Single Image Super-Resolution via Internal Gradient Similarity

Yang Xian and Yingli Tian*

The Graduate Center and the City College of New York,

City University of New York, New York, NY 10016 USA

Email: yxian@gradcenter.cuny.edu, ytian@ccny.cuny.edu.

Abstract

Image super-resolution aims to reconstruct a high-resolution image from one or multiple low-resolution images which is an essential operation in a variety of applications. Due to the inherent ambiguity for super-resolution, it is a challenging task to reconstruct clear, artifacts-free edges while still preserving rich and natural textures. In this paper, we propose a novel, straightforward, and effective single image super-resolution method based on internal across-scale gradient similarity. The low-resolution gradients are first upsampled and then fed into an optimization framework to construct the final high-resolution output. The proposed approach is able to synthesize natural high-frequency texture details and maintain clean edges even under large scaling factors. Experimental results demonstrate that out method outperforms existing single image super-resolution techniques. We further evaluate the super-resolution performance when both internal statistics and external statistics are adopted. It is demonstrated that generally, internal statistics are sufficient for single image super-resolution.

Keywords — Image Super-Resolution, Patch Similarity, Across Scale, Gradient Similarity, Optimization

* Corresponding author. E-mail: ytian@ccny.cuny.edu.

1. Introduction

Image super-resolution (SR) is to predict a fine-resolution image from one or multiple coarse-resolution image(s). It is a fundamental operation in image editing software and plays a crucial role in a variety of applications such as video surveillance, desktop publishing, movie restoration, and object tracking in satellite images.

Image SR techniques can be applied to regular 2D images and depth images [5, 18, 20, 21, 25]. Broadly speaking, within the scope of 2D images, SR tasks can be divided into two categories: multi-image SR and single-image SR. Multi-image SR methods [1, 3, 8, 10, 17, 24, 27] utilize the non-redundant information of multiple frames of the same scene to reconstruct one fine-resolution image. Single image SR, on the other hand, only has one low-resolution input image at disposal which leads to a numerically ill-posed problem. It is possible to generate numerous high-resolution images given the same low-resolution image. Therefore, single image SR task relies on strong image priors or assumptions to eliminate the ambiguities and to finalize a visually pleasing output image among all the possible candidate solutions.

Single image SR approaches can be categorized into three classes: interpolation-based, example-based and reconstruction-based methods. Interpolation-based methods are based on data invariant linear filtering. Commonly used interpolation-based methods such as bilinear and bicubic interpolations are simple and efficient to obtain upsampled images and thus are widely used in the commercial software. These linear interpolation-based SR methods hinge on the inaccurate assumption that natural images are always smooth. However, with discontinuities existing in natural images, the generated results are over-smoothed and with obvious visual artifacts such as blurring, aliasing and jaggies. More sophisticated interpolation-based methods were proposed [22, 28] to suppress artifacts and restore sharper edges compared with interpolation with simple linear filters.

Example-based image upscaling methods were proposed by Freeman *et al.* [12, 13] to learn the relationship between high-resolution images and their corresponding low-resolution versions through an external image dataset. In general, with lack of relevance between testing images and a universal training dataset, the produced results are noisy with irregularities along curved edges. The performance may be improved with an increment in the size of the external image dataset. However, it would lead to heavy computational cost as

well as the increasing ambiguity among patch correspondences [39]. Various methods have been proposed later after Freeman to improve the SR performance and the computational speed. Coupled high-resolution and low-resolution dictionaries [15, 31, 35, 36, 38, 39] are popular representations for the external patch exemplars where optimization techniques could apply. Yang *et al.* [35, 36] learnt a compact dictionary based on sparse signal representation which allows the possibility to adaptively choose the most relevant reconstruction neighbors. Built upon Yang's work, Zeyde *et al.* [38] introduced several modifications to further improve the execution speed. Timofte *et al.* [31] temped to combine the benefits of both neighbor embedding [2, 4] and sparse coding. Zhu *et al.* [39] proposed a method based on deformable patches which lead to a more "expressive" dictionary without increasing the size of the dictionary. Deep network has recently been explored to structure external example-based learning [6, 7].

With all the attempts made in external example-based image SR methods, the inadequate relevance between certain testing images and the universal training dataset remains unsolved. Self-example-based image SR provides a solution to build a tailored 'training dataset' for each input low-resolution image. Self-example-based image SR methods [11, 14, 37] have been proposed based on the observation that for small image patches in a natural image, self-similarities exist within the image itself and across different resolutions. Glasner *et al*. [14] proposed a patch searching scheme based on a patch pool formed with internal patches collected through a pyramid structure with only the input image at different resolutions. Freedman *et al.* [11] utilized a real time multi-step coarse-to-fine algorithm which adopts a local search instead of a global search. Yang *et al.* [37] combined learning from self-examples as well as an external dataset into a regression model based on in-place examples. These patch-based approaches are capable of generating natural-looking textures estimated from across-scale self-similarities with/without the assistance of external statistics under small magnification factors. However, it is difficult to handle the visual artifacts introduced during the estimation process especially when the upscaling factor is relatively large due to the lack of a robust global constraint.

Reconstruction-based image SR methods tend to form global constraints to enforce the fidelity between the predicted high-resolution image and the provided low-resolution image. In [26], Shan *et al.* built a feedback-control framework that enforces the output image to be consistent with the input image when downscaling to the input resolution. Recently, reconstructed gradient profile has been a popular prior utilized during the

restoration of the target HR image [9, 29] due to its heavy-tailed distribution [16]. Fattal [9] proposed an image SR method which generates the gradient field of the target HR image other than determining its pixel intensities directly. Reconstruction of the high-resolution gradient field is based on a statistical edge dependency relating certain edge features of two different resolutions. Sun *et al.* [29] modeled the image gradient by a parametric profile model. A gradient field transformation was learnt to constrain the gradients of the HR image given the low-resolution gradients. These reconstruction-based approaches are often referred to as "edge-directed SR" [30] and are capable of creating sharp edges even under large magnification factors. However, they tend to produce "unnatural" or "unrealistic" patterns within detailed texture regions due to their emphasis on preserving sharp edges. Moreover, it is extremely difficult to capture the complicated local features of natural images with a limited number of parameters.

In this paper, we propose a novel and accurate single image SR method based on internal across-scale gradient similarity. Given an input low-resolution image, its gradients in horizontal and vertical directions are first calculated and upscaled through the proposed internal gradient similarity-based upsampling algorithm. The target high-resolution image is then reconstructed based on the two upsampled gradients and the input low-resolution image. Our proposed approach combines the advantages of both self-similarity-based methods and the gradient-based techniques.

(a) (b)

Figure 1. Our image SR result of "child" (× 4). (a) Input low-resolution image; (b) Our SR result. Our method produces natural contours along eyes and face. Rich textured regions, e.g., the hat area, are also well reconstructed. For a better visual presentation, the input image is upscaled to the target resolution utilizing nearest-neighbor interpolation. This figure is better viewed on screen with high-resolution display.

Internal across-scale gradient similarity is based on the observation that small patches in a natural image tend to recur redundantly across different resolutions. Therefore, we should expect the similar redundancy for gradient patches. In this paper, we refer to the usage of image patch redundancy as “image similarity” and the gradient patch redundancy as “gradient similarity”.

As shown in Fig. 1, our proposed method successfully generates visually pleasing result in both edges and textures. From the zoom-in comparisons between the input low-resolution image (the input image is magnified to the target resolution by nearest-neighbor interpolation for better illustration) and the generated high-resolution result, it is clearly demonstrated that our approach is capable of restoring clear and natural face and eyes contours while still preserving realistic knit textures. As later shown in Fig. 7, it is a challenging task to restore the textures within the hat region for most of the recent state-of-the-art SR algorithms due to the complicated patterns. However, our generated result rebuilds this region naturally with minimal artifacts closest to the ground-truth.

