Article

Learning Weighted Forest and Similar Structure for Image Super Resolution

Ziwei Lu [1,2], Chengdong Wu [3,]* and Xiaosheng Yu [3,]*

[1] College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; luziwei530@126.com

[2] School of Computer and Communication Engineering, Liaoning Shihua University, Fushun 113001, China

[3] Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China

* Correspondence: wuchengdong@ise.neu.edu.cn (C.W.); yuxiaosheng@ise.neu.edu.cn (X.Y.); Tel.: +86-130-1930-6281 (C.W.)

check for updates

Featured Application: The work has many potential applications, such as small target detection, remote sensing investigation, video surveillance and medical imaging.

Abstract: Image super resolution (SR) based on example learning is a very effective approach to achieve high resolution (HR) image from image input of low resolution (LR). The most popular method, however, depends on either the external training dataset or the internal similar structure, which limits the quality of image reconstruction. In the paper, we present a novel SR algorithm by learning weighted random forest and non-local similar structures. The initial HR image patches are obtained from a weighted forest model, which is established by calculating the approximate fitting error of the leaf nodes. The K-means clustering algorithm is exploited to get a non-local similar structure inside the initial HR image patches. In addition, a low rank constraint is imposed on the HR image patches in each cluster. We further apply the similar structure model to establish an effective regularization prior under a reconstruction-based SR framework. Comparing with current typical SR algorithms, the results of comprehensive experiments implemented on three publicly datasets show that peak signal-to-noise ratio (PSNR) has been effectively promoted by the presented SR approach, and a better visual effect has been realized.

1. Introduction

Image SR technique aims at reconstructing a HR image from one or multiple LR images. The SR technique is closely related to computer vision problems with extensive applications ranging from astronomical and medical imaging to video surveillance. It has received widespread attention since put forward by Tsai and Huang [1]. The study focuses on single image SR approaches. Mainstream approaches for SR reconstruction can be subdivided into three categories: Interpolation-based approaches [2–6], reconstruction-based approaches [7–11], and example learning-based approaches [12–29].

The SR approaches based on interpolation generally exploit the kernel function [2,3] to estimate a large number of unknown pixels in HR meshes. Although they offer simplicity and rapidity, the restored HR image usually has obvious blurring and jagged artifacts. Therefore, the performance of such approaches is unattractive in reality.

The reconstruction-based approaches generally assume that the observation LR image results from multiple degeneration factors like warping, blurring, down-sampling, and noise [9]. Since a

LR pixel can correspond to multiple HR pixels, the SR reconstruction is an ambiguous and ill-posed problem. For the purpose of obtaining a credible and unique HR image, certain prior knowledge such as edge-directed prior [11] needs to be added to the reconstructed image. However, these approaches cannot impose adequate novel details on the target HR image if up-sampling factor is greater than 2, and result in quality decline of the resultant image.

Example learning-based approaches usually exploit a variety of machine-learning algorithms to obtain a mapping relationship from LR to HR image by using training dataset including millions of LR-HR exemplar patch pairs [12]. Using the co-occurrence LR-HR patches as priors, more high frequency details can be reconstructed and imposed on the LR test image to promote the quality of the reconstructed image. However, if external training dataset lacks a correlation with the LR input image, example learning-based methods tend to bring unpleasing artifacts to the restored image.

For single image SR reconstruction, only one LR input is available. The degradation process from the original HR image to the LR image is formulated as:

$$X = DBY + n \qquad (1)$$

where X is observation LR image, Y expresses objective HR image, B represents blurring matrix, D stands for the down-sampling matrix, and n is noise. Because of factors such as down-sampling, blurring and noise, an input LR image is able to match multiple varying HR images, so the issue of SR reconstruction is seriously underdetermined. Thus, to accurately realize estimation for HR images, some priors or regularization constraint items need to be introduced into the reconstruction process. The reconstruction prediction for SR based on regularization constraint is formulated as:

$$\hat{Y} = \underset{Y}{\mathrm{argmin}}\gamma\|X - DBY\|_F^2 + R(Y) \qquad (2)$$

where the first term $\|X - DBY\|_F^2$ denotes the reconstruction error, and $\|\cdot\|_F^2$ stands for Frobenius norm. The second term $R(Y)$ denotes the regularization constraint item. Here γ is the balance parameter, which is used to adjust weight between the first and second items.

