Single-image superresolution based on local regression and nonlocal self-similarity

Jing Hu
Yupin Luo

Single-image superresolution based on local regression and nonlocal self-similarity

Jing Hu* and Yupin Luo
Tsinghua University, Department of Automation, Tsinghua National Laboratory for Information Science and Technology (TNList),
Beijing 100084, China

Abstract. The challenge of learning-based superresolution (SR) is to predict the relationships between low-resolution (LR) patches and their corresponding high-resolution (HR) patches. By learning such relationships from external training images, the existing learning-based SR approaches are often affected by the relevance between the training data and the LR input image. Therefore, we propose a single-image SR method that learns the LR-HR relations from the given LR image itself instead of any external images. Both the local regression model and nonlocal patch redundancy are exploited in the proposed method. The local regression model is employed to derive the mapping functions between self-LR-HR example patches, and the nonlocal self-similarity gives rise to a high-order derivative estimation of the derived mapping function. Moreover, to fully exploit the multiscale similarities inside the LR input image, we accumulate the previous reconstruction results and their corresponding LR versions as additional example patches for the subsequent estimation process, and adopt a gradual magnification scheme to achieve the desired zooming size step by step. Extensive experiments on benchmark images have validated the effectiveness of the proposed method. Compared to other state-of-the-art SR approaches, the proposed method provides photorealistic HR images with sharp edges. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JEI.23.3.033014]

1 Introduction

Image superresolution (SR) aims to generate a high-resolution (HR) image from a single image or a set of low-resolution (LR) images. With the fast emergence of high-definition displays at consumer electronics market, such as Nexus 7 (1920×1200) and iPad3 (2048×1536), there is a great demand to super-resolve existing LR images, so that they can be enjoyably viewed on such readily available HR devices without any visual artifacts. Therefore, the SR problem has gained popularity in image processing community in recent years. Ever since the pioneer work proposed by Huang and Tsai in 1984,[1] many improvements have been made in this particular research field. Generally, SR methods that are implemented in spatial domain can be roughly categorized into three classes: the interpolation-based methods,[2–8] the reconstruction-based methods,[9–11] and the learning-based methods,[12–27,28,29,30] each of which has its own distinctive advantages, prior assumptions, and requirements for additional information. For a comprehensive literature survey, the interested reader is referred to Refs. 31 and 32.

The simplest interpolation-based SR methods, such as bicubic and Lanczos, use linear interpolation kernels to predict unknown pixels in a higher grid. Because these linear kernels are usually derived under the assumption that natural images are either spatially smooth or band-limited, the interpolation-based SR methods tend to produce smooth HR images with several visual artifacts, such as ringing, aliasing, blocking, and blurry.[2,4] In recent years, some sophisticated methods,[2–8] in which the kernel coefficients are locally adaptive, have been

proposed to reduce these unwanted artifacts. For instance, the regression model used in Refs. 5–7 and the adaptive normalized convolution technique used in Ref. 8 are very helpful to approximate local structures and obtain somewhat sharp edges.

The reconstruction-based SR methods regard the SR problem as an inverse imaging procedure. Moreover, considering that far more pixels than the number given need to be determined, this inverse problem is intrinsically ill-posed. To cope with this, certain priors are incorporated. For example, the gradient profile prior[10] and the soft edge smooth prior[11] are commonly used for yielding sharp edges. Besides the prior knowledge, the embedded back-projection policy guarantees the consistency between the resultant HR and the LR input images. In general, methods in this category perform better than the interpolation-based methods, but it is still difficult to recover high-frequency (HF) details. Besides, Lin and Shum[33] have validated that only under a 2× magnification, can such reconstruction-based methods effectively generate fine details, which, of course, greatly limits their applications.

To exceed the limits of the reconstruction-based SR approaches, the learning-based SR methods aim at estimating the HF details that are not explicitly found in the LR input images. In the pioneer work proposed by Freeman et al.,[13] these lost HF details are learned from a set of universal LR-HR patch pairs. Several extensions[14–21] have been proposed thus far to better predict the correspondences between LR-HR patches from varieties of training images. Although their successful SR capabilities have been substantiated by extensive experimental results, several researchers[10,22–24] have pointed out that false HF details are very likely to be introduced if the LR input image is incompatible with the training images. Therefore, many self-learning

*Address all correspondence to: Jing Hu, E-mail: j-hu09@mails.tsinghua.edu.cn

algorithms[22–27,29] have been proposed recently to use the LR input image itself as the only source of LR-HR example patches. Such approaches are all based on the observation that the small patches tend to redundantly repeat themselves many times within the original image scale[34] as well as across different image scales.[22–24,27,35] Compared to SR approaches using an external database, these self-learning approaches are able to exploit more relevant examples.[22,24]

In this paper, we propose a self-learning SR method by taking advantage of local structural information and nonlocal patch redundancy (self-similarity) property[22–24,27,35] in natural images. First, a set of self LR-HR patch pairs is formed by splitting the LR input image and its smoothened version into overlapped patches. Based on these example patch pairs, a regression model is established to derive the mapping functions that characterize relationships between the LR-HR patches. By exploiting the multiscale patch redundancy in LR image, we successfully obtain the high-order derivation estimation as well as the approximation error of each LR-HR mapping function. Moreover, since cross-scale self-similarity holds better for small scaling factors,[23] we adopt a gradual up-sampling scheme, and accumulate the previous reconstructed HR images and their corresponding LR versions for succeeding estimation.

Contributions of this paper are twofold:

1. Reliable SR reconstruction. In our scheme, both local structural information and nonlocal self-similarity are intimately related in a unified framework. With the aid of the patch redundancy from multiple image scales, we give a closed-form solution to a high-order derivative estimation of the LR-HR mapping function.

2. Adaptive structure preservation. Since the mapping function, from low to high resolution grid, is particularly specified for each patch in the initial LR image, the local structure is therefore implicitly preserved. Besides, the high-order derivative estimation also leads to the preservation of complex structures, such as edge orientations (first-order derivatives) and curvatures (second-order derivatives).

The remainder of this paper is organized as follows. The next section gives a brief overview of related work on local regression model and nonlocal similarity in self-learning SR approaches, respectively. Section 3 details the proposed algorithm. Experimental results and comparisons are demonstrated in Sec. 4. Finally, Sec. 5 concludes this paper.