Compared with the image similarity-based SR approaches, our proposed method has the following advantages:

- The proposed upscaling scheme is robust with stable SR performance by ensuring both local and global fidelity between the input low-resolution and the output high-resolution correspondences in both gradient level and image level.
- An easily-optimized energy function is utilized to combine constraints from different domains into one uniformed framework.
- Gradients of natural images have been modeled by a heavy-tailed distribution [16]. Computational cost can be reduced by upscaling the patches with small variance directly through bicubic interpolation.

The rest of the paper is organized as follows: Section 2 provides a detailed description of the proposed image SR method. Section 3 demonstrates the feasibility of reconstructing a high-quality image from its gradients and the advantages of the proposed gradient similarity-based SR algorithm. Section 4 discusses why only the internal statistics is adopted in our approach and the role that external statistics could play in image SR tasks. Experimental results and discussions are presented in Section 5. The conclusions are drawn in Section 6.

2. Internal Gradient Similarity-based Super-Resolution

In this section, we introduce the proposed image SR method based on internal across-scale gradient similarity. Fig. 2 illustrates the schematic pipeline of our approach. Since human eyes are more sensitive to brightness changes than color changes, therefore, same as the majority of other SR approaches [2, 4, 6, 7, 9, 11, 14, 15, 26, 29, 31, 33, 35, 36, 37, 39], for a given input low-resolution color image, our proposed SR algorithm is only performed in the luminance channel of the YUV color space while the other two channels are upsampled through bicubic interpolation.

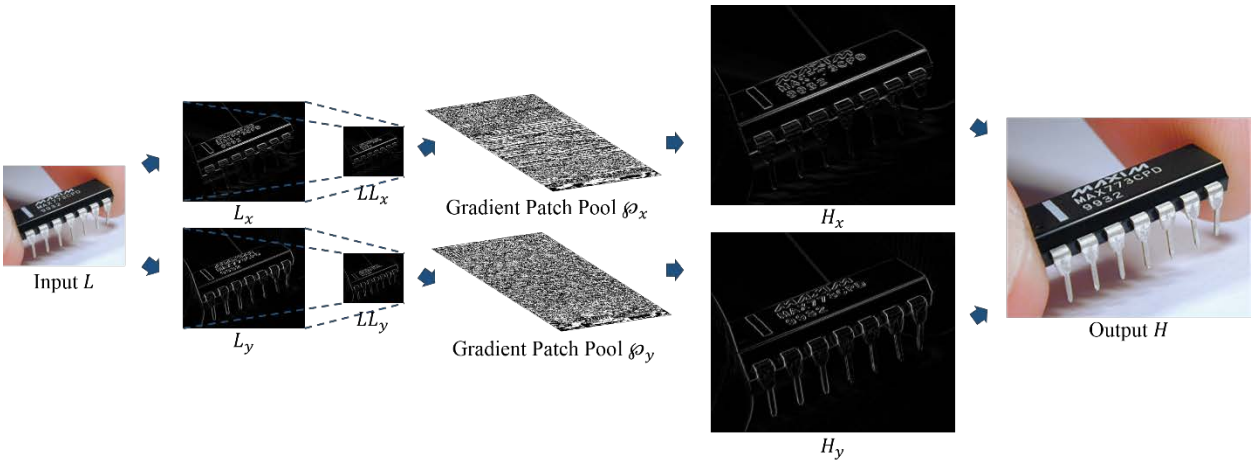

Figure 2. Flowchart of the proposed algorithm. After calculating the low-resolution gradients of the input image in horizontal and vertical directions (represented as L_x and L_y), for each gradient patch, the top k most similar patches are searched within the corresponding gradient patch pool. Patch pool $\wp_x$ is composed of all gradient patches in the downsampled version of L_x (represented as LL_x). $\wp_y$ is built up in a similar manner utilizing L_y. After obtaining the high-resolution gradients H_x and H_y, the output image is constructed based on them and the input image L.

Given a grayscale low-resolution image L, to upsample L by a magnification factor s, we first calculate the gradients L_x and L_y of L in horizontal and vertical directions. After that, based upon internal across-scale gradient similarity, L_x and L_y are upsampled individually by s to obtain the gradients in high-resolution as H_x and H_y; the second step is to reconstruct the target high-resolution image H from L, H_x and H_y through optimizing a uniformed energy function which incorporates the constraints in both image and gradient levels. Details of these two steps will be presented in the following subsections respectively.

2.1. Upscale Low-Resolution Gradients

In order to upscale the low-resolution gradient in x direction by factor s, L_x is firstly decomposed into a set of overlapping patches at size $a \times a$ ($a = 5$ in our implementation) with stride equals to 1. LL_x is calculated by downsampling L_x by s. A gradient patch pool $\wp_x$ is constructed with all the patches ($a \times a$) in LL_x. In order to form a more expressive patch pool, all the patches are normalized to have zero mean and uniform variance to better preserve the structural information rather than abstract values.

Gradients of natural images have been modeled by a heavy-tailed distribution. Therefore, generally for natural images, gradient patches form a sparse distribution. Majority of the gradient patches will be flat with small variances. For these patches, bicubic interpolation will be effective enough without compromising the final SR performance. For each patch p in L_x, we calculate its variance and compared it with a pre-set threshold θ. If the variance is smaller than θ, p is upscaled directly through bicubic interpolation; otherwise, after patch normalization, its top k most similar patches are searched within gradient patch pool $\wp_x$. The similarity between two instance patches is measured in their mean square error (MSE).

Within a pair of images representing the same scene but at different resolutions, given an instant patch in the coarse-resolution image, the corresponding patch in the fine-resolution image is referred as its 'parent' patch. After obtaining the top k most similar patches within $\wp_x$ for query patch p, their 'parent' patches in L_x are extracted, normalized to have zero mean and unit variance, and then combined weightedly. Patches that are more similar to the query patch are assigned with larger weights. The weights used to combine the k patches are computed with:

$$w_i = \frac{\exp\left(-\frac{M_i}{\sum_{j=1,\dots,k} M_j}\right)}{R} \tag{1}$$

Therein, w_i represents the weight for patch i during the combination, M_i stands for the MSE between the query patch p and patch i, R is a normalization factor to ensure the summation of the k weights equals to 1.

The combined patch is then adjusted according to the original mean and variance value of the input patch p and 'pasted' to H_x in the right position. The overlapped regions are simply averaged. H_y is calculated in a similar manner utilizing L_y and $\wp_y$.

2.2. Reconstruct High-Resolution Image

With L, H_x and H_y, the target high-resolution image H is reconstructed through minimizing the following energy function:

$$C = |(H * G)\downarrow_s - L|^2 + \lambda|\nabla H - \nabla H_D|^2, \quad (2)$$

where ∇H_D indicates the desired high-resolution gradients H_x and H_y obtained after the previous step. $*$ represents the convolution operation. $\downarrow$ is the downsampling operator. G stands for a Gaussian kernel with standard variance varies for different scaling factors s. We set the standard variance σ same as in [29]: $\sigma = 0.8$ if $s = 2$; $\sigma = 1.2$ if $s = 3$; $\sigma = 1.6$ if $s = 4$. λ is the weighting factor.

We integrate constraints in both image level and gradient level into a single cost function: the first term ensures the consistency between the output high-resolution image and the input low-resolution image. It has been demonstrated in [41] that a global constraint in the fidelity between input and output images is critical in image SR; the second term constrains the gradients of the reconstructed image to be close to the generated high-resolution gradients. The parameter λ controls the weight between these two terms. The cost function can be minimized through the gradient descent algorithm with:

$$H^{t+1} = H^t - \delta \cdot \left(\left((H^t * G)\downarrow_s - L\right)\uparrow^s * G - \lambda \cdot (\nabla^2 H^t - \nabla^2 H_D)\right), \quad (3)$$

where H^t represents the output after the t-th iteration and δ indicates the step width. Details regarding the parameter settings are introduced in Section 5.1. Feasibility of the proposed reconstruction process is demonstrated in Section 3.1.