Example learning-based SR can successfully obtain high-frequency details by depending on the external image training dataset. Simultaneously, the reconstruction-based SR algorithm is able to successfully and effectively utilize reasonable prior knowledge. To obtain better reconstructed quality, we propose a learning weighted forest from an external training dataset and use a similarity structure inside the initial HR image patches as a reasonable regularization prior. Specially, we first learn a random forest, and compute the approximate fitting error of the predictive HR patches so as to determine the weighted model in each leaf node, and then obtain the weighted predicted patches as the initial HR image patches. Next, K-means clustering is performed on the initial HR patches according to structural similarity. In addition, a low rank constraint is imposed on the HR image patches in each cluster. We further apply a similar structure model to establish an effective regularization prior in SR framework based on reconstruction. To sum up, the contribution of this work is embodied in three aspects:

- We construct an error model from the approximate fitting error of each leaf node, and propose a weighted forest SR algorithm, which promotes the performance of the original SR forest method.
- Clustering and low rank constraint are performed according to the structural similarity of the initial HR image patches, and the similarity information is encapsulated into a regularization term.
- Comprehensive experimental results on quantitative and qualitative benchmarks indicate that our SR method is superior to other competing methods.

If the non-local similarity structure and low rank constraint priors are incorporated into the initial HR patches, the quality of restored image can be further improved. The analogous priors have also been exploited in Jiang's approach [30] and Zeng's approach [31]. In contrast to the above two methods,

the innovation of our approach concentrates on how to reasonably encapsulate the weighted forest and the similarity priors into a reconstruction-based SR framework.

The structure of the paper is arranged as follows: In Section 2, main work related to the paper is briefly reviewed. In Section 3, the SR algorithm based on random forest is introduced. In Section 4, the proposed SR algorithm is described in detail. In Section 5, experimental results of comparison with competing SR methods are conducted. Section 6 summarizes the whole paper.

2. Related Work

The SR technique is a significant computer vision problem with a long history. Several works of literature have been published on SR research. We briefly survey the works most relevant to our approach.

Example learning-based-approaches can be classified as external and internal learning methods, according to the source of dataset. The two classical methods about external learning SR are those based on dictionary and regression. Dictionary learning-based methods are typically established on sparse coding [32]. Yang et al. [15] adopted sparse representation formulation to jointly identify a compact and powerful LR-HR dictionary pair with shared coefficients for sparse reconstruction. Zeyde et al. [16] promoted a method in [15], which reduced feature dimensionality by PCA, performed sparse coding using a K-SVD algorithm in [33] and Orthogonal Matching Pursuit, and further improved the efficiency of dictionary training and inference. However, these approaches are confronted by computational bottlenecks. Lately, several effective regression learning-based methods have obtained extensive attention. These approaches directly study the mapping relationship between LR patches and relevant HR patches. Timofte et al. proposed a quick and effective SR algorithm named anchored neighborhood regression (ANR) [17], which exploited ridge regression to understand sample neighborhoods offline and pre-calculate the mapping to transform LR patches into HR space, further improving a variant named A+ [18]. Yang et al. [19] sought to solve the complicated regression by partitioning sample space into multiple subspaces, and then selected enough samples to study an effective regression for each subspace. Dai et al. [20] proposed a similar method to jointly train a group of local regressions and resolve each input LR patch via its most suitable regression. However, these methods need to set the number of clusters, then perform regression, which influence the balance of the up-sample quality and inference time. Schulter et al. [21] proposed to train local linear regression between LR and HR patches by random forest, which establishes a regularization quality evaluation function operating on input and out space simultaneously, and complete effective prediction without increasing inference time. In addition, Dong et al. [22] created a novel deep learning algorithm for SR. This method trains end-to-end mapping to directly convert LR patches into the HR domain, demonstrating perfect performance.

Internal database-driven SR methods usually employ statistical priors [23], which have shown strong capacity to resolve the SR problem. Protter et al. [24] applied the nonlocal means filter to restore video sequences with normal motion modes. Another classical approach was presented by Glasner et al. [25], who used multi-scale similarity to solve the SR problem, and combined the multi- and single image reconstruction under the unified framework. Freedman et al. [26] further demonstrated that self-similarity patches usually repeat in a finite spatial neighborhood; therefore, computational acceleration could be gained. Michaeli et al. [27] exploited self-similarity to jointly estimate fuzzy kernels and HR images. Similar to [25], Zhang et al. [28] employed similar redundancy to restore HR image from only one LR input image. To restore missing details, the algorithm acquired similarity across various scales and estimated the mapping relationship between LR and HR patch pairs using neighbor embedding (NE) [13]. Huang et al. [29] exploited geometric deformation to expand patch space, and achieved the closest patch search by plane positioning and perspective geometry detection in the scene.

Recently, a combination model based on reconstruction and example learning have drawn more attention to solve SR problems. Dong et al. [34] proposed to apply centralized sparse representation for image reconstruction under a unified adaptive framework. Wang et al. [35] presented another

combing approach by using sparse Gaussian process regression and nonlocal mean filter, which shows perfect performance.

