR methods.

5.1. Experimental Settings

The 91 training images from [25] are used in our experiments for training. The testing datasets include *Set5* [25] (5 images), *Set14* [26] (14 images), and *Urban100* [8] (100 urban images with rich similar patches). Quantitative evaluations of the super-resolved images are performed using the Peak Signal-to-Noise Ratio (PSNR), and the Structural Similarity (SSIM) index. Since human eyes are more sensitive to intensity than color, color images are converted from RGB color space to YCbCr color space. The luminance channel Y is upscaled using SR algorithms, while the chrominance channels Cb and Cr are up-sampled using bicubic interpolation. Upscaling factors 2 and 4 are considered in this paper.

For upscaling factor s, HR training images are first down-sampled by a factor s and then up-sampled using bicubic interpolation followed by BP to generate the LR training images. The patch size for both LR and HR patches is $3s \times 3s$. HR and LR patch features are extracted from HR and LR training images at the positions containing patch pixels. For LR patches, features are computed as the 1^{st} and the 2^{nd} order gradients of intensities; while for HR patches, features are residuals between the HR patch and the LR patch. As the dimension of the LR patch feature is too large, they are compressed using principal component analysis (PCA), preserving 99.9% energy. During testing, the LR image is up-sampled using bi-cubic interpolation and enhanced by BP. The predicted HR image is reconstructed by adding the estimated HR feature onto the LR image.

The minimum number of samples for leaf node construction N_{min} is set to be $2D_L$. The regularization parameter λ is set to be 0.01 by cross-validation. The maximum depth of a decision tree is set to be $\xi = 20$. Implementation of the proposed methods is based on [17]. As demonstrated in [17] that the alternating regression forests [18] can optimize the global loss over random forests, we adopt the same technique for training.

5.2. Random Forests for SR

The use of random forests for SR has been proposed in [9, 16, 17]. The SR performance can be improved by aggregating regression models retrieved from multiple decision trees. We want to demonstrate that there is a potential for further improvement by directly using the random forests learned as in [9, 16, 17]. Figure 4 shows the average PSNR of SRRF [17] and SRRF with model fusion (SRRFf) evalu-

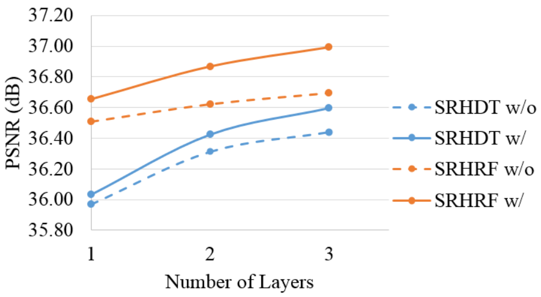


Figure 5: Comparison between SRHDT and SRHRF with and without model fusion using different number of layers on *Set5* with upscaling factor 2.

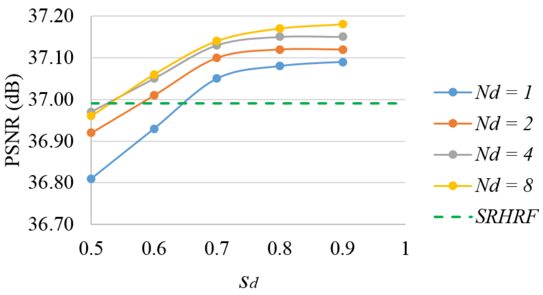


Figure 6: Self-example random forests enhanced results with different down-sampling factor s_d and number of images in the pyramid N_d for upscaling factor 2.

ated on *Set5* for upscaling factor $s = 2$ with different number of decision trees (i.e. varying from 1 to 3, 5, 10 and 15). We can find that it is beneficial to fuse regression models. When $T = 1$, SRRFf provides over 0.5 dB improvement in PSNR compared with SRRF. When $T = 5$, the improvement is around 0.3 dB. As T increases, the gap between SRRFf and SRRF slowly drops, but remains around 0.2 dB. In the following experiments, we set the number of decision trees in a random forests to $T = 5$.