3. Why Gradient Level Works Better?

In the proposed SR framework, patch-based self-similarity is utilized. Different from the existing state-of-the-art SR methods [11, 14, 37], which perform the upscaling algorithms using the similarity of image patches, we instead calculate the high-resolution gradients based on internal gradient patch similarity. The target high-resolution image is then restored based on the constructed high-resolution gradients and the input low-

resolution image.

In this section, two questions are answered:

- It is easy and straightforward to calculate the gradients in horizontal and vertical directions given an image. How well can we estimate a high-resolution image given its corresponding low-resolution image and the high-resolution gradients?
- While the proposed algorithm can be utilized to upscale an image directly, why instead taking a detour to upsample the gradients first?

3.1. Reconstruct Image from Gradients

In this subsection, we investigate the feasibility of recovering images from their gradients. The reconstruction process is based on Eqs. (2) and (3) (see Section 2.2 for details). We verify that the target high-resolution image can be well reconstructed from the input low-resolution image and the corresponding high-resolution gradients.

We conduct an experiment based on the Berkeley Segmentation Dataset (BSDS300) [23]. For each high-resolution image H_{GT}, its gradients in horizontal and vertical directions are calculated and serve as ∇H_D (shown in Eqs. (2)). The low-resolution image L is obtained through downsampling H_{GT} by the magnification factor s. Then H is restored iteratively according to Eq. (3). H^0 is initialized as the bicubic interpolated version of L. After the reconstruction, we calculate the difference between the reconstructed image H and the ground truth image H_{GT}.

Fig. 3(a) demonstrates the similarity of the original high-resolution images and the reconstructed images utilizing corresponding low-resolution images and high-resolution gradients. The experiments are performed over 100 natural images in database BSDS300 [23] at different scales s and weights λ. We adopt three different criterions in indicating the similarity between the ground truth and the reconstructed images. The pixel-wise averaged error is calculated as the absolute difference between the reconstructed image and the ground truth image followed by a division of the total number of pixels within each image. For upscaling factors of 2 and 4, an increase in weight λ consistently results in the improvement of the reconstruction performance measured in the increment of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM)

[32] and the decrement in the error due to the absolute accuracy of ∇H_D. However, in practice, it is impossible to restore the perfectly accurate high-resolution gradients from the low-resolution gradients. To better control the visual artifacts in the output high-resolution images, we set λ as 0.2 in our final SR experiments (more details regarding the implementation can be found in Section 5.1).

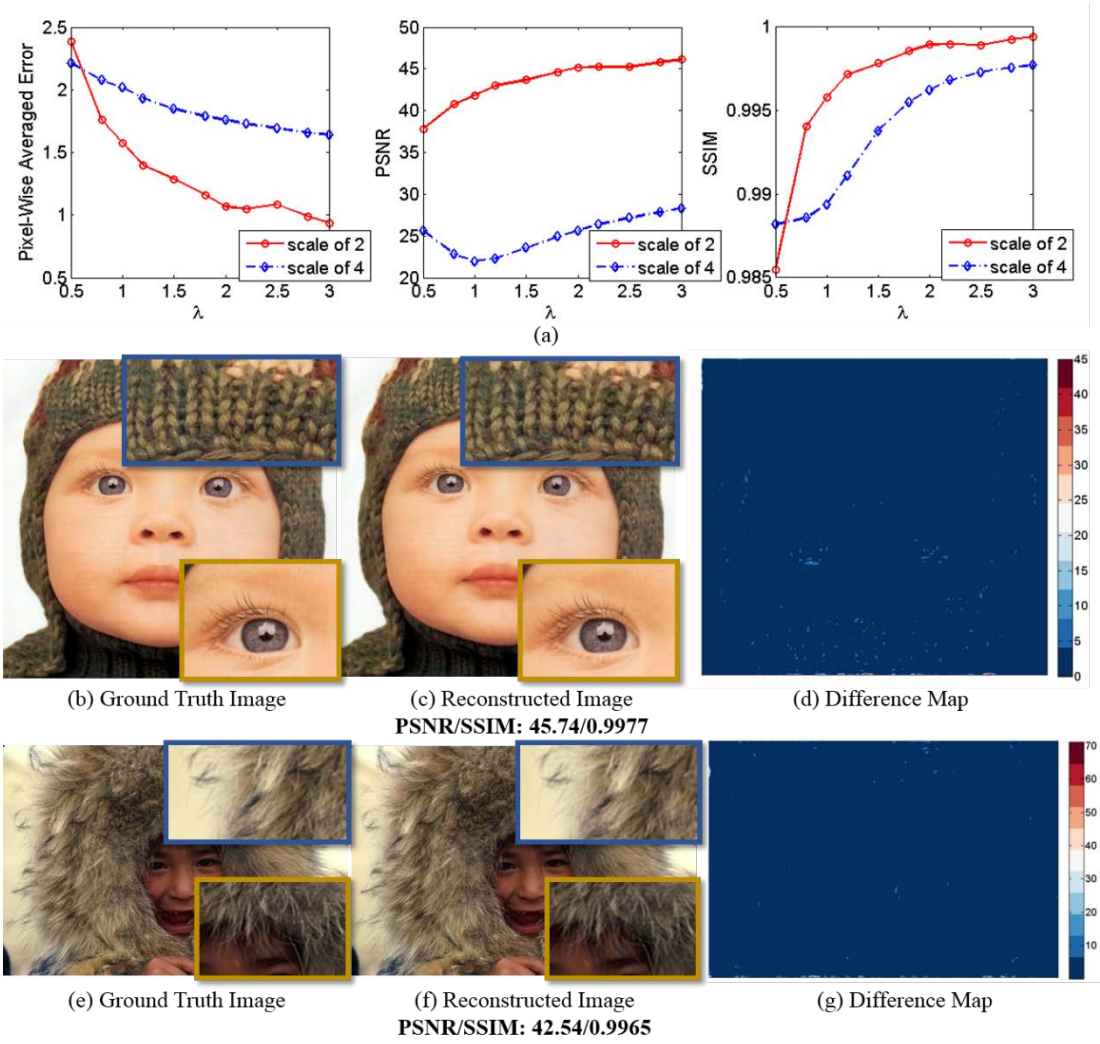

Figure 3. Demonstration of the feasibility to restore the high-resolution image from accurate horizontal & vertical gradients and the corresponding low-resolution image using the proposed reconstruction framework. (a) Experiments over the ground truth images and the reconstructed images based on 100 natural images in database BSDS300 [23] at different scales and weights (λ). Left: pixel-wise averaged error; Middle: PSNR; Right: SSIM. (b)-(g) illustrate the reconstruction results on images 'child', 'smile' along with the ground-truth images and the correspondent difference maps. Error mainly exists along image boundaries.

The effectiveness of restoring an image from its gradients is further demonstrated in Fig. 3(b)-(g) with two specific examples. Here the scaling factor is 4 and weight is set to be 1. Fig. 3(b) is the ground truth 'child' image. The low-resolution image is generated by downsampling the ground truth image by factor 4. Fig. 3(c) is the reconstructed image based on the ground truth gradients. As can be seen more clearly in the zoom-in areas, edges and textures are perfectly restored even for the eyelashes. Visually we cannot differentiate (b) from (c). Fig. 3(d) presents the pixel-wise differences between Fig. 3(b) and (c) which further demonstrates that the ground truth image and the reconstructed high-resolution image are nearly identical. Fig. 3(e)-(g) provide another set of results over image 'smile'. The ground truth image is a little noisy. Still, the reconstruction image is nearly the same as the original one. In both examples, as shown in Fig. 3(d) and (g), majority of the reconstruction error lies in the boundary of the images due to the fact that we do not have adequate information in those regions. We observe the similar pattern for all the other images involved in the experiment. Extremely good PSNR and SSIM numbers are listed to illustrate that a near-perfect high-quality image can be well reconstructed from accurate gradients and the corresponding low-resolution image.