In summary, the external example learning-based methods can receive novel high-frequency information from the training dataset. However, the correlation degree between varying test samples and training samples is different, so there is no guarantee that each LR patch can obtain a matching HR patch; thus, the reconstructed image tends to be too smooth. The internal learning-based method can utilize similarity redundancy to get the intrinsic prior of an image structure, but for some image patches without the repeated structure or the irregular texture, the reconstructed image is likely to generate texture artifacts. Motivated by the assembling method of [34,35], we propose to learn weighted random forest and similar structures for image SR, and build a joint optimization model, which synthesizes external example learning and internal similar structure prior under a reconstruction-based SR framework.

3. The Image Super Resolution Based on Random Forest

Schulter et al. [21] proposed to study mapping relationship from LR to HR patches based on random forest. The training dataset was partitioned into N groups of the training sample. Let $X = \{x_l{}^i | i = 1, 2, \ldots, N\}$, $Y = \{y_h{}^i | i = 1, 2, \ldots, N\}$, which denotes LR patches and corresponding HR patches training datasets, respectively. The estimation of HR patches can be considered the local linear regression problem for LR patches as below:

$$y_h = W(x_l) \cdot x_l \tag{3}$$

where x_l denotes the LR patch, y_h denotes the corresponding HR patch, and $W(\cdot)$ is the local linear regression function.

The least square model can be adopted to study the local linear regression function as below:

$$\underset{W(x_l)}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| y_h{}^i - W(x_l{}^i) \cdot x_l{}^i \right\|_2^2 \tag{4}$$

This optimization is implemented using the SR forest algorithm, which utilizes a splitting function to recursively split the input data into disjoint subspaces to obtain a tree structure, and establishes the linear regression function to express data dependence between LR and HR patches.

When random forest is trained, the target HR image patch y_h is estimated as below:

$$y_h = \frac{1}{T} \sum_{t=1}^{T} m_{l(t)}(x_l) \tag{5}$$

where T denotes the number of decision tree, and $m_{l(t)}$ denotes the local linear regression function for the sampled LR image patches x_l in the t-th tree.

4. The Proposed Super Resolution Method

Our SR method consists of four parts: (1) Studying on the weighted predictive model, which guarantees the smaller error with the bigger weighted coefficient; (2) achieving a similar structure clustering of the initial SR image patch and adding a low rank constraint for each similar structure clustering; (3) accomplishing a reconstruction-based SR framework, which integrates weighted forest and a similar structure, and transforms the similarity structure model into an effective regularization term for the objective HR image; and (4) summarizing the proposed SR algorithm.

4.1. The Weighted Predictive Model based on Random Forest

As there are T trees in random forest, for each LR test patch, T different candidate HR patches are obtained. Considering that these LR test image patches have different degrees of fitting error

in each prediction model, the application of weighted prediction can estimate HR image patches more accurately.

Firstly, use the k-d tree method to quickly search the K-nearest neighbors' patches for each LR test image patch x_l in its leaf node. The number of neighborhoods is expressed as $N_K(x_l) = \{n_1, n_2, \ldots, n_k\}$. Secondly, calculate the cumulative fitting error of its K-nearest neighbor image patches as the approximate error of the test image patch in the local regression model. Then, perform the weighted prediction to estimate the HR image patch according to the approximate error in each regression model. The predictive HR image patch is expressed as follows:

$$\widetilde{y_{h,t}} = m_{l(t)}(x_l) \tag{6}$$

where $\widetilde{y_{h,t}}$ is the estimation HR patch from the *t-th* tree. Its SR reconstruction error is described as below:

$$e_t = \|\widetilde{y_{h,t}} - y_h\|^2 \tag{7}$$

then the cumulative fitting error of the K-nearest neighbor image patches in leaf node for LR test patch x_l is approximated as:

$$e_{N_K(t)} = \frac{\sum\limits_{k=1}^{K} \frac{1}{k} e_{t,n_k}}{\sum\limits_{k=1}^{K} \frac{1}{k}} \tag{8}$$

where $\frac{1}{k}$ can be exploited to adjust the contribution of neighbors to calculate cumulative error, e_{t,n_k} denotes the reconstruction error resulting from the kth neighbor image patch, and K stands for the number of neighbor image patches.

Finally, implement the weighted prediction for the desired HR image patch according to the cumulative fitting error sum of the K-nearest neighbor image patch in each local regression model:

$$y_h = \frac{\sum\limits_{t=1}^{T} w_t m_{l(t)}(x_l)}{\sum\limits_{t=1}^{T} w_t} = p(x_l) \tag{9}$$

where $p = \dfrac{\sum\limits_{t=1}^{T} w_t m_{l(t)}}{\sum\limits_{t=1}^{T} w_t}$ represents the weighted projection model, and $w_t = \dfrac{\frac{1}{e_{N_K(t)}}}{\sum\limits_{t=1}^{T} \frac{1}{e_{N_K(t)}}}$ satisfied $\sum\limits_{t=1}^{T} w_t = 1$.