5.3. Hierarchical Random Forests for SR

This section compares the proposed SRHRF method with the SRHDT method [10]. To perform training for SRHDT and SRHRF with 3 layers, we divided the 91 training images into three subsets. Each subset contains around 30 images for training a single layer. We evaluate these two methods on *Set5* for upscaling factor 2. For SRHDT, we set $N_{min} = 50D_L$ as suggested in [10] to learn stable decision trees which have low variance error. Using the same amount of training data, each decision tree in SRHRF has more leaf nodes than the decision trees in SRHDT as the N_{min} was set to be much smaller. Figure 5 shows the average PSNR in dB of the super-resolved images using SRHDT and SRHRF with and without model fusion. SRHRF with model fusion offers around 0.3 dB improvement, while the average PSNR of SRHDT with fusion model is only 0.15 dB higher than that without model fusion. When model fusion is applied in both methods, SRHRF has around 0.4 dB higher PSNR than SRHDT. In the following experiments, the proposed SRHRF method uses fusion model and is with three layers of random forests by default.

5.4. Self-Example Random Forests

There are two important parameters for self-example random forests, i.e. down-sampling factor s_d and number of images for the image pyramids N_d. The factor s_d controls the similarity between the self-example training data and the SRHRF enhanced image as well as the amount of training data available in the image pyramid pair. The image number N_d is also related to the training data size. Based on hierarchical random forests learned in previous section, a SERF with 3 decision trees is learned using the self-example training data extracted from the image pyramid pair. The performance of SERF is evaluated by varying s_d and N_d on *Set5* for upscaling factor 2. In Figure 6, the green line is the baseline result provided by SRHRF. When s_d is too small, we obtain deteriorated results. This is due to a mismatch between the structures in the down-sampled images and the super-resolved image by SRHRF. The average PSNR increases fast between $s_d = 0.5$ and $s_d = 0.7$ and plateaus after 0.7. It is also clear that more images in the pyramids give better performance. Although $N_d = 8$ provides slightly better result, it requires longer online learning time. Based on the numerical results, we set $s_d = 0.8$ and $N_d = 4$ for $s = 2$ to balance the reconstruction quality and processing time. For upscaling factor 4, similar trends can be observed and we set $s_d = 0.8$ and $N_d = 4$.

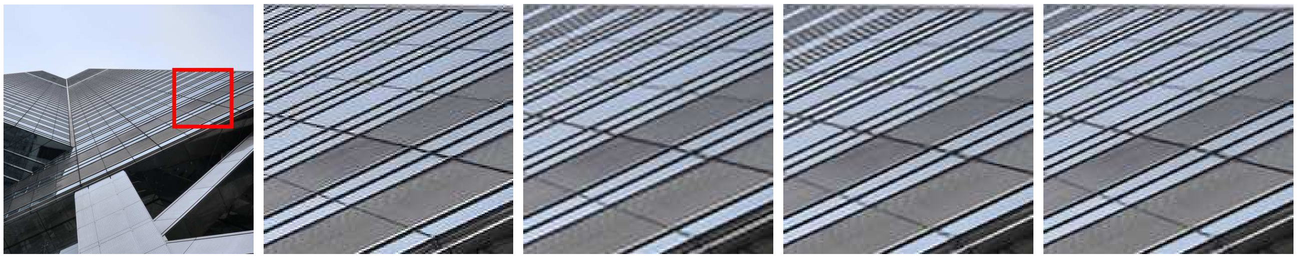
5.5. Comparison with State-of-the-art Methods

We compare our proposed methods i.e. SRHRF and SRHRF+ with the state-of-the-art SISR methods, including A+ [21], RFL [17], SRHDT [10], transformed self-example (SelfEx) [8], SRCNN [4]. We note that A+ [21] and RFL [17] are recent representative work of the "first-classification-then-regression" based methods. SRHDT method [10] cascades decision trees to boost the performance. SelfEx [8] is the state-of-the-art self-example based method. SRCNN [4] is the recent deep-learning based method. All the comparison results are obtained by using publicly available implementations with default settings.