3.2. Gradient Similarity vs. Image Similarity

Accurate reconstruction of edges and textures are critical to SR since they are the most perceptually essential features in a natural image. However, it is difficult to automatically hallucinate both edges and textures within one framework due to the different characteristics revealed by these two features. In this paper, the edges which provide structural information of the objects in the images are referred as structural edges. Therefore, in our proposed method, we aim to create sharp structural edges and reasonable textures.

Moreover, a key element in SR application is that the output high-resolution image should be consistent with the input low-resolution image according to the definition of SR. It is demonstrated in [41] that ensuring the consistency during high-resolution image reconstruction is at least as important as the usage of a proper image prior. Self-similarity based SR approaches ensure a local consistency between low-resolution patch instance and corresponding high-resolution patch. A global fidelity constraint is often ignored in the upsampling scheme. Although back-projection is commonly utilized in self-similarity based approaches, it

only provides a limited solution. In the proposed framework, both local and global fidelities are ensured in one uniformed framework represented by two different terms in the cost function.

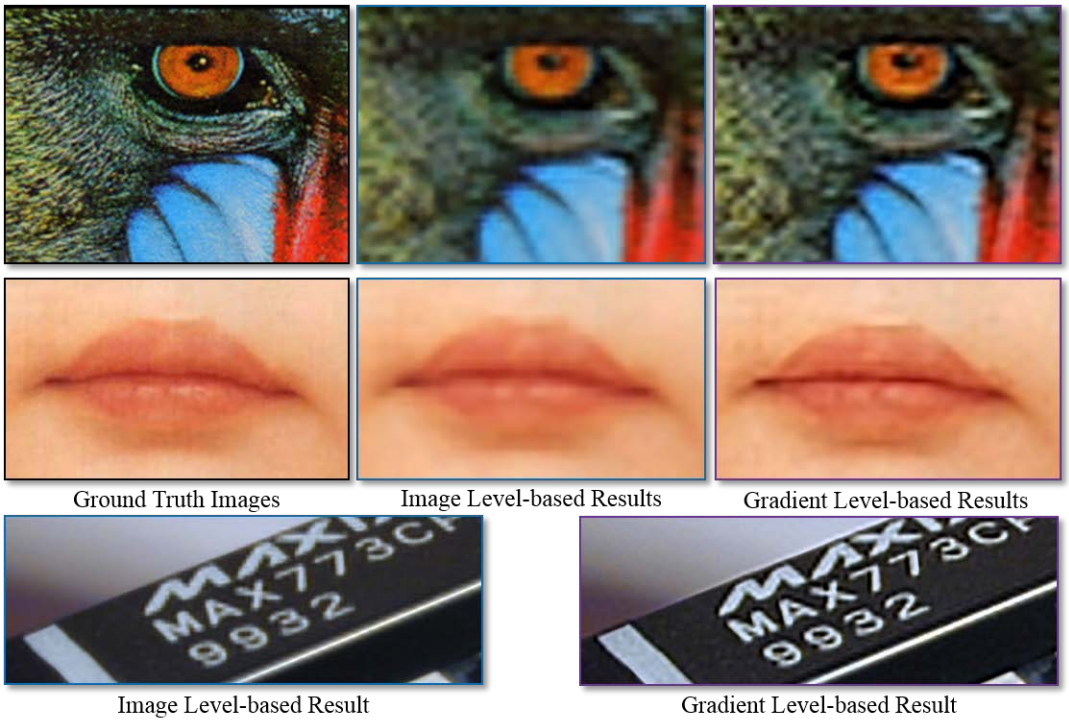

Figure 4. Comparisons of results in upsampling images by 4 in the image level and the gradient level respectively using the same framework. Three examples are presented and ground truth images are included for two cases shown in the top and middle row. Ground-truth is not available for the example in the bottom row. In all three cases, it is clearly demonstrated that the results produced in the gradient level reveal sharper edges and more natural textures.

Image similarity-based SR methods have been successful in synthesizing rich details. Compared with edge-directed approaches, methods adopt internal image similarity [11, 14] hallucinate more natural textures. However, it is difficult to control the artifacts introduced along structural edges especially under large scaling factors. Instead, gradients emphasize more on the intensity changes. Modeled by a marginal distribution, image gradient is often combined with L_2 norm or sparsity regularization and has been a success in a variety of image restoration tasks. As illustrated in [29], reconstruction based on only image level or only gradient level introduces artifacts. Combining constraints in both levels provides better and more stable SR performance. Therefore, gradient patches combined with the proposed reconstruction framework are more robust compared with traditional self-similarity approaches utilizing only image patches in preserving

structural edges and complicated textures for generic natural images. However, gradient-based approaches are known to be noise sensitive. Under circumstances where the input images are very noisy, the proposed SR framework could be combined with denoising algorithms to enhance the overall image quality in a sequential manner.

To further demonstrate the effectiveness of the proposed gradient similarity-based SR scheme, we reconstruct the high-resolution images from the same low-resolution input image by an upscaling factor of 4 using the proposed 'search and paste' framework based on image similarity and gradient similarity respectively. The two cases use the same set of parameters to upsample a low-resolution image or gradient. As demonstrated in Fig. 4, the results generated utilizing gradient similarity produce sharper edges and more natural details than those based on image similarity. For the two examples (shown in the top and middle rows in Fig. 4) with the ground truth images available, results generated used the proposed gradient-based reconstruction SR scheme are closer to the ground truth images and the image-level results are over-blurred in both edges and textures.

4. Internal vs. External Statistics

Recently, the usage of external natural images from a large dataset for reconstructing high-resolution images has raised many discussions. In this section, we investigate the contribution of internal and external statistics for gradient similarity-based SR reconstruction. We define the internal statistics to be the image or gradient patches extracted from only the input image at different resolutions without using any external sample images or statistics; the external statistics represent image or gradient patches extracted from images of an external natural image dataset.

Zontak and Irani [40] have demonstrated the effectiveness of internal statistics both in "Expressiveness" (how similar between a small patch and its most similar patches found internally or externally) and "Predictive Power" (how well can the found similar patches be used in image restoration tasks given a prediction model). For SR tasks, in order to achieve comparable results with methods adopt internal statistics, it usually requires a large external image dataset with hundreds or thousands of natural images. However, by

the usage of a large external image dataset, the computational cost will increase dramatically. Moreover, increasing the size of the external dataset makes the patch correspondences even more ambiguous [39].

We aim at exploring, with the presence of internal statistics, whether the external gradient statistics are helpful in boosting the performance of gradient similarity-based SR. Since including a large external dataset will be infeasible in practice, our experiments in this section collect external statistics from a small dataset (with 5 or 10 images). Even though the number of external images used is limited, under a small patch size (5 × 5), still hundreds of thousands of patch instances are collected.

We compare the reconstructed high-resolution images from the same low-resolution input image using only internal gradient statistics and both internal and external statistics. For both cases, the image SR follows the same pipeline as introduced in Section 2. The only difference lies in the formed gradient patch pools. We evaluate two different kinds of external statistics: general external gradient statistics and class-specific external gradient statistics.

4.1. General External Gradient Statistics

General external statistics are collected from 10 fine-resolution natural images[1]. We evaluate the contribution of the external gradient statistics on constructing high-resolution outputs from 100 images of BSDS300 [23]. All the images are rescaled to three different sizes: 40 × 40, 80 × 80, and 160 × 160. The SR task is to upsample these images by a scaling factor of 2 using only internal gradient statistics and by using both internal and external statistics. In both cases, the upscaling schemes follow the same pipeline as shown in Section 2. The only difference lies in the formation of the gradient patch pool. Only internal gradient patches are utilized to form the patch pool if internal gradient statistics are adopted. For the latter case, besides internal collected patches, gradient patches from the downsampled version of the 10 external images are also included in the patch pool. Since the original sizes of the 100 images are larger than 320 × 320, the ground truth images are available for all three input sizes.

[1]The images are downloaded from Flickr (www.flickr.com). Example images will be available after the paper publication.

Evaluation of the SR performance based on different statistics is measured with SSIM between the generated high-resolution images and the ground truth images. Here we assume that a larger SSIM indicates the better SR performance given a uniformed pipeline.