The weighted prediction mode ensures that each local regression model with a smaller error has a higher weight, which improves accuracy of the weighted prediction.

4.2. Similar Structure Clustering and Low Rank Constraint

It is well known that nature images contain many local similar structures. In order to express the inherent geometry structures well, we propose to partition the obtained HR image patches into several clusters so that each cluster has similar structures. For this purpose, we exploit a structural clustering conception enlightened by [36] and further implement a low rank constraint. The feature of local image patch is represented by normalized pixel intensity. For the obtained HR image Y, let y_h^i express image patch vector in the *i-th* two dimensional position. A multi-cluster union consisting of all feature vectors is represented as follows:

$$Y = \bigcup_{k=1}^{C} \left\{ y_h^i \middle| i \in \Omega_k \right\} \tag{10}$$

where C is cluster number, and Ω_k denotes the index set *in k-th* cluster. The k-means clustering operates quickly, and is able to accurately partition the initial HR patches to appropriate subsets, so k-means

clustering is adopted to realize clustering. In the process of clustering, we utilize l_2-norm as distance metric and minimize inter-cluster variance to partition the obtained HR patches into multiple clusters.

$$J = \sum_{k=1}^{C} \sum_{i \in \Omega k} \left\| \widetilde{y_h^i} - \overline{y_{(k)}} \right\|_2^2 \tag{11}$$

where $\widetilde{y_h^i}$ denotes normalized vector, and $\overline{y_{(k)}}$ denotes mean vector for the *k-th* cluster. In our research, set C = 12, and the initial HR patches are partitioned into 12 clusters by 20 iterative operation.

The above clustering algorithm can partition the obtained HR image patches with similar structures into the same cluster. $F_k Y$ is defined as a matrix consisting of image patch vectors in the *k-th* cluster.

Since the image patches in the same cluster have a high degree of similarity, it can be considered that the matrix formed by each set of vectorized image patches is approximated as a low rank matrix, which is described as below:

$$F_k Y = A_k + E_k \tag{12}$$

where A_k is low rank matrix, and E_k is error matrix. Because of the difficulty in solving a low rank matrix, the problem is transformed into a kernel norm minimization problem for approximate solution, which is expressed as follows:

$$Y = \underset{Y}{\text{argmin}} \| F_k Y - A_k \|_F^2 + \lambda \| A_k \|_* \tag{13}$$

where $\| \cdot \|_*$ denotes the kernel norm. Equation (13) indicates a high similarity for patches in the *k-th* cluster.

4.3. Reconstruction-based Optimization

Expression $\widetilde{Y_k} = F_k Y$ indicates the matrix consisting of image patches vectors in the *k-th* cluster. Expression $\widetilde{X_k} = P_k x_l^k$ stands for the output of weighted predictive model of image patch vectors in the *k-th* cluster, where P_k denotes the weighted mapping matrix obtained from Equation (9), and x_l^k is the correlational LR patches. It is extended to all clusters of image patches, and combined with weighted random forest algorithm to obtain the target formula as follows:

$$\{\{\widetilde{Y_k}\}, \{A_k\}\} = \underset{\{\widetilde{Y_k}\}, \{A_k\}}{\text{argmin}} \gamma \sum_{k=1}^{C} \left\| \widetilde{Y_k} - \widetilde{X_k} \right\|_F^2 + \sum_{k=1}^{C} \left(\left\| \widetilde{Y_k} - A_k \right\|_F^2 + \lambda \| A_k \|_* \right) \tag{14}$$

Equation (14) comprehensively utilizes the extrinsic example learning and the intrinsic structural similarity prior. We perform the optimization motivated by [31], and further decompose Equation (14) into C sub-problems as follows:

$$\{\widetilde{Y_k}, A_k\} = \underset{\widetilde{Y_k}, A_k}{\text{argmin}} \gamma \| \widetilde{Y_k} - \widetilde{X_k} \|_F^2 + \| \widetilde{Y_k} - A_k \|_F^2 + \lambda \| A_k \|_* \tag{15}$$

Equation (15) represents the *k-th* sub-problem, which can be resolved by an alternate iterative algorithm. Firstly $\widetilde{Y_k}$ is fixed to solve A_k, and Equation (15) can be further expressed as:

$$A_k = \underset{A_k}{\text{argmin}} \| \widetilde{Y_k} - A_k \|_F^2 + \lambda \| A_k \|_* \tag{16}$$