As the 91 training images have been divided into multiple subsets and the training data in each subset is small, the performance limit of the hierarchical methods is not fully exploited. For example, the SRRFf method in Section 5.2 which is trained using the whole 91 training images provides over 0.2 dB higher PSNR than SRHRF method with

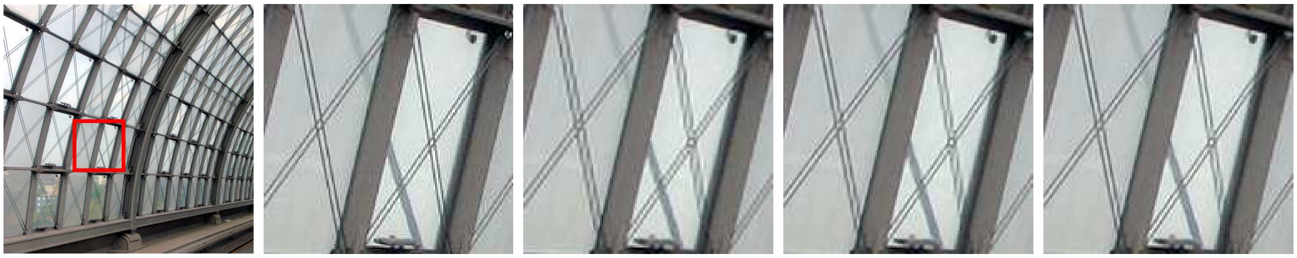
			Bicubic	A+ [21]	RFL [17]	SRHDT [10]	SRCNN [4]	SelfEx [8]	SRHRF	SRHRF+
Set5	×2	PSNR	33.66	36.54	36.54	36.92	36.66	36.49	37.19	**37.29**
		SSIM	0.9299	0.9544	0.9537	0.9546	0.9542	0.9537	0.9568	**0.9574**
	×4	PSNR	28.42	30.28	30.14	-	30.48	30.31	30.74	**30.82**
		SSIM	0.8104	0.8603	0.8548	-	0.8628	0.8619	0.8706	**0.8710**
Set14	×2	PSNR	30.24	32.28	32.26	32.67	32.42	32.22	32.85	**32.91**
		SSIM	0.8683	0.9056	0.9040	0.9069	0.9063	0.9034	0.9097	**0.9104**
	×4	PSNR	26.00	27.32	27.24	-	27.49	27.40	27.69	**27.74**
		SSIM	0.7027	0.7491	0.7451	-	0.7503	0.7518	0.7574	**0.7582**
Urban100	×2	PSNR	26.86	29.20	29.11	29.75	29.50	29.54	30.13	**30.77**
		SSIM	0.8395	0.8938	0.8904	0.8985	0.8946	0.8967	0.9038	**0.9110**
	×4	PSNR	23.14	24.32	24.19	-	24.52	24.79	24.70	**25.10**
		SSIM	0.6577	0.7183	0.7096	-	0.7221	0.7374	0.7305	**0.7422**

Table 1: PSNR (dB) and SSIM of different SISR methods on *Set5*, *Set14* and *Urban100* with upscaling factor 2 and 4.



(a) Ground-truth HR (b) HR (PSNR) (c) SelfEx [8] (25.61dB) (d) SRHRF (24.91dB) (e) SRHRF+ (27.09dB)

Figure 7: Reconstructed HR images of "*img059*" from *Urban100* by different SISR methods for *upscaling factor 2*.



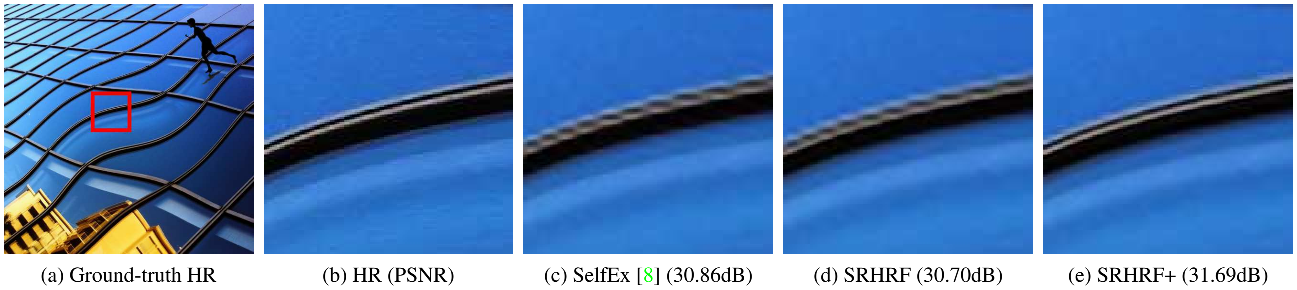
(a) Ground-truth HR (b) HR (PSNR) (c) SelfEx [8] (30.04dB) (d) SRHRF (31.18dB) (e) SRHRF+ (31.85dB)