Table 1 illustrates the statistical result that for each size of the input images, the percentage of images which have an increase in SSIM when external statistics are introduced. For most of the input images with size 160×160, including external statistics will not boost the SR performance. On the contrary, 67% of the images have a larger SSIM without external statistics. An increase in the patch pool will definitely increase the "Expressiveness". However, it reduces the "Predictive Power" for the SR task due to the increasing of ambiguity between patch correspondences. As the size of the input image decreases, the external statistics become more useful since the patch pool formed with only internal statistics is very limited.

TABLE 1: Results of the percentage of images which have higher SSIM when external statistics are introduced at three different input sizes

Input Size	Percentage of images with larger SSIM with general external statistics
40×40	85%
80×80	56%
160×160	33%

Generally, for SR applications, the size of the input image is measured in hundreds by hundreds pixels where the external statistics seems not that useful. Therefore, in our proposed SR model, only internal statistics are adopted without usage of general external statistics.

4.2. Class-Specific External Gradient Statistics

Zontak and Irani [40] have described the class-specific external database to be "extremely useful, even if small". Here, we refer to class-specific external database as a set of sample high-resolution images which are highly similar to the target image (images representing the same scene or sharing the same textures). We also evaluate the SR performance by using only internal statistics and both internal and class-specific external statistics.

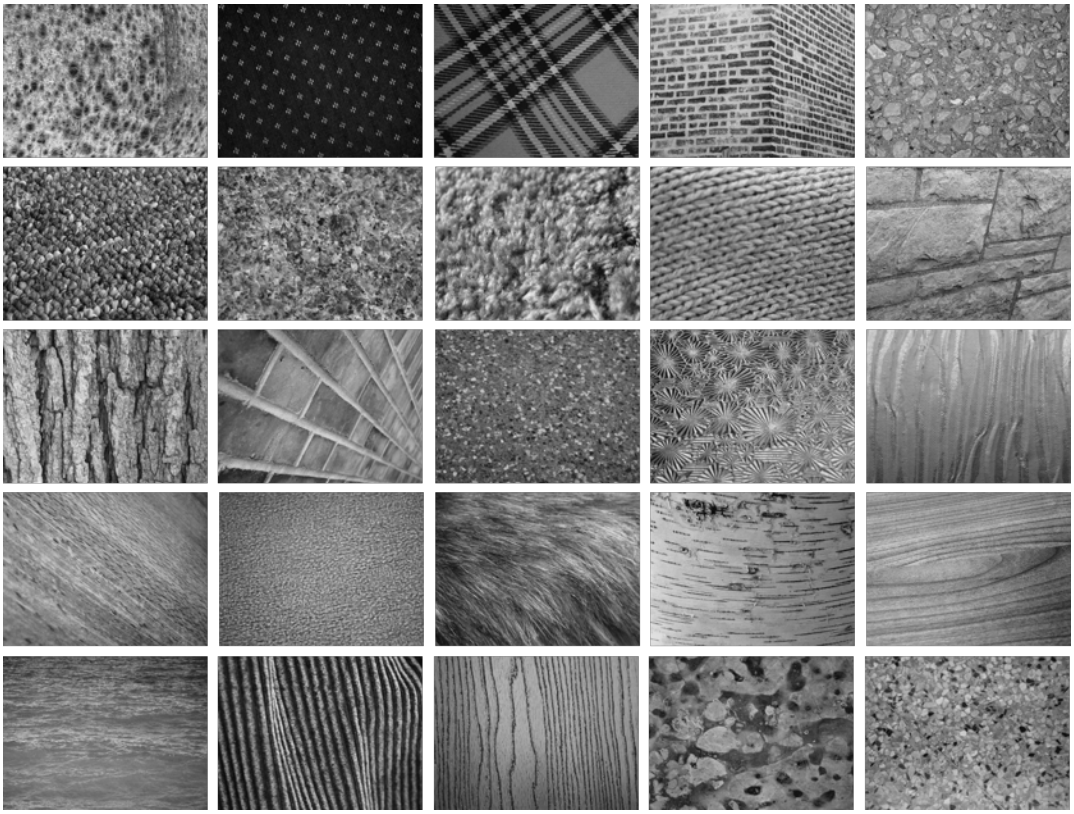

Figure 5. Example images of all the 25 classes in the UIUC Texture Dataset [19] (sorted in descending order measured in "percentage of images with larger SSIM with class-specific external statistics". Refer to Table 2 for more details).

The experiment is conducted on the UIUC Texture Dataset [19]. The database includes 25 texture classes, 40 samples for each class, all in grayscale (see Fig. 5 for examples of different texture patterns.) We run the experiment on all the 25 classes. For each class, the first 5 images are selected to collect the external gradient statistics; the remaining 35 images are downsampled by magnification factor 2 and then serve as the input images for the comparison. After the rescaling, the generated low-resolution images all have size 320×240.

Table 2 illustrates the percentage of images which have an increase in SSIM when class-specific external statistics are introduced. The result demonstrates that with the presence of small scale class-specific external statistics, the SR performance varies dramatically for different classes. For the class "bark2", all images have better performance when external statistics are introduced. However, for the class "floor2", none of the

images achieves better performance with external statistics. Fig. 5 presents the examples of all the 25 texture patterns from the UIUC texture database. The 25 texture patterns are sorted in descending order measured in "percentage of images with larger SSIM with class-specific external statistics" as presented in Table 2. There is no clear criterion to differentiate the classes which benefit from class-specific external statistics with those do not. As observed from Fig. 5, very roughly speaking, images with sparse regular or square patterns tend to make more contributions when they are introduced as class-specific external statistics; on the other hand, external gradient statistics from dense irregular or parallel patterns degrade the SR performance.

To sum up, our experiments indicate that external gradient statistics collected from a small dataset normally does not improve the performance for general SR tasks. However, under certain circumstances, such as when the size of the input image is very small or the external dataset contains fine-resolution images very much similar to the target high-resolution image, external statistics might be helpful. Our proposed method adopts internal statistics only and can be easily scaled to include external statistics when needed.

TABLE 2: Results of the percentage of images which have higher SSIM when class-specific external statistics are introduced at 25 different classes.

Image Class	Percentage of images with larger SSIM with class-specific external statistics
bark1	74.29%
bark2	**100.00%**
bark3	45.71%
wood1	37.14%
wood2	22.86%
wood3	65.71%
water	34.29%
granite	80.00%
marble	14.29%
floor1	71.43%
floor2	**0%**
pebbles	82.86%
wall	77.14%
brick1	88.57%
brick2	74.29%
glass1	71.43%
glass2	57.14%
carpet1	82.86%
carpet2	80.00%
upholstery	91.43%
wallpaper	71.43%
fur	48.57%
knit	80.00%
corduroy	28.57%
plaid	91.43%

5. Experimental Results

In this section, we evaluate the proposed SR method on multiple natural images at different upsampling factors. The greyscale images are directly upscaled using our proposed algorithm. For the color images, as we mentioned previously in Section 2 that since human eyes are more sensitive to luminance changes, we only perform the proposed SR algorithm on the luminance channel in YUV color space while the rest two color channels are upscaled directly through bicubic interpolation. We also compare our results with existing state-of-the-art approaches in single image SR.

5.1. Parameter Selection

Same as many existing state-of-the-art SR methods, large upscaling factors require a coarse-to-fine scheme in our proposed model. However, instead of adopting a relatively small scaling factor per-step, our method upsamples the image by a factor of 1.5 or 2 progressively, i.e., upscaling factor of 3 follows a 1.5×2 manner; upscaling factor of 4 takes steps 2×2; upscaling factor of 8 is calculated as $2 \times 2 \times 2$; and so forth.