Equation (16) can be solved using a singular value threshold algorithm. Then A_k is fixed to solve $\widetilde{Y}_k$. Equation (15) can also be further expressed as:

$$\widetilde{Y}_k = \underset{\widetilde{Y}_k}{\operatorname{argmin}} \gamma \|\widetilde{Y}_k - \widetilde{X}_k\|_F^2 + \|\widetilde{Y}_k - A_k\|_F^2 \tag{17}$$

Equation (17) becomes a quadratic optimization problem, and its optimal solution is:

$$\widetilde{Y}_k = (\gamma + 1)^{-1}(\gamma \widetilde{X}_k + A_k) \tag{18}$$

4.4. Summary of the Present Super Resolution Algorithm

To summarize, the SR method adopted in this paper is drawn in Algorithm 1.

Algorithm 1. The weighted forest and similar structure-based SR algorithm

---

Input: LR test image X, magnification factor S;
Output: Objective HR image Y.
Step 1: Magnify X factor S times by using bi-cubic interpolation to reach the same size with target HR image Y.
Step 2: Obtain the initial mapping relationship from LR to HR patches in each leaf node by using random forest.
Step 3: Compute the weighted coefficients according to the approximate cumulative error obtained from Equation (8) of the LR image patch in the corresponding leaf nodes of each decision tree.
Step 4: Acquire the predicted HR patches by using Equation (9).
Step 5: Divide the obtained HR image patches into C different clusters by using Equations (10) and (11) to get internal clustering of the HR image.
Step 6: Employ Equations (14)–(18) to optimize each cluster of image patches, and obtain the objective HR image Y.

---

5. Experimental Results

5.1. Datasets

For training, we exploit the standard dataset containing 91 images proposed by Yang [14]. In the test phase, we employ three standard datasets for SR quality evaluation. They are Set 5, Set 14 and B 100, including commonly used images with wide coverage.

5.2. Comparisons

We compare our algorithm with most algorithms included in [17], SR forests method (RFL) [21], and SRCNN [22]. The methods in [17] include Bi-cubic, NE+LLE [13], Zeyde's [16], GR and ANR. Bi-cubic interpolation is the benchmark comparison method in the experiment. The implementation codes of other algorithms are offered by their authors. These approaches share the same training dataset. They are compared quantitatively according to PSNR. The result shows that PSNR generally associates well with visual quality.

5.3. Experimental Setting

The size of LR patches is 3×3 pixels. If the magnified factor is set at 3 or 4, we engage a patch size of 9×9 or 12×12 pixels. In order to ensure the compatibility of each reconstructed patch, the 3×3 patches are extracted with two overlapping pixels between adjacent patches. The HR image is scaled by factor 1/3 using Bi-cubic interpolation to obtain the corresponding LR image, and then LR image is scaled three times to get a magnified image. The magnified image is still called an LR image due to a lack of high frequency information. The first- and second-order derivative of patches express an image patch. These derivatives are concatenated and their dimensions become reduced by

adopting the method used in [17]. Unless otherwise noted, the setting for our method is T = 15, K = 16, $D_{\max} = 15$, where $D_{\max}$ denotes maximum depth of the tree.

5.4. Performance

In the paper, PSNR is used as a quantitative assessment to evaluate the performance of the comparison methods. We compute PSNR on standard test datasets for all comparison methods, including our proposed method, as shown in Table 1. The results prove that PSNR from our algorithm is clearly superior to those from the compared approaches.

Table 1. Average PSNR Results for Comparison Methods.

Benchmark		Bi-cubic	NE+LLE	Zeyde	GR	ANR	RFL	SRCNN	Proposed
Set5	× 3	30.39	31.84	31.90	31.41	31.92	32.46	32.39	32.73
	× 4	28.42	29.61	29.69	29.34	29.69	30.15	30.09	30.45
Set14	× 3	27.54	28.60	28.67	28.31	28.65	29.05	29.00	29.41
	× 4	26.00	26.81	26.88	26.60	26.85	27.24	27.20	27.62
B100	× 3	27.15	27.85	27.87	27.70	27.89	28.39	28.10	28.47
	× 4	25.92	26.47	26.51	26.37	26.51	26.86	26.66	26.98

5.5. Visual Quality

Just like many other SR works, our method also uses a single luminance channel because the high-frequency chrominance changes are less sensitive than luminance changes for human visual systems. The color image is first converted from the RGB space into a YCbCr space, and then an HR output image is restored only by luminance component and directly using bi-cubic interpolation for chrominance components.

To prove the effectiveness of our algorithm, visual quality of the SR-reconstructed results of three typical test images (monarch, zebra and airplane) with up-sampling factor of 3, is displayed in Figures 1–3.