2 Related Work

2.1 Local Regression Model

The local regression model relates two dependent measurements, y_i and x_i, via an unspecified regression function $f(\cdot)$

$$y_i = f(x_i) + \varepsilon_i, \tag{1}$$

in which ε_i is the estimation error and ideally $\varepsilon_i \to 0$ under the smoothness assumption. One typical application for this model is to predict the function value $f(x)$ at any point x from the sample pairs (x_i, y_i) $i = 1, \cdots, J$. Without any specific assumptions on the samples or regularization terms, this prediction model is usually solved by an N-term Taylor series, which approximates the observation y_i as follows:

$$y_i = f(x_i + x - x) + \varepsilon_i$$
$$= f(x) + f'(x)(x_i - x) + \frac{1}{2} f''(x)(x_i - x)^2 + \cdots$$
$$+ \frac{1}{N!} f^{(N)}(x)(x_i - x)^N + \varepsilon_i, \tag{2}$$

where $f'(\cdot)$ and $f^{(N)}(\cdot)$ denote the first and Nth derivatives of the regression function. Because Eq. (2) is derived from a local signal representation of the regression model, it is reasonable to estimate the unknown coefficients $\{f^{(n)}(x)\}_{n=0}^{N}$ by using the samples in the vicinity. With respect to this notion, Eq. (2) is reformulated as a weighted least-squares optimization problem

$$\min_{\{f^{(n)}(x)\}_{n=0}^{N}} \sum_{i=1}^{J} \left[y_i - f(x) - f'(x)(x_i - x) - \cdots \right. $$
$$\left. - \frac{1}{N!} f^{(N)}(x)(x_i - x)^N \right]^2 w_h(x_i - x), \tag{3}$$

where J is the number of samples that are near x, and $w_h(x_i - x) = \exp[-(x_i - x)^2 / 2h^2]$ represents the similarity between x_i and x. If we denote $\mathbf{y} = [y_1, y_2 \cdots, y_J]^T$, $\mathbf{b} = [f(x), f'(x), \cdots, f^{(N)}(x)]^T$, $\mathbf{W} = \mathrm{diag}\{[w_h(x_1 - x), w_h(x_2 - x), \cdots, w_h(x_J - x)]^T\}$, and

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 - x & \cdots & \frac{1}{N!}(x_1 - x)^N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_J - x & \cdots & \frac{1}{N!}(x_J - x)^N \end{bmatrix}.$$

Eq. (3) is then expressed in the following matrix form:

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_{\mathbf{w}}^2. \tag{4}$$

The function $\mathrm{diag}(\cdot)$ defines a diagonal operator that

$$\mathrm{diag}(\mathbf{A}) = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_J \end{bmatrix}$$

for a vector $\mathbf{A} = [a_1, a_2, \cdots, a_J]^T$. The solution of Eq. (4) is easily given by the weighted least-squares estimation

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}. \tag{5}$$

The approaches in Refs. 5–7 and 28 have validated the effectiveness of the local regression model for recovering complex structures in the SR problem by achieving higher-order derivative approximations ($N \geq 1$),[5] i.e., the first-order derivatives correspond to edge orientations and the second-order derivatives are related to curvatures.

2.2 Nonlocal Similarity in Self-Learning Superresolution Approaches

As illustrated before, the self-learning SR approaches have been developed to circumvent the false HF details from irrelevant training image by exploiting the multiscale

self-similarities in natural images.[22–24,27,35] In particular, Glasner et al.[22] have unified the reconstruction-based SR and the self-learning SR in the same framework, where the same-scale patch recurrence forms the constraints used in reconstruction-based SR and the across-scale patch recurrence generates exemplar LR-HR patch pairs used in the learning-based SR. By confirming that the local self-similarity assumption holds better for small zooming factors, Freedman and Fattal[23] have performed multiple magnification steps of small zooming factors to achieve the desired magnification size, wherein the example patch pairs were extracted from the LR input image and its smoothed version.

Both Glasner's and Freedman's works have confirmed the effectiveness of the internal patch recurrence in self-learning SR approaches. Further, Zontak and Irani[24] have shown that the image statistic that is derived from the internal patch recurrence is more predictive and powerful than the statistic derived from the external images. Accordingly, this powerful internal statistic is adopted by several self-learning SR approaches[25–27] as an effective regularization term. For example, by introducing the nonlocal similarity constraint, the nonlocal autoregression model (NARM)[25] has ensured the incoherence and sparsity of representation dictionary and consequently made the sparse representation model more suitable for the SR problem. Moreover, under the maximum *a posteriori* probability framework, Zhang et al.[26] have successfully assembled both the nonlocal prior that is obtained from the self-similarity in the same scale and the local prior that is derived from a specific regression model[6] to regularize the ill-posed SR problem. Later on, they have extended to multiscale self-similarity for a better image detail synthesis.[27]

3 Proposed Approach

The proposed approach is based on a local regression model and the multiscale nonlocal self-similarity, which will be described respectively in Secs. 3.1 and 3.2. Further, an adaptive patch processing strategy is provided in Sec. 3.3 to reduce the time consumption.

In the following, we use matrices $\mathbf{I}$ and $\mathbf{H}$ to represent the LR input and the HR output images; $\mathbf{L}_n$ is the smoothed version of $\mathbf{I}$, and $\mathbf{L}_{n+1}$ is an enlarged version of $\mathbf{I}$. The bolded lowercase $\mathbf{p}$ and $\mathbf{q}$ represent the column vectors of two $s \times s$ image patches that are extracted from $\mathbf{I}$ and $\mathbf{H}$, respectively; $\mathbf{m}$ and $\mathbf{n}$ denote the column vectors of two $s \times s$ image patches that are extracted from $\mathbf{L}_n$ and $\mathbf{L}_{n+1}$, respectively. Without specific notification, we assume that $\mathbf{p}, \mathbf{q}, \mathbf{m}$, and $\mathbf{n}$ are related. That is, $\mathbf{m}$ is the most similar patch to $\mathbf{n}$ among all other patches in image $\mathbf{L}_n$; patches $\mathbf{m}(\mathbf{n})$ and $\mathbf{p}(\mathbf{q})$ have the same coordinate for the center pixel, as presented by the dashed lines in Fig. 1. Given that image $\mathbf{I}$ has more HF details than $\mathbf{L}_n$, $\{\mathbf{m}, \mathbf{p}\}$ constitutes the self-exemplar LR-HR patch pairs in the proposed method. In addition, because $\mathbf{L}_{n+1}$ lacks the HF details in $\mathbf{H}$, $\{\mathbf{n}, \mathbf{q}\}$ is another group of LR-HR patch pairs.

3.1 *Local Patch Representation Based on a Local Regression Model*

As mentioned above, the task of learning-based SR approaches is to find the relations between LR-HR patches. In the proposed method, this relation is interpreted as an
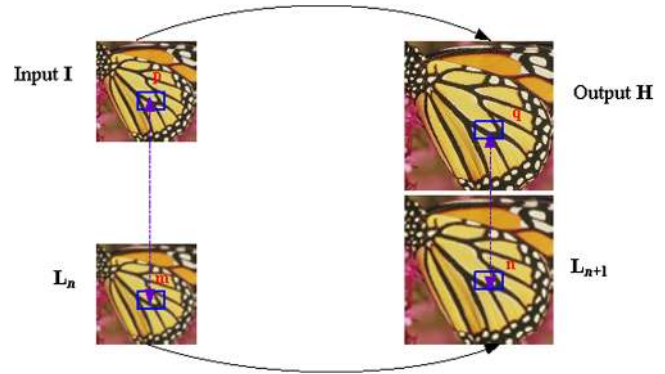

Fig. 1 Patch relations in the proposed method.

unspecified mapping function that associates each LR-HR patch pair $\{\mathbf{n}, \mathbf{q}\}$ in the following way $\mathbf{q} = f(\mathbf{n})$. The assumption that $\mathbf{m}$ and $\mathbf{n}$ are similar patches can also be interpreted as that $\mathbf{m}$ is "near" $\mathbf{n}$ in terms of geometric layout. Therefore, as in Sec. 2.1, this mapping function f can also be estimated via Taylor series

$$\begin{aligned} \mathbf{q} = f(\mathbf{n}) &= f(\mathbf{m} + \mathbf{n} - \mathbf{m}) \\ &= f(\mathbf{m}) + f'(\mathbf{m}) \odot (\mathbf{n} - \mathbf{m}) \\ &\quad + \frac{1}{2} f''(\mathbf{m}) \odot (\mathbf{n} - \mathbf{m}) \odot (\mathbf{n} - \mathbf{m}) + \cdots, \end{aligned} \quad (6)$$

where $\odot$ denotes the Hadamard product (also known as element-wise product). By neglecting noise and blurry in LR image, we come to $\mathbf{p} = f(\mathbf{m})$ (which is a very basic assumption in several self-learning based SR approaches, such as Refs. 23 and 28), Eq. (6) is therefore reformulated as

$$\mathbf{q} = \mathbf{p} + f'(\mathbf{m}) \odot (\mathbf{n} - \mathbf{m}) + \frac{1}{2} f''(\mathbf{m}) \odot (\mathbf{n} - \mathbf{m}) \odot (\mathbf{n} - \mathbf{m}) + \mathbf{r}_m, \quad (7)$$

where $\mathbf{r}_m$ represents the residual error. Although $\mathbf{m}$ and $\mathbf{n}$ are both taken from smooth images, it is agreed that $\mathbf{n}$ comparably contains more HF details than $\mathbf{m}$ does,[23] and therefore $\mathbf{n} - \mathbf{m}$ leaves only the HF information.[13,23,28] In view of this, Eq. (7), in fact, generates HR patches by imposing HF details to the related LR patches, which is actually the basic scheme of learning-based SR approaches.

From Eq. (7), we have to infer $f'(\mathbf{m})$ and $f''(\mathbf{m})$ first so as to obtain each HR patch $\mathbf{q}$. It is worth noting that for many existing SR approaches, only the first-order derivative is inferred. For example, in Refs. 13 and 23, the first-order derivative is assumed to be a constant, whereas in Ref. 28, this first-order derivative is estimated from an external database. However, it is generally agreed that only using the first-order derivative is insufficient to recover complex structures, such as curvatures; moreover, using external images is highly probable to introduce unfaithful image details. In the next subsection, we will present how to extend the first-order derivative estimation to the second order by exploiting the multiscale self-similarity property in a given LR image, and also provide a closed-form solution to this second-order estimation.

3.2 Derivative Estimation Based on Multiscale Local Self-Similarity

According to nonlocal patch redundancy, each patch is very likely to have many "neighbors," which are geometrically similar to it but are spatially far from it. Suppose $\mathbf{m}_i$ is a similar patch to $\mathbf{m}$ in image $\mathbf{L}_n$, as in Eq. (6), $f(\mathbf{m}_i)$ can be approximated as follows:

$$
\begin{aligned}
\mathbf{p}_i = f(\mathbf{m}_i) &= f(\mathbf{m} + \mathbf{m}_i - \mathbf{m}) \\
&= f(\mathbf{m}) + f'(\mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \\
&\quad + \frac{1}{2} f''(\mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) + \cdots \\
&= \mathbf{p} + f'(\mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \\
&\quad + \frac{1}{2} f''(\mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) + \xi_i.
\end{aligned} \tag{8}
$$

The column vector $\mathbf{p}_i$ is $\mathbf{m}_i$'s paired HR patch in image $\mathbf{I}$, and the column vector ξ_i also refers to the approximation error as $\mathbf{r}_m$ does in Eq. (7). As in Sec. 2.1, by incorporating the J-most similar patches $\{\mathbf{m}_i\}_{i=1}^{J}$ and their paired HR patches $\{\mathbf{p}_i\}_{i=1}^{J}$, Eq. (8) can be transformed into the following optimization problem:

$$
\begin{aligned}
\min_{f'(\mathbf{m}), f''(\mathbf{m})} \sum_{i=1}^{J} &\left\| \mathbf{p}_i - \mathbf{p} - f'(\mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \right. \\
&\left. - \frac{1}{2} f''(\mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \odot (\mathbf{m}_i - \mathbf{m}) \right\|_2^2 w_h(\mathbf{m}_i - \mathbf{m}),
\end{aligned} \tag{9}
$$

where $w_h(\mathbf{m}_i - \mathbf{m}) = \exp(-\|\mathbf{m}_i - \mathbf{m}\|_2^2 / 2h^2)$ measures the similarity between patches $\mathbf{m}$ and $\mathbf{m}_i$. Similar to Eq. (3), the above optimization problem could be easily solved by the weighted least-squares estimation that $\hat{\mathbf{b}} = [f'(\mathbf{m}), f''(\mathbf{m})]^T = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}$ as shown in Eq. (5). We should remind the reader that unlike Eq. (3), in which all the variables are scalars, the variables inside $\|\cdot\|_2^2$ in Eq. (9) are all in the column vector forms. Thus, we should redefine the specific formation of $\mathbf{X}$, $\mathbf{W}$, and $\mathbf{y}$ by taking account of the vector forms and the Hadamard product in Eq. (9). In the proposed method, these three variables are devised as follows:

$$
\mathbf{X} = \left\{
\begin{array}{cc}
\mathrm{diag}(\mathbf{m}_1 - \mathbf{m}) & \mathrm{diag}[(\mathbf{m}_1 - \mathbf{m}) \odot (\mathbf{m}_1 - \mathbf{m})] \\
\mathrm{diag}(\mathbf{m}_2 - \mathbf{m}) & \mathrm{diag}[(\mathbf{m}_2 - \mathbf{m}) \odot (\mathbf{m}_2 - \mathbf{m})] \\
\vdots & \vdots \\
\mathrm{diag}(\mathbf{m}_J - \mathbf{m}) & \mathrm{diag}[(\mathbf{m}_J - \mathbf{m}) \odot (\mathbf{m}_J - \mathbf{m})]
\end{array}
\right\},
$$

$$
\mathbf{W} = \left\{
\begin{array}{c}
\mathrm{diag}[w_h(\mathbf{m}_1 - \mathbf{m}) \times \mathbf{1}] \\
\mathrm{diag}[w_h(\mathbf{m}_2 - \mathbf{m}) \times \mathbf{1}] \\
\vdots \\
\mathrm{diag}[w_h(\mathbf{m}_J - \mathbf{m}) \times \mathbf{1}]
\end{array}
\right\}, \quad \text{and}
$$

$$
\mathbf{y} = \begin{bmatrix}
\mathbf{p}_1 - \mathbf{p} \\
\mathbf{p}_2 - \mathbf{p} \\
\vdots \\
\mathbf{p}_J - \mathbf{p}
\end{bmatrix}, \tag{10}
$$

where $\mathbf{1}$ is an $s^2 \times 1$ column vector with all elements equal to one, and $\mathbf{y}$ is of size $Js^2 \times 1$ that cascades each column vector $\mathbf{p}_i - \mathbf{p}$ $i = 1, \cdots J$.

After plugging the least-squares solution of Eq. (9) back into Eq. (8), the residual error ξ_i is easily obtained. Further, by simply averaging all the ξ_i $i = 1, \cdots J$ can the approximation error $\mathbf{r}_m$ in Eq. (7) be also obtained. In summary, we not only provide a closed-form solution of the derivative estimation, but also approximate the residual error that is caused by omitting higher-order derivative terms.

Intuitively, for the derivative estimation to be reliable, the number of independent equations in Eq. (9) must exceed the number of unknown derivative coefficients. With the assumption that $f'(\mathbf{m})$ and $f''(\mathbf{m})$ are not constant within the patch $\mathbf{m}$ (it is a reasonable assumption, for example, in an edge patch, the pixels across this edge are very likely to have larger first-order derivative than pixels along it), the number of unknown derivative coefficients is then related to the patch size and the derivative order. For instance, if a patch contains $s \times s$ pixels and a second-order derivative estimation is required, then the derivative coefficients are up to $2s^2$, implying that the demanded number of equations or the number of similar patches $\mathbf{m}_i$ should be above $2s^2$. Naturally, one question arises: are we able to find such enough similar patches in a single image? A statistical analysis made in Ref. 22 shows that the smooth patches recur much more frequently than the detailed patches do. However, in our case, we have to find the same patch numbers to ensure a reliable second-order derivative estimation, no matter whether these patches are highly detailed or not. Recently, Zontak and Irani[24] have shown that the patches with different gradient magnitudes need to search at different distance to ensure the same number of similar patches, and have also provided an empirical function. According to this function, we surprisingly find that the search region should be even beyond the image size for some highly textured patches so as to get a reliable second-order derivative estimation. This reveals that for highly textured patches, search for their similar patches in the same image scale is very insufficient.

To overcome this insufficient patch problem, we extend the same-scale local search to multiple scales. More specifically, based on the gradual magnification scheme, not only the down-scaled images of the current image resolution, but also all the images from the previous estimates, including the initial input image, are employed to find the similar patches. Take a two-iteration estimate for example. In the first iteration, we respectively search the $\lceil 2(s^2 + 1)/3 \rceil$ similar patches in $\mathbf{L}_n^0$ and its two down-scaled images to estimate $f'(\mathbf{m})$ and $f''(\mathbf{m})$ (these two down-scaled images of $\mathbf{L}_n^0$ are obtained by down sampling with the bicubic interpolation under scale 0.9^{sd}, sd = 1, 2). After that, the intermediate HR image $\mathbf{L}^1$ is generated by merging all the reconstructed HR patches according to Eq. (7), with a fusion of the overlapping regions between the adjacent patches. In the second iteration, we search the $\lceil (s^2 + 1)/2 \rceil$ similar patches, respectively, in $\mathbf{L}_n^0$, $\mathbf{L}_n^1$ and the two down-scaled versions of $\mathbf{L}_n^1$ to estimate the second derivatives. Subsequently, we obtain the second intermediate HR image $\mathbf{L}^2$, and generate the HR output image by resizing $\mathbf{L}^2$ to the desired size using the bicubic interpolation. We can see that for the derivative estimation in higher resolution levels, the required number of similar

patches with respect to each image in the training set is reduced because more images in lower resolutions can be utilized.

Although incorporating multiscale images guarantees a reliable derivative estimation, the additional exhaustive search for similar patches in other images would inevitably increase time complexity. Therefore, similar patches in coarser scales and lower resolutions are selected at the "relative" coordinates. For instance, in the second iteration process of the previous example, if the similar patch $\mathbf{m}_i$ found in $\mathbf{L}_n^1$ is centered at (j, k), then the additional similar patch $\mathbf{m}_i^{ds}$ in the down-scaled image of $\mathbf{L}_n^1$ is centered at $(\lfloor j \times 0.9^{sd} \rfloor \lfloor k \times 0.9^{sd} \rfloor)$, where sd = 1, 2 denotes the down-scale level, and the similar patch $\mathbf{m}_i^{ls}$ in the lower-resolution image $\mathbf{L}_n^0$ is centered at $(\lfloor j/t \rfloor \lfloor k/t \rfloor)$, where $t > 1$ denotes the gradual up-scaling factor.

3.3 Computational Speed-Ups

In Sec. 3.1, we translate the estimation of the relations between LR-HR patches to the estimation of the regression function derivatives by leveraging a local regression model. In Sec. 3.2, we further adopt the nonlocal similarity across multiple image scales to provide a closed-form solution.

In this subsection, an adaptive patch processing strategy is introduced to reduce the time consumption.

Intuitively, there is no need to estimate the high-order derivatives for smooth regions since their intensity values are nearly constant. Moreover, since natural images generally contain much more smooth regions than the detailed ones,[22,28,36] it is efficient to process different patch types (smoothed or detailed) in different ways. To discriminate the textured areas from the flat areas, an effective pixel classification criterion[37] TM is adopted, which provides a quantitative assessment of image geometric structures using the structure tensor and the intensity standard deviation of each patch. More specifically, $\text{TM} = [(\tilde{s}_{i,1}^2 - \tilde{s}_{i,2}^2)/(\tilde{s}_{i,1}^2 + \tilde{s}_{i,2}^2)] \times \{1 - [1/(1 + \zeta_i/255^2)]\}$, in which $\tilde{s}_{i,1}$ and $\tilde{s}_{i,2}$ are the eigen values of the structure tensor computed from each image patch, and ζ_i is the second moment of a gray-level cumulative histogram of that patch. By analyzing the cumulative TM histogram, three regions with different texture magnitudes are obtained:

$$(i, j) \in \begin{cases} c_1 & \text{TM}(i, j) > T_1 \\ c_2 & T_2 < \text{TM}(i, j) \leq T_1 \\ c_3 & \text{TM}(i, j) \leq T_2 \end{cases}, \qquad (11)$$

Algorithm 1 The proposed method.

Input: LR image $\mathbf{I}$, patch size $s \times s$, overlapping step o, search region size $K \times K$, gradual up-scaling factor t, and desired up-scaling factor r.

Output: HR image $\mathbf{H}$.

(1) **Initialization:**

    (a) Set $T = \lceil \log(r)/\log(t) \rceil$ and $\mathbf{L}^0 = \mathbf{I}$.

(2) **Iteration**:

    For iter = 0 to $T - 1$ do

    (a) Calculate $\mathbf{L}_n^{\text{iter}}$ by convolving $\mathbf{L}^{\text{iter}}$ with a Gaussian filter.

    (b) Calculate TM for $\mathbf{L}_n^{\text{iter}}$, and obtain thresholds T_1 and T_2 from the cumulative TM histogram.

    (c) Calculate $\mathbf{L}_{n+1}^{\text{iter}}$ by upscaling $\mathbf{L}^{\text{iter}}$ with the bicubic interpolation under the $t\times$ magnification factor.

    (d) Partition $\mathbf{L}^{\text{iter}}$ and $\mathbf{L}_n^{\text{iter}}$ into $s \times s$ image patches in raster-scan order to construct the HR and LR training set $P^{\text{iter}} = \{\mathbf{p}\}$ and $Q^{\text{iter}} = \{\mathbf{m}\}$, respectively; partition $\mathbf{L}_{n+1}^{\text{iter}}$ into $s \times s$ image patches with o pixels overlapped in raster-scan order to construct the LR test patch set $R^{\text{iter}} = \{\mathbf{n}\}$.

    (e) For each patch $\mathbf{m}$ in Q^{iter}: if its center pixel's TM is lower than T_2, assign $f'(\mathbf{m}) = 0$, $f''(\mathbf{m}) = 0$, and $\mathbf{r}_m = \mathbf{0}$; if its center pixel's TM is larger than T_1, search a total number of $\lceil (2s^2 + 2)/(\text{iter} + 3) \rceil \times (\text{iter} + 3)$ K-NN neighbors in $\mathbf{L}_n^{\text{iter}}$, the two down-scaled versions of $\mathbf{L}_n^{\text{iter}}$, and the lower-resolution images $\{\mathbf{L}_n^i\}_{i=0}^{\text{iter}-1}$, after that, these similar patches and their paired HR patches are used to compute $f'(\mathbf{m})$, $f''(\mathbf{m})$, and ξ_i according to Eqs. (10), (5), and (8), and $\mathbf{r}_m$ is obtained by averaging all ξ_i s; if its center pixel's TM is in between, only $f'(\mathbf{m})$ and ξ_i are estimated with a total number of $\lceil (s^2 + 1)/(\text{iter} + 2) \rceil \times (\text{iter} + 2)$ similar patches in $\mathbf{L}_n^{\text{iter}}$, its one down-scaled image, and the lower-resolution images $\{\mathbf{L}_n^i\}_{i=0}^{\text{iter}-1}$, assign $f''(\mathbf{m}) = 0$, and $\mathbf{r}_m$ is obtained by averaging all ξ_i s.

    (f) For each patch $\mathbf{n}$ in R^{iter}, search its nearest neighbor $\mathbf{m}$ in Q^{iter}, generate HR patch $\mathbf{q}$ using Eq. (7). Current HR image $\mathbf{H}_{\text{temp}}$ is obtained by merging all $\mathbf{q}$ s.

    (g) Image estimate update: $\mathbf{L}^{\text{iter}+1} = \mathbf{H}_{\text{temp}}$.

    End for

(3) **Interpolation:** Use bicubic interpolation to resize image $\mathbf{L}^T$ to the desired size for the final HR image $\mathbf{H}$.

---

where TM (i, j) is the TM value of the pixel at location (i, j); T_1 and T_2 are the bin values that correspond, respectively, to 70% and 30% in the cumulative TM histogram; c_1, c_2, and c_3 respectively represent the hard-detailed, medium-detailed, and smooth regions.

After the classification, an adaptive derivative order estimation scheme is implemented: for hard-detailed regions, we estimate their second-order derivatives to recover complex structures; for medium-detailed regions, a first-order derivative estimation is sufficient; for smooth regions, we simply paste the results from the previous HR estimation. Experiments show that such an adaptive patch processing strategy greatly reduces the time consumption of second-order scheme, which estimates the second-order derivative for both hard- and medium-textured regions, to 0.625 times without any obvious visual quality compromise. Algorithm 1 describes the entire SR process of the proposed method.

4 Experimental Results and Analysis

To evaluate the effectiveness of the proposed method, experiments on 14 standard images (shown in Fig. 2) are conducted. Several state-of-the-art SR approaches, including Glasner's method,[22] Fattal's method,[10] Freedman's method,[23] the GPR method,[29] the SC-based method,[19] Shan's method,[9] Kim's method,[21] the ANR method,[30] and the NARM method,[25] are selected as comparison baselines. The results of the first three approaches are straightly downloaded from http://www.cs.huji.ac.il/~raananf/projects/lss_upscale/sup_images/index.html and for the rest six approaches we run the source codes that are available at their authors' homepage.

Two quantitative measures, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index,[38] are adopted to evaluate the objective performance. A high PSNR score indicates that the reconstructed HR image contains little distortion, and a SSIM value near 1 implies that the resultant HR image has a very similar structure to the ground truth. It is worth noting that due to the gradual magnification scheme, any minor structure distortions and estimation errors may propagate and accumulate over several iterations.[28] Thus, the HR image obtained by the proposed method is inconsistent with the initial LR image. To alleviate inconsistency, a simple back-projection policy[19] is adopted to postprocess the reconstructed HR image. Unless otherwise specified, the presented SR results are postprocessed by the back-projection.

4.1 Experimental Configuration

In the following experiments, we straightly use the images shown in Fig. 2(b) as the LR input images, while the other LR input images are generated by down sampling (using the bicubic interpolator) the images shown in Fig. 2(a). In the proposed method, the patch size is 5×5 and four pixels are overlapped between the adjacent patches. To reduce the computational complexity, similar patches are found within the 13×13 searching region. The parameter h in Eq. (9), which is used to compute the similarity, is fixed to 10.14. The gradual upscaling factor is set to 1.5. The smooth image $\mathbf{L}_n^{\text{iter}}$ is generated by convolving with a Gaussian filter with a standard deviation of 0.65. The upscaled image $\mathbf{L}_{n+1}^{\text{iter}}$ is obtained by bicubic interpolation.

(a)

(b)

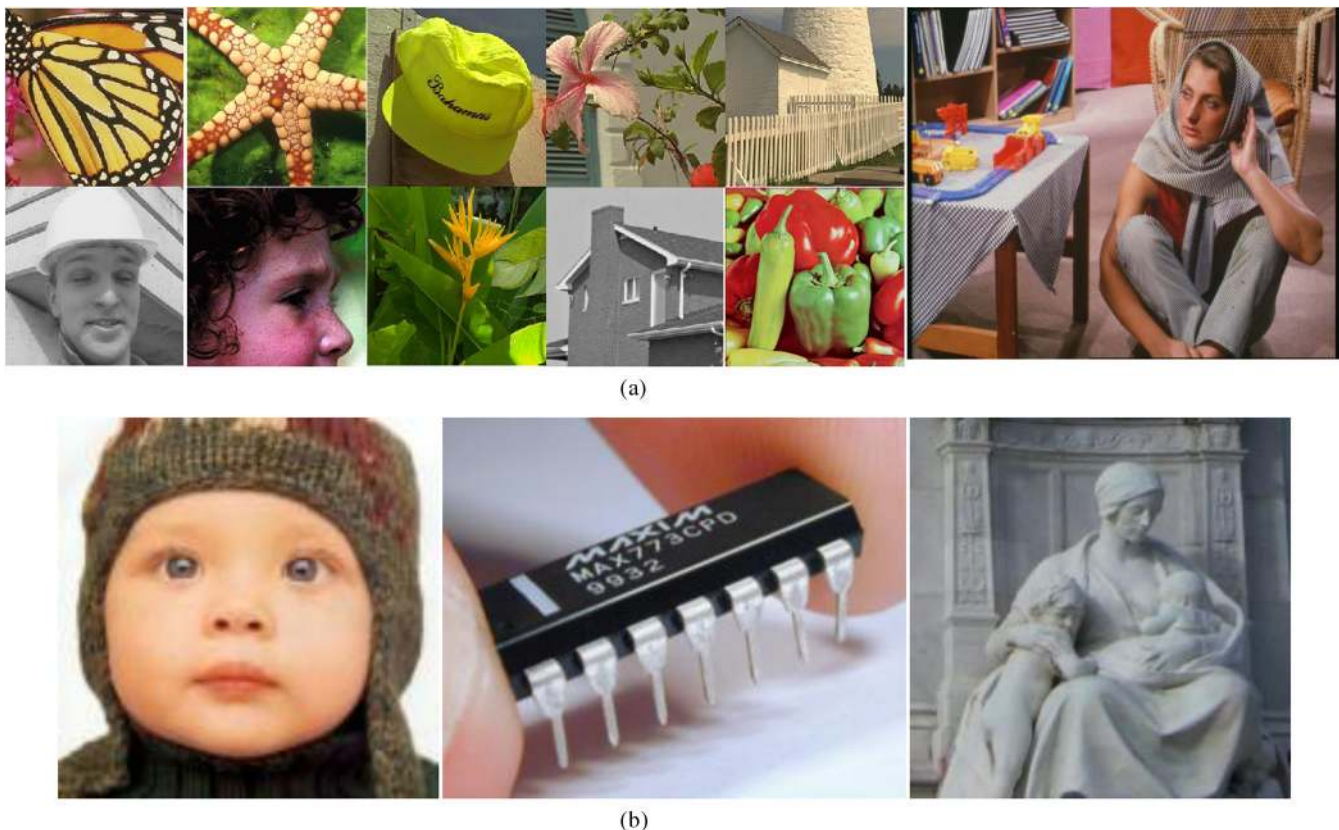**Fig. 2** Test images used in our experiments. (a) Original HR images used in 2× and 3× magnification case. (b) LR images used in 4× magnification case.

4.2 Experimental Results

For color images, we only super resolve their luminance channel since human eyes are more sensitive to illuminant changes. Accordingly, we compare the quantitative difference only on the luminance channel between the ground truth and the HR output image.

Table 1 summarizes the quantitative comparisons on images shown in Fig. 2(a). Among those listed approaches, ANR,[30] Kim's method,[21] and SC-based[19] are all training-based SR approaches, whereas GPR,[29] Shan's method,[9] NRAM,[25] and the proposed approach are training-free. It is worth noting that Kim's method not only uses external dataset as ANR[30] and SC-based[19] do, but also postprocesses the resultant image by imposing an image edge prior. As in Table 1, Kim's method on average gives the best PSNR and SSIM scores. However, even for the image Butterfly, where Kim's method presents nearly the highest gain in both PSNR and SSIM over the proposed method, we observe no apparent visual difference between these two methods (see Fig. 3 for details), despite that there are some jaggies in our method. Nevertheless, such jaggy effects are in fact introduced by the back-projection process, as illustrated in Figs. 3(b) and 3(c). Since jaggies are highly probable to cause the degradation of the objective performance of the proposed method, we believe that using certain de-jagged techniques could further improve our method, a most recent one of which can be found in Ref. 39.

In general, among all the training-free SR approaches, our method gives the best performance in terms of PSNR and SSIM. As for the other training-based SR approaches, we have to point out that they use thousands of training images to assure that enough relevant information can be utilized in the SR reconstruction. However, by using a single image only, the proposed method is able to reach a comparable objective performance, which validates that the relevant example LR-HR patch pairs can be found by fully exploiting the multiscale patch similarity.

For visual comparison, we carry out 3× magnification on LR images obtained from Fig. 2(a) and 4× magnification on images in Fig. 2(b). Since the pretrained dictionaries in the SC-based[19] approach are only for a zoom factor of 2, this approach is excluded in the following visual comparisons. In the first 3× magnification case, the proposed method is compared with the GPR method,[29] Shan's method,[9] Kim's method,[21] the ANR method,[30] and the NARM method.[25] Figure 4 shows the experimental results on images Barbara, Girl, Fence, and Hat, in which the local regions of interest (ROIs) in blue boxes are presented at the lower corner for providing a better comparison. Based on a universal dictionary, which is pretrained from thousands of images, the ANR method[30] appears to generate natural-looking results. Nevertheless, it is clumsy in preserving sharp edges, e.g., the blurry eyelashes in image Girl, due to the fact that a universal dictionary fails to accurately represent the complex details. Leveraged with a generic image prior model, Kim's method can effectively preserve some tiny details (see the ROI in image Girl), but it is prone to create ringing artifacts along some complicated edges, e.g., the characters in the image Hat. As for the NARM[25] and Shan's methods,[9] they are both somewhat prone to smooth HF details. The GPR method[29] seems to have difficulty in synthesizing faithful details, i.e., the characters in image Hat are badly recovered. By contrast, the proposed algorithm is able to provide sharper edges and more photorealistic details than other methods. For example, in images Barbara and Fence, the details shown in our ROIs are more distinctive in comparison with other methods; in images Girl and Hat, very fine details,

Table 1 Comparisons on peak signal-to-noise ratio (PSNR) values (dB) and structural similarity (SSIM) values (PSNR/SSIM) for 2× magnification.

|	ANR[30]	GPR[29]	Kim[21]	The proposed	SC-based[19]	NARM[25]	Shan[9]
Starfish	30.958/0.933	25.842/0.857	31.638/0.941	32.053/0.944	31.527/0.939	26.358/0.855	25.568/0.843
Butterfly	29.176/0.946	23.047/0.880	30.448/0.959	29.943/0.956	29.926/0.955	24.055/0.889	23.2920.877
Fence	24.990/0.805	21.223/0.683	24.897/0.809	25.146/0.814	24.940/0.810	22.483/0.708	21.280/0.698
Foreman	37.806/0.962	32.918/0.929	38.469/0.962	38.133/0.963	38.535/0.965	33.113/0.926	32.460/0.921
House	34.390/0.911	28.907/0.857	35.050/0.914	34.445/0.912	34.790/0.914	30.129/0.863	29.271/0.852.
Girl	32.353/0.843	31.739/0.810	32.328/0.845	32.201/0.843	32.305/0.844	29.537/0.753	29.073/0.753
Barbara	27.213/0.865	23.863/0.769	27.173/0.868	27.237/0.868	27.342/0.867	25.219/0.792	24.590/0.766
Flower	31.238/0.925	26.160/0.830	31.819/0.934	31.625/0.932	31.521/0.929	27.010/0.840	26.250/0.831
Pepper	32.703/0.902	29.218/0.865	33.662/0.904	33.904/0.906	32.875/0.903	29.969/0.860	29.128/0.852
Hat	32.348/0.917	28.805/0.866	33.061/0.927	32.630/0.922	32.715/0.922	29.152/0.863	28.617/0.849
Plants	35.563/0.959	29.796/0.894	36.657/0.963	36.161/0.961	36.092/0.959	30.280/0.890	29.670/0.883
Average	31.704/0.906	27.410/0.840	32.291/0.911	32.134/0.911	32.048/0.909	27.937/0.840	27.200/0.829
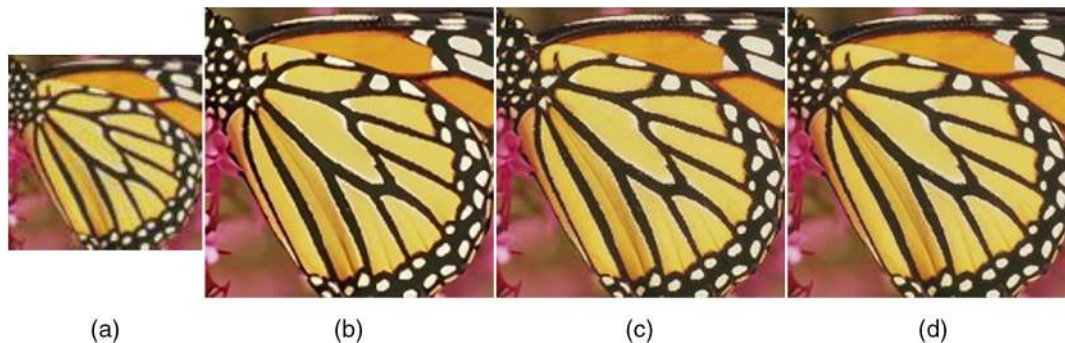
Fig. 3 Quantitative and qualitative SR comparisons (2×) on image Butterfly. (a) LR image. (b) Our result before back-projection (PSNR: 25.902 dB, SSIM: 0.926). (c) Our result after back-projection. (PSNR: 29.943 dB, SSIM: 0.956). (d) Kim's method[21] (PSNR: 30.448 dB, SSIM: 0.959). Note that the artifacts in (c) are, in fact, introduced by back-projection.

such as eyelashes and the contour of characters, are well preserved.

In addition, the objective comparisons of the HR results under 3× magnification are listed in Table 2, where our result appears not as superior as in the case of 2× magnification. However, we also find that our method, in fact, provides the best appealing visual details, i.e., the ROIs in Fig. 4(d) are consistently more vivid than those in Fig. 4(b), although the latter gives the almost highest PSNR and SSIM values; Hat image is poorly reconstructed by GPR but it has higher PSNR value than ours. This paradox suggests that the objective quality degradation of our method can be only attributed to the inconsistency caused by the gradual magnification scheme. Although this inconsistence can be alleviated by the back-projection policy, Freedman and Fattal[23] have further claimed that such inconsistency should be alleviated at an early stage. Their claim also explains why our method gives superior quantitative performance in 2× magnification case but fails to do so in 3× magnification case, since in 2× magnification, only two iterations are required in the proposed method.

To further inspect the effectiveness of our method, we continue to compare with more algorithms in Figs. 5–7 under a 4× magnification. Since the code of the NARM[25] method is only devised for zoom factors of 2 and 3, we also exclude this method in the subsequent visual comparisons. Among all the reconstructed HR images shown in Figs. 5–7, Freedman's method[23] tends to blur HF details since it simply assumes that the first-order derivative of the mapping function is a constant value. Although Fattal's method[10] applies imposed edge statistics to generate sharp edges, it produces staircase artifacts along edges as shown in Fig. 6(g). Glasner's method[22] avoids these unwanted staircase artifacts and provides somewhat pleasant results. Nevertheless, this particular method fails to reproduce plausible details, such as the character "9" in Fig. 7(f). On the contrary, as reflected in all the ROIs, the proposed method outperforms the others by providing more reasonable details. Although there are some jaggies in our results, such as Fig. 7(e), we should point out that they are mainly introduced by the simple back-projection process. As reflected in Fig. 8(b), prior to the back-projection, our result presents more distinguishable details. Although Fig. 8(b) presents a kind of overestimation, we in fact cause this artifact on purpose since the subsequent back-projection step weakens the details. Moreover, quantitative comparisons are also made. Since no ground truth is available in 4× magnification, we adopt a no-reference image assessment natural image quality evaluator (NIQE)[40] to indicate the quality of reconstructed HR image. NIQE reflects how the perceived image deviates from the statistical regularities calculated from natural images, a small value of which suggests a better image quality. From all the results in Figs. 5–7, we see that our method provides the lowest NIQE values.

To sum up, the advantages of the proposed method are attributed to the combination of the local regression model and the multiscale patch redundancy property: (1) we estimate the higher-order derivative of the mapping function, so that complex details such as curvature can be well preserved and enhanced; (2) exploiting the self multiscale redundancy helps to find more related patches than using an external database, without introducing false details.

4.3 Discussion on Down-Sampling Method

As illustrated in Ref. 7, the down-sampling method also affects the SR algorithms, because in the synthesis experiments, the LR input images are usually obtained by down sampling the original image. Therefore, if a proper down-sampling method is applied, the LR image will have fewer artifacts, which lead to a better reconstructed HR image. In Table 3, the influences of four down-sampling methods on the proposed method, namely, "Nearest," "Bilinear," "Bicubic," and "Lanczos," are quantified in terms of PSNR and SSIM, using the test images in Fig. 2(a) under a 3× magnification. Note that we do not post-process the SR results (back-projection) in this case, since we want to examine how the down-sampling method influences our algorithm. Surprisingly, the listed numerical results seem to be a violation of the claim in Ref. 7, since the simple bilinear method outperforms others. However, we notice that the SR results on images down sampled by bicubic are always sharper than the results by bilinear method, which, we believe, causes the degradation of objective performance of the bicubic method since overestimation may occur in some highly textured regions. However, we also found in our experiments that after the back-projection, the average PSNR gain obtained by bicubic over bilinear is higher than 4 dB. Therefore, in all of our experiments, the bicubic interpolation is used as the down-sampling method.
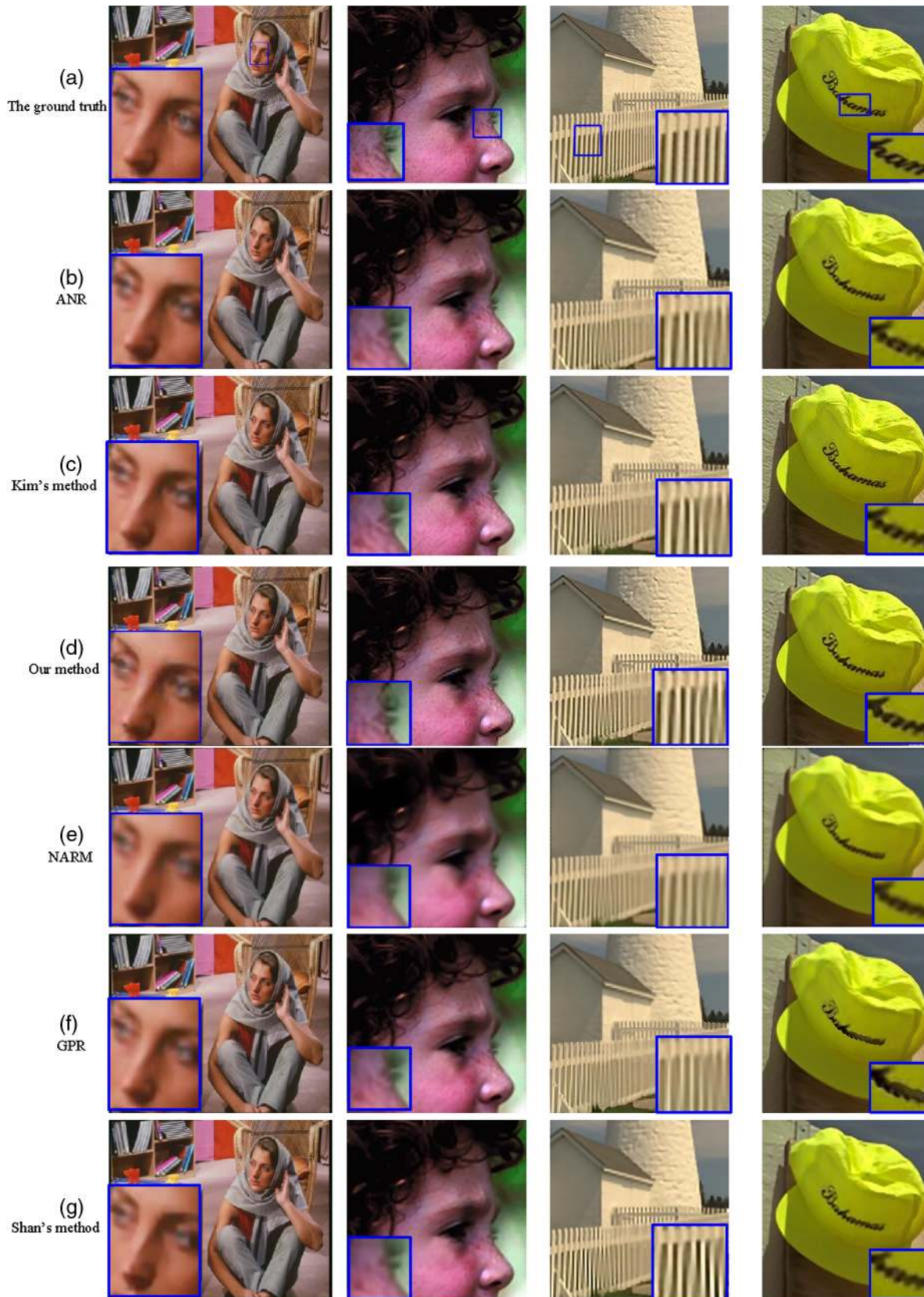
Fig. 4 Superresolution (3×) comparison on four images: Barbara, Girl, Fence, and Hat. (a) The ground truth; (b)–(g) the reconstructed HR results by