Patch size a during the gradient patch upsampling is set to be 5 and during the searching, the top 5 most similar patches are selected. Threshold θ in differentiating smooth patches from non-smooth patches is set to 10. To reconstruct the final output image from the upscaled gradients, the weight λ is assigned to be 0.2. As mentioned earlier in Section 3.1, although with the presence of ground truth gradients, a larger weight leads to a better reconstruction result, in practice, we do not have the gradients that are extremely accurate available. Output images generated by a relatively large λ suffer from visual artifacts of over-sharped edges and unrealistic textures. Setting λ to a large number will also increase the noise sensitivity of the proposed SR framework. Therefore, at scales 2 and 4, a set of high-resolution images are generated in datasets BSDS200 [23] and SET5 [2] utilizing different λ ranging from 0.05 to 0.5 at a stride of 0.05. We observe that λ between 0.1 and 0.2 gives fairly similar outputs with stable SR performance of minimal artifacts. In Eq. (3), with a fixed step, a larger lambda will lead to a faster convergence. Therefore, we set λ to 0.2 to obtain the best SR results.

Figure 6. SR of image 'comic' (× 8). (a) Input low-resolution image. (b) Our result. SR of image 'butterfly' (× 16). (c) Input low-resolution image. (d) Our result. Our method synthesizes realistic details and restores clear edges. For a better presentation, the input image is upscaled to the target resolution utilizing nearest-neighbor interpolation. This figure is better viewed on screen with high-resolution display.

5.2. Visual Results

We test our proposed method on a variety of natural images with different upscaling factors. Results are compared with recent state-of-the-art approaches [9, 11, 14, 26, 31, 33, 35, 37, 38]. Single image SR is very challenging when the scaling factor is large due to too much missing information. We need to estimate numerous unknown values based on a small amount of given pixels. In the coarse-to-fine scheme, error

accumulates as the scaling factor increases. Fig. 6 demonstrates the feasibility of the proposed SR framework when the magnification factors are large (8 for the image 'comic' and 16 for the image 'butterfly'). Our proposed algorithm is robust with satisfactory SR performance and is capable of reconstructing realistic high-frequency regions.

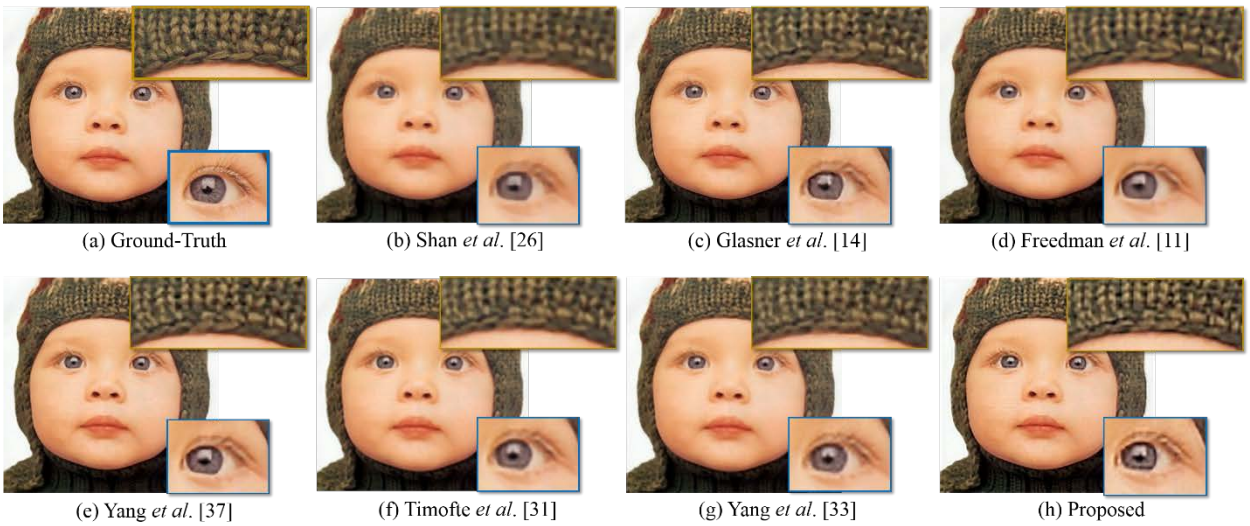

Figure 7. SR of image 'child' (× 4). Our proposed method successfully reconstructs the knit textures in the **hat** region and maintains sharp and natural eyes, facial and lip contours so that the hat is actually above the face, not vice versa as observed in peer results. This figure is better viewed on screen with high-resolution display.

We further compare our results with other recently published approaches as illustrated from Figs. 7-10. Fig. 7 presents a set of SR results on image "child" under an upscaling factor of 4. 'Child' has been a popular test image in single image SR papers. Although recent state-of-the-art approaches are able to generate clean edges along eyes, face and lip contours, it is still very challenging to restore the knit textures in the hat region. As clearly illustrated in the zoom-in areas, our approach can produce more realistic hat textures with minimal artifacts compared with peer results. Moreover, in our generated high-resolution image, the contour between face and hat is more natural without being too "sharp" so that the hat is actually above the face, not vice versa. Papers [31, 33, 37] represent the very recent state-of-the-art SR methods. Yang *et al.* [37] were able to restore sharp and clean structural edges but tend to blur the hat area which includes complicated knit patterns. Timofte *et al.* [31] and Yang *et al.* [33] are capable of generating natural-looking outputs but there are

noticeable artifacts along the face contour and within eyes areas. The hat region is also over-smoothed with blurry visual artifacts. In contrast, our algorithm successfully produces clean face, lip contours and reconstructs hat textures closest to the ground truth image.

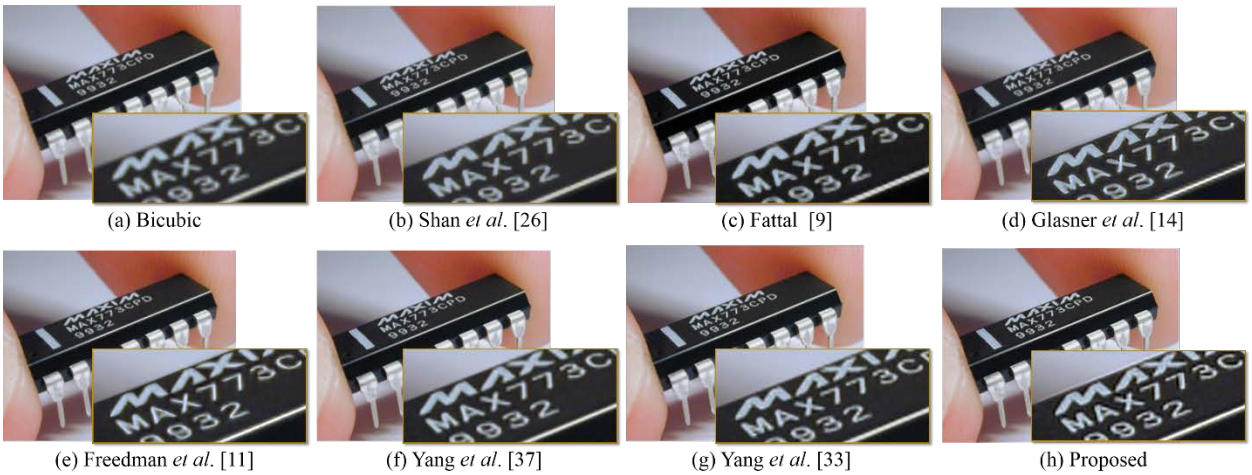

Figure 8. SR of image 'chip' (× 4). Our proposed method retains clean and sharp structural edges. This figure is better viewed on screen with high-resolution display.

Fig. 8 provides another set of SR results over image 'chip' which is also a commonly used test image. The magnification factor is 4. The zoom-in areas clearly indicate that our method creates natural results among the characters and along the structural edges of the chip.

In Fig. 9, we compare our results with two dictionary-based approaches [35, 38] with the existence of the ground truth images under magnification factor 3. The three example images contain different kinds of challenging tasks for SR including fur, whiskers, eyes, fabrics, feather, and object shadows. As shown in the zoom-in areas, our produced results are vivid and realistic with more natural details compared with the ground truth images. For example, in image 'Lenna', the edges in the hat fabrics generated by [35, 38] are over-smoothed and our result reconstructs patterns very much similar to the ground truth. Moreover, the reconstructed feather in the hat by our proposed method is natural and clean.