Figure 1. Visual quality comparison of a monarch image: (**a**) Original; (**b**) Bi-cubic; (**c**) Zeyde; (**d**) GR; (**e**) ANR; (**f**) NE+LLE; (**g**) SRCNN; (**h**) RFL; (**i**) Proposed.
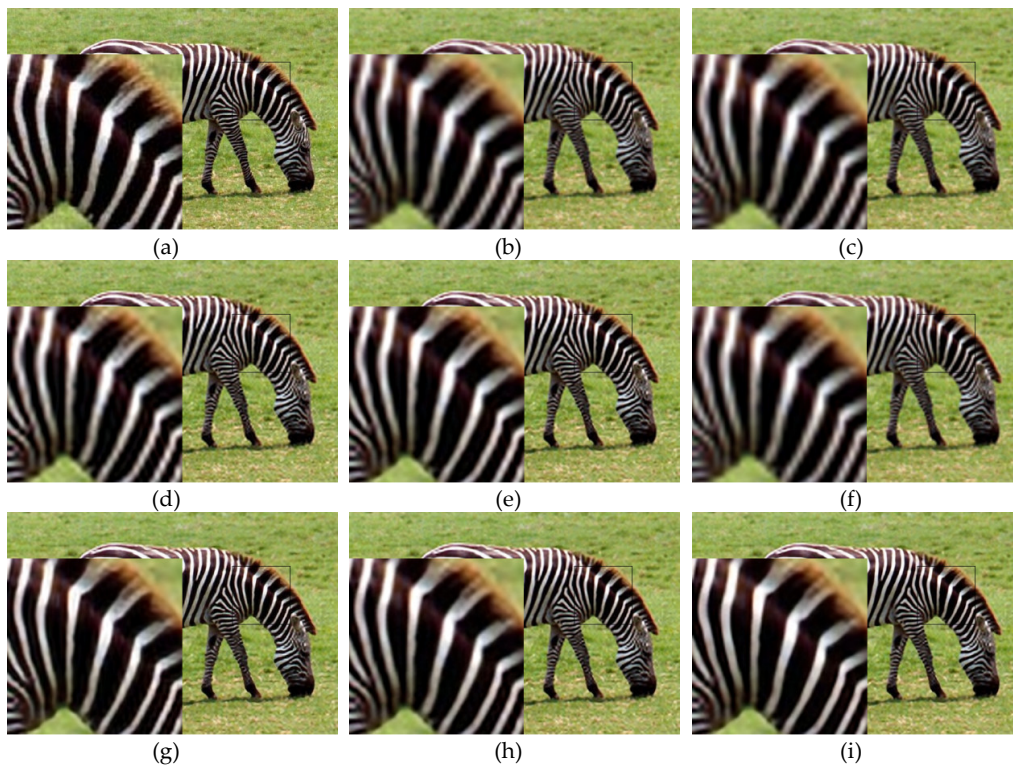
Figure 2. Visual quality comparison of a zebra image: (**a**) Original; (**b**) Bi-cubic; (**c**) Zeyde; (**d**) GR; (**e**) ANR; (**f**) NE+LLE; (**g**) SRCNN; (**h**) RFL; (**i**) Proposed.
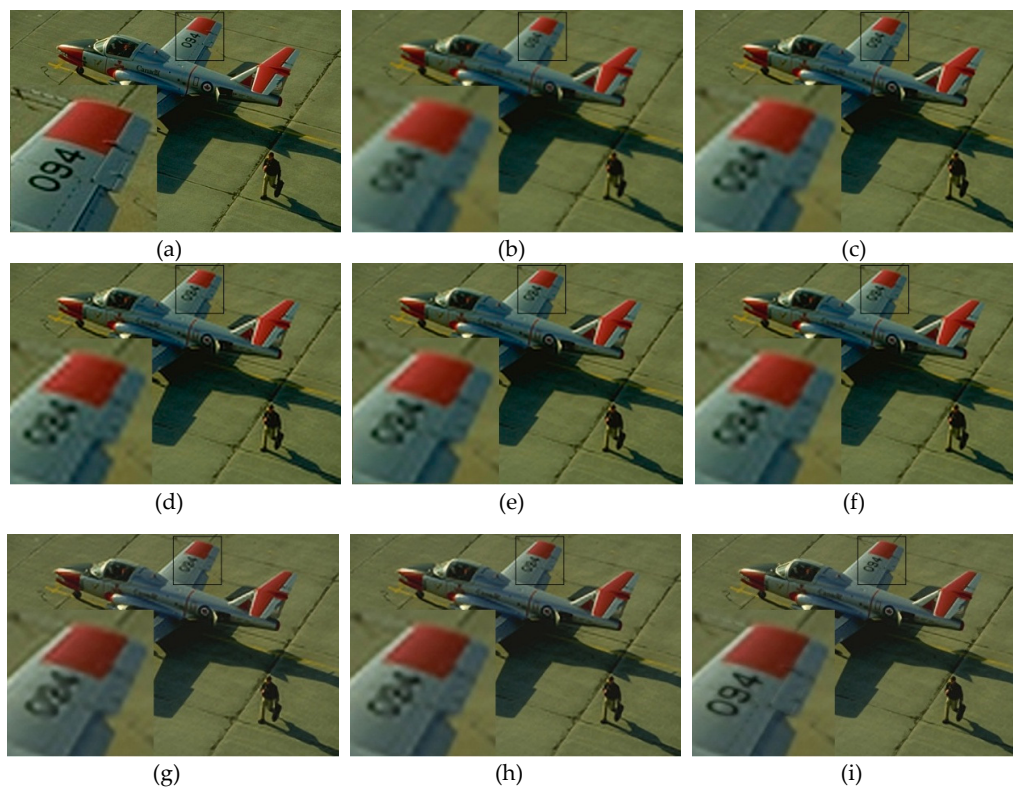
Figure 3. Visual quality comparison of an airplane image: (**a**) Original; (**b**) Bi-cubic; (**c**) Zeyde; (**d**) GR; (**e**) ANR; (**f**) NE+LLE; (**g**) SRCNN; (**h**) RFL; (**i**) Proposed.