Figure 8: Reconstructed HR images of "*img002*" from *Urban100* by different SISR methods for *upscaling factor 2*.

1 layer of random forests. In this section, 200 training images from Berkeley Segmentation Dataset (BSD) [15] will be added to enrich the training data as in [17]. Image rotations and flipping are also performed for data augmentation.

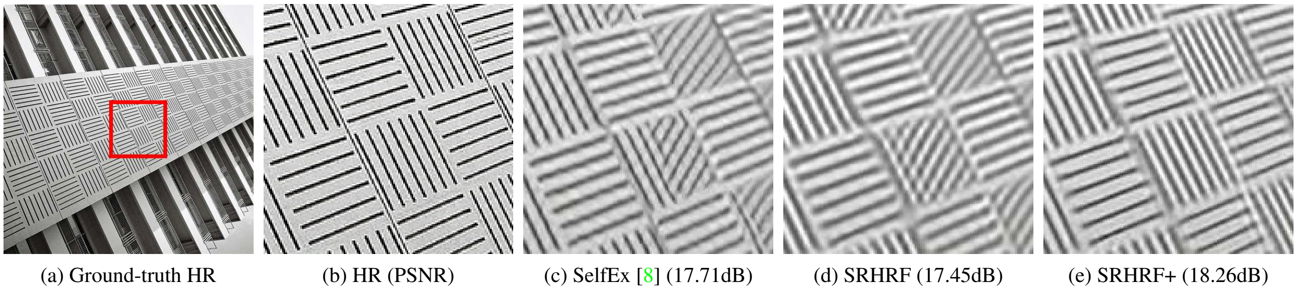
Table I reports the average PSNR (dB) and SSIM of different SISR methods evaluated on *Set5*, *Set14*, and *Urban100* for upscaling factor 2 and 4. SelfEx [8] has on par or slight lower performance than those example-based SISR methods on *Set5* and *Set14,* while achieves state-of-the-art performance on *Urban100*. This indicates that the example-based SISR methods have more stable performance on natural scene images, while self-example based methods can take advantage of the input images and achieve superior results on images with self-similar patterns. Our proposed SRHRF+ method tries to make use of the good features of both example-based methods and self-example based methods. Table I shows that the proposed SRHRF method achieves the higher PSNR and SSIM in most scenarios compared to other methods. SRHRF+ further improves SRHRF by different amounts in dB depending on the content of the image. For natural images in *Set5* and *Set14*, the average PSNR improvement is around 0.1 dB. For the urban images with rich self-similarity in *Urban100*, the SRHRF+ method provides 0.6 dB and 0.4 dB higher PSNR than our baseline SRHRF method for upscaling factor 2 and 4, respectively.

Figure 7 and Figure 8 present the subjective comparisons



(a) Ground-truth HR (b) HR (PSNR) (c) SelfEx [8] (30.86dB) (d) SRHRF (30.70dB) (e) SRHRF+ (31.69dB)

Figure 9: Reconstructed HR images of "*img082*" from *Urban100* by different SISR methods for *upscaling factor 4*.