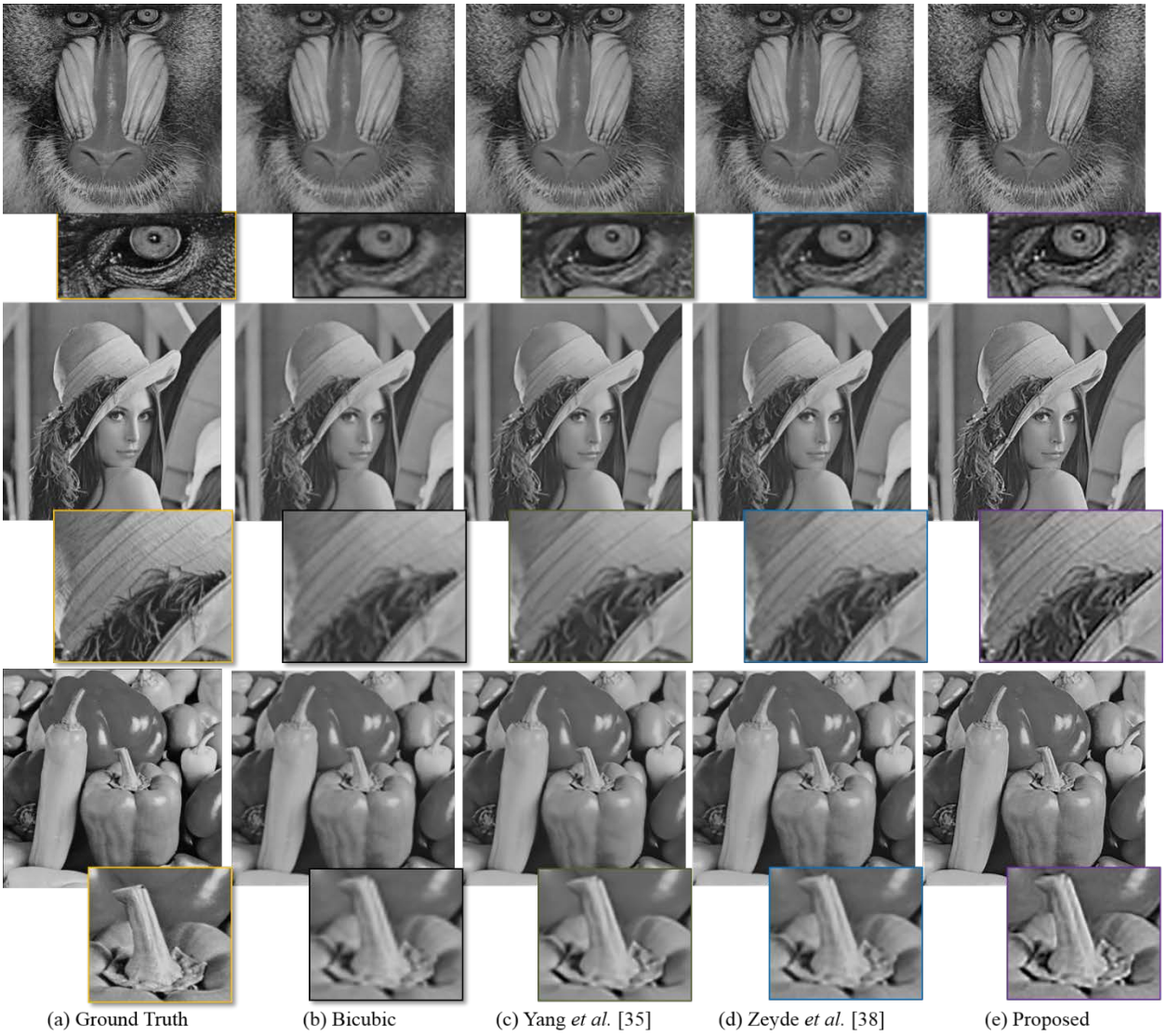

Figure 9. SR of images 'baboon', 'Lenna', and 'pepper' (× 3). In all three images, our proposed method generates results with clear edges and realistic textures closest to the ground truth images. This figure is better viewed on screen with high-resolution display.

More results are shown in Fig. 10 with comparisons to Shan *et al.*[2] [26], Sun *et al.* [29] and Yang *et al.* [33]. The ground truth images all come from BSDS300 [23]. In the image 'mushroom', the results generated by [26, 29] tend to blur the patterns within the mushrooms while there are noticeable visual artifacts in [33]. Our result better synthesizes the complicated details without over-sharpen the edges. Our method also produces

[2]The results are generated using the executable file provided by the authors. We use the default parameters.

sharper edges with minimal artifacts in the 'flower' image as illustrated in the zoom-in. For the 'fish' image, all the other three methods fail to recover the white line along the body contour of the fish. But our method nicely reconstructs this structural edge. Similarly, for the tire part in 'car' and the cloth shown in 'girl', our results are clearer and more visually pleasing.

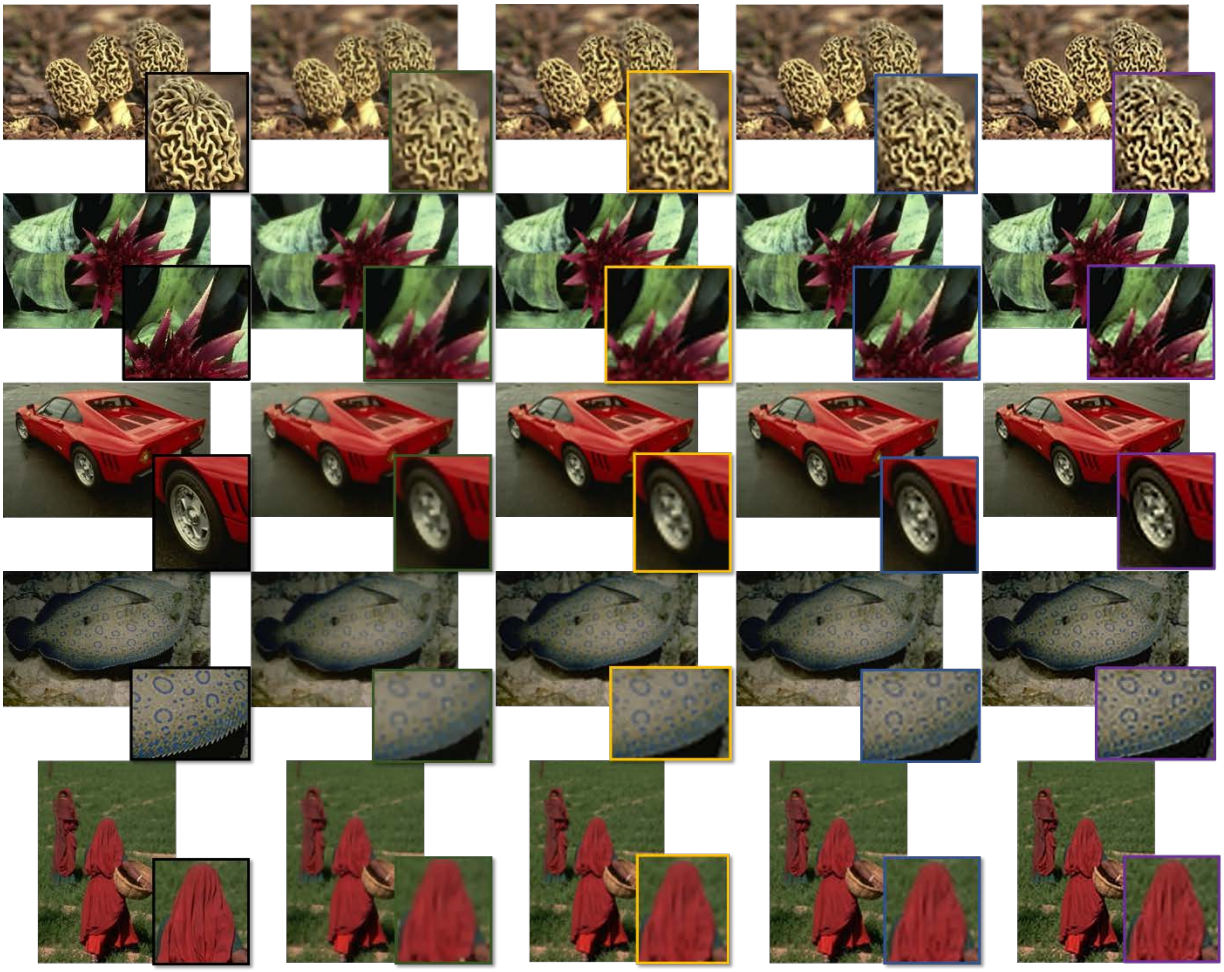

Figure 10. SR of images 'mushroom', 'flower', 'car', 'fish', and 'girl' (× 4). In 'mushroom', our result best reconstructs the complicated contours without the blurring artifacts. For image "fish", only our method well recovers the white line along the fish body contour. Among all five cases, our proposed method produces results with clear edges and natural details compared to other state-of-the-art algorithms. This figure is better viewed on screen with high-resolution display.

5.3. Quantitative Evaluation

The SR performance is further evaluated quantitatively in this subsection. Recently, in [34], a variety of image quality metrics are evaluated by investigating the correlation with visual perception by human experts. Among the 8 criterions evaluated, Information Fidelity Criterion (IFC) [42] seems to be the most proper metric that could be utilized to perform comparison among different SR frameworks. Therefore, we compare our results with [26] and [36] on dataset SET5 [2] (with 5 images, i.e., baby, bird, butterfly, head, and woman) under a scaling factor of 4 utilizing IFC. The high-resolution images for [36] are generated with the released code provided by the authors. Table 3 presents the corresponding comparison results. As observed, our method outperforms [26, 36] in all 5 images measured in IFC.

TABLE 3: IFC comparisons with peer SR methods on SET5 [2]

SET5	Shan [26]	Yang [36]	Ours
baby	1.8414	1.7387	**2.3597**
bird	1.9941	2.2013	**2.5754**
butterfly	2.3266	2.4080	**2.5282**
head	1.4845	1.4956	**1.8091**
woman	1.8733	1.9373	**2.0316**

6. Conclusions

In this paper, we have proposed a novel image super-resolution model via internal gradient similarity. As demonstrated by the extensive experimental results, our method is capable of producing sharp edges while maintaining natural and realistic textures compared with recent state-of-the-art SR methods. By adopting a uniformed reconstruction framework to pose constraints in both gradient and image domains and ensure the fidelity between input and output images locally and globally, the proposed approach is robust with stable SR performance under different input images and scaling factors in reconstructing visually pleasing results.

Acknowledgments

This work was supported in part by ONR grant N000141310450 and NSF grants EFRI-1137172, IIP-1343402.

Reference

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):1167-1183,2002

[2] M. Bevilacqua, A. Roumy, C. Guillemot, and M. A. Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012

[3] J. Boulanger, C. Kervrann, and P. Bouthemy. Space-time adaptation for patch-based image sequence restoration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(6):1096-1102, 2007

[4] H. Chang, D. Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR*, 2004

[5] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 3D Shape Scanning with a Time-of-Flight Camera. In *CVPR*, 2010

[6] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen. Deep Network Cascade for Image Super-resolution. In *ECCV*, 2014

[7] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a Deep Convolutional Network for Image Super-Resolution. In *ECCV*, 2014

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and Robust Multiframe Super Resolution. *IEEE Trans. Image Processing*, 13(10): 1327-1344, 2004

[9] R. Fattal. Image Upsampling via Imposed Edge Statistics. In *ACM SIGGRAPH*, 2007

[10] R. Fransens, C. Strecha, and L. V. Gool. Optical flow based super-resolution: A probabilistic approach. *Computer Vision and Image Understanding*, 106(1):106-115, 2007

[11] G. Freedman, and R. Fattal. Image and Video Upscaling from Local Self-Examples. *ACM Trans. Graph*, 28(3): 1-10, 2010

[12] W. Freeman, E. Pasztor, and O. Carmichael. Learning Low-Level Vision. *International Journal of Computer Vision*, 40(1):25-47, 2000

[13] W. Freeman, T. Jones, and E. Pasztor. Example-based Super-Resolution. *Computer Graphics and Applications*, 22(2):56-65, 2002

[14] D. Glasner, S. Bagon, and M. Irani. Super-Resolution from a Single Image. In *ICCV*, 2009

[15] L. He, H. Qi, and R. Zaretzki. Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-Resolution. In *CVPR*, 2013

[16] J. Huang, and D. Mumford. Statistics of natural images and models. In *CVPR*, 1999

[17] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231-239, 1991

[18] M. Kiechle, S. Hawe, and M. Kleinsteuber. A Joint Intensity and Depth Co-Sparse Analysis Model for Depth Map Super-Resolution. In *ICCV*, 2013

[19] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(8):1265-1278, 2005

[20] H. S. Lee and K. M. Lee. Simultaneous Super-Resolution of Depth and Images using a Single Camera. In *CVPR*, 2013

[21] J. Li, Z. Lu, G. Zeng, R. Gan, and H. Zha. Similarity-Aware Patchwork Assembly for Depth Image Super-Resolution. In *CVPR*, 2014

[22] X. Li, and M. Orchard. New Edge-Directed Interpolation. *IEEE Trans. Image Processing*, 10(10):1521-1527, 2001

[23] 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 *ICCV*, 2001

[24] M. Protter, M. Elad, H. Tekeda, and P. Milanfar. Generalizing the non-local-means to super-resolution reconstruction. *IEEE Trans. Image Processing*, 18(1):36-51, 2009

[25] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. LidarBoost: Depth Superresolution for ToF 3D Shape Scanning. In *CVPR*, 2009

[26] Q. Shan, Z. Li, J. Jia, and C. Tang. Fast Image/Video Upsampling. In *ACM SIGGRAPH Asia*, 2008

[27] B. Shi, H. Zhao, M. Ben-Ezra, S. Yeung, C. Fernandez-Cull, R. H. Shepard, C. Barsi, and R. Raskar. Sub-Pixel Layout for Super-Resolution with Images in the Octic Group. In *ECCV*, 2014

[28] D. Su, and P. Willis. Image interpolation by pixel-level data-dependent triangulation. *Computer Graphics Forum*, 23(2):189-201, 2004

[29] J. Sun, J. Sun, Z. Xu, and H. Shum. Image Super-Resolution using Gradient Profile Prior. In *CVPR*, 2008

[30] Y. Tai, S. Liu, M. Brown, and S. Lin. Super Resolution using Edge Prior and Single Image Detail Synthesis. In *CVPR*, 2010

[31] R. Timofte, V.D. Smet, and L.V. Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, 2013

[32] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600-612, 2004

[33] C. Yang and M. Yang. Fast Direct Super-Resolution by Simple Functions. In *ICCV*, 2013

[34] C. Yang, C. Ma, and M. Yang. Single-Image Super-Resolution: A Benchmark. In *ECCV*, 2014

[35] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, 2008

[36] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Processing*, 19(11):2861-2873, 2010

[37] J. Yang, Z. Lin, and S. Cohen. Fast Image Super-resolution Based on In-place Example Regression. In *CVPR*, 2013

[38] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. *Curves and Surfaces*, 6920:711-730, 2010

[39] Y. Zhu, Y. Zhang, and A. L. Yuille. Single image super-resolution using deformable patches. In *CVPR*, 2014

[40] M. Zontak, and M. Irani. Internal Statistics of a Single Natural Image. In *CVPR*, 2011

[41] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin. Accurate Blur Models vs. Image Priors in Single Image Super-Resolution. In *ICCV*, 2013

[42] H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Tran. Image Processing*, 14(12): 2117-2128, 2005