As can be seen from the figures above, the bi-cubic interpolation usually generates obviously blurring artifacts around the edges and produces unpleasing visual effects in the texture and text regions. Zeyde's method implements SR based on sparse coding, which cannot accessibly restrain the undesirable sawtooth effects. The GR method and ANR method exploit the pre-computed regressions to transform a LR patch into an HR patch. Although they can produce more details, there exist a great many unsatisfying jaggy artifacts around the edge and texture regions. The NE+LLE method assumes that the HR patch is restored, relying on the weighted linear combination of its K nearest neighbors estimated by corresponding LR neighbors, which also produces blurring edges and unpleasing details. The SRCNN method and RFL method are capable of generating clearer edges and relatively better visual effects in comparison with the previous methods. The reconstruction results of our algorithm with regard to edge, texture, and text regions are superior to the previous six algorithms, and the visual effects are further improved. For example, in the wing regions of the monarch image in Figure 1, the white spotted edges of the butterfly reconstructed by other algorithms are blurred. In comparison to other algorithms, the present algorithm removes the blurring phenomenon, and the reconstructed white spotted edges of butterfly are much clearer. For the texture part of the zebra image in Figure 2, there are different degrees of blurring in the white stripe edge regions of the reconstructed image by Zeyde's method, GR method, ANR method and NE+LLE method. Compared with other methods, the reconstructed zebra stripes using our algorithm are much brighter and clearer. For the text portions of the airplane image in Figure 3, compared with other methods, the reconstructed image obtained in this paper is much clearer, the blurring in text characters is eliminated, and the visual effect become better. Based on the above analysis, the reconstructed image obtained by our algorithm has a noticeable edge and texture improvement, eliminates the jagged artifacts, and can reconstruct more detail information. As the PSNR shows, it is also proved that the figures from the proposed algorithm have better visual quality.

6. Conclusions

In the paper, we propose a novel image SR method by learning weighted forest and similar structures. External example learning and internal similar structure prior are synthesized under a reconstruction-based SR framework. We construct an error model based on the approximate fitting error of each leaf node to improve performance of the initial SR forest algorithm. Clustering and low rank constraint are further performed based on the structural similarity of the initially obtained HR image patches. The structural similarity information is integrated into the regularized prior term. Extensive experiments on quantitative and qualitative benchmarks demonstrate that the experimental result of our SR algorithm is obviously superior to that of other competing SR algorithms.

References

1. Tsai, R.Y.; Huang, T.S. Multi-frame image restoration and registration. In *Advance in Computer Vision and Image Processing*; JAI Press, Inc.: Greenwich, CT, USA, 1984; pp. 317–339.
2. Keys, R.G. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [CrossRef]

3. Takeda, H.; Farsiu, S.; Milanfar, P. Kernel regression for image processing and reconstruction. *IEEE Trans. Image Process.* **2007**, *16*, 349–366. [CrossRef] [PubMed]

4. Li, X.; Orchard, M.T. New edge-directed interpolation. *IEEE Trans. Image Process.* **2001**, *10*, 1521–1527. [PubMed]

5. Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans. Image Process.* **2006**, *15*, 2226–2238. [CrossRef] [PubMed]

6. Li, M.; Nguyen, T.Q. Markov random field model-based edge-directed image interpolation. *IEEE Trans. Image Process.* **2008**, *17*, 1121–1128. [PubMed]

7. Sun, J.; Zhu, J.; Tappen, M.F. Context-constrained hallucination for image super-resolution. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 231–238.

8. Tai, Y.; Tai, Y.W.; Liu, S.; Brown, M.S.; Lin, S. Super resolution using edge prior and single image detail synthesis. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2400–2407.

9. Zhang, K.B.; Gao, X.B.; Tao, D.; Li, X. Single image super-resolution with non-local means and steering kernel regression. *IEEE Trans. Image Process.* **2012**, *21*, 4544–4556. [CrossRef] [PubMed]

10. Xu, H.T.; Zhai, G.; Yang, X. Single image super-resolution with detail enhancement based on local fractal analysis of gradient. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1740–1754. [CrossRef]

11. Wang, L.F.; Xiang, S.M.; Meng, G.F.; Wu, H.Y.; Pan, C.H. Edge-directed single-image super-resolution via adaptive gradient magnitude self-interpolation. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1289–1299. [CrossRef]

12. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Gr. Appl.* **2002**, *22*, 56–65. [CrossRef]

13. Chang, N.H.; Yeung, D.Y.; Xiong, N.Y. Super-resolution through neighbor embedding. In Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; pp. 275–282.

14. Yang, J.C.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

15. Yang, J.C.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [CrossRef]

16. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In Proceedings of the 7th International Conference on Curves and Surfaces, Avignon, France, 24–30 June 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 711–730.

17. Timofte, R.; De, V.; Gool, L.V. Anchored neighborhood regression for fast example-based super-resolution. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1920–1927.

18. Timofte, R.; Smet, V.D.; Gool, L.V. A+: Adjusted anchored neighborhood regression for fast super-resolution. In Proceedings of the 12th Asian Conference on Computer Vision, Singapore, 1–5 November 2014; pp. 111–126.

19. Yang, C.Y.; Yang, M.H. Fast direct super-resolution by simple functions. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 561–568.

20. Dai, D.G.; Timofte, R.; Gool, L.V. Jointly optimized regressors for image super-resolution. *Comput. Gr. Forum* **2015**, *34*, 95–105. [CrossRef]

21. Schulter, S.; Leistner, C.; Bischof, H. Fast and accurate image upscaling with super resolution forest. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3791–3799.

22. Dong, C.; Loy, C.C.; He, K.M.; Tang, X.O. Learning a deep convolutional network for image super-resolution. In Proceedings of the 13th the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 184–199.

23. Zontak, M.; Irani, M. Internal statistics of a single natural image. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 20–25 June 2011; pp. 977–984.

24. Protter, M.; Elad, M.; Takeda, H.; Milanfar, P. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Trans. Image Process.* **2009**, *18*, 36–51. [CrossRef] [PubMed]

25. Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 349–356.

26. Freedman, G.; Fattal, R. Image and video upscaling from local self-examples. *ACM Trans. Gr.* **2011**, *30*, 474–484. [CrossRef]

27. Michaeli, T.; Irani, M. Nonparametric blind super-resolution. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 945–952.

28. Zhang, K.B.; Gao, X.B.; Tao, D.C.; Li, X.L. Single image super-resolution with multiscale similarity learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 1648–1659. [CrossRef]

29. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.

30. Jiang, J.J.; Ma, X.; Chen, C.; Lu, T.; Wang, Z.Y.; Ma, J.Y. Single image super-resolution via locally regularized anchored neighborhood regression and nonlocal means. *IEEE Trans. Multimed.* **2017**, *19*, 15–26. [CrossRef]

31. Zheng, X.T.; Yuan, Y.; Lu, X.Q. Single image super-resolution restoration algorithm from external example to internal self-similarity. *Acta Opt. Sin.* **2017**, *37*, 64–70.

32. Olshausen, B.A. Sparse Coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.* **1997**, *37*, 3311–3325. [CrossRef]

33. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [CrossRef]

34. Dong, W.S.; Zhang, L.; Shi, G.M.; Wu, X.L. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. Image Process.* **2011**, *20*, 1838–1857. [CrossRef]

35. Wang, H.J.; Gao, X.B.; Zhang, K.B.; Li, J. Fast single image super-resolution using sparse Gaussian process regression. *Signal Process.* **2017**, *134*, 52–62. [CrossRef]

36. Zhang, K.B.; Li, J.; Wang, H.J.; Liu, X.P.; Gao, X.B. Learning local dictionaries and similarity structures for single image super-resolution. *Signal Process.* **2018**, *142*, 231–243. [CrossRef]