(a) Ground-truth HR (b) HR (PSNR) (c) SelfEx [8] (17.71dB) (d) SRHRF (17.45dB) (e) SRHRF+ (18.26dB)

Figure 10: Reconstructed HR images of "*img092*" from *Urban100* by different SISR methods for *upscaling factor 4*.

between SelfEx [8] and our proposed methods on testing images from *Urban100* for upscaling factor 2. In Figure 7, SelfEx restores fine structures but reconstructs less accurate large structures (the thick lines near right bottom). It achieves 0.7 dB higher PNSR than our baseline method SRHRF. By exploiting self-examples, our enhanced method SRHRF+ recovers even better fine structures than SelfEx and achieves 2.18 dB improvement over SRHRF. In Figure 8, SRHRF has less ringing artifacts and already offers higher PSNR than SelfEx. The PSNR of SRHRF+ is 0.78 dB and 1.71 dB higher than that of SRHRF and SelfEx, respectively. Visually, there are still some artifacts around the lines in the result by SelfEx and SRHRF, while most visible artifacts are removed in the result of SRHRF+.

Figure 9 and Figure 10 further show the subjective comparisons for upscaling factor 4. In Figure 9, both SelfEx and SRHRF reconstruct blurry and discontinuous edges, while the reconstructed image by SRHRF+ is very close to the ground-truth HR image. The PSNR improvement of SRHRF+ over SRHRF is around 1 dB. In Figure 10, the fine structures are severely aliased after down-sampling by a factor 4. It is difficult for example based methods to remove the aliasing artifacts due to high ambiguity. Self-example based methods can alleviate the aliasing by exploiting self-similarity. The reconstructed patterns by SelfEx is better than that of SRHRF. However, both of them are still in the aliased direction. SRHRF+ produces sharper edges and is able to correct the aliasing artifacts in the result of SRHRF using the internal training data. This validates the effectiveness of our proposed self-example enhanced approach.

6. Conclusions

In this paper, we proposed SRHRF+ method which exhibits the stability of the example-based SISR methods and the adaptability of the self-example based SISR methods. The hierarchical random forests approach clearly demonstrates its superiority in learning statistical priors from external training images over random forests and hierarchical decision trees for image super-resolution. The self-example random forests approach exploits self-similarity from an image pyramid pair which is constructed using the down-sampled super-resolved image instead of the input LR image and further improves the baseline results. Overall, our proposed SRHRF+ method achieves state-of-the-art SISR performance and shows excellent results for images with rich self-similar patterns.

For future work, the self-example enhancement idea can be applied in other SISR methods. More efficient training strategies can be further exploited, for example, re-using the existing random forests structure and updating the leaf node using self-exemplar training data.

Acknowledgements

This work was supported by the EU H2020 TERPSICHORE project "Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects" under the grant agreement 691218.

References

[1] 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*, 23(12):5334–5347, 2014.

[2] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[3] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004.

[4] 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*, pages 184–199. Springer, 2014.

[5] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[6] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013.

[7] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 349–356. IEEE, 2009.

[8] 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*, pages 5197–5206, 2015.

[9] J.-J. Huang and W.-C. Siu. Practical application of random forests for super-resolution imaging. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2161–2164. IEEE, 2015.

[10] J.-J. Huang and W.-C. Siu. Learning hierarchical decision trees for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[11] J.-J. Huang, W.-C. Siu, and T.-R. Liu. Fast image interpolation via random forests. *IEEE Transactions on Image Processing*, 24(10):3232–3245, 2015.

[12] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[13] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1645, 2016.

[14] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE transactions on pattern analysis and machine intelligence*, 32(6):1127–1133, 2010.

[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423. IEEE, 2001.

[16] J. Salvador and E. Pérez-Pellitero. Naive bayes super-resolution forest. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 325–333, 2015.

[17] S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3791–3799, 2015.

[18] S. Schulter, C. Leistner, P. Wohlhart, P. M. Roth, and H. Bischof. Alternating regression forests for object detection and pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 417–424, 2013.

[19] Y. Tian, F. Zhou, W. Yang, X. Shang, and Q. Liao. Anchored neighborhood regression based single image super-resolution from self-examples. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2827–2831. IEEE, 2016.

[20] 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*, pages 1920–1927, 2013.

[21] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126. Springer, 2014.

[22] X. Wei and P. L. Dragotti. FRESH: FRI based single image super resolution algorithm. *IEEE Transactions on Image Processing*, 25(8):3723–3735, 2016.

[23] C.-Y. Yang and M.-H. Yang. Fast direct super-resolution by simple functions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 561–568, 2013.

[24] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1059–1066, 2013.

[25] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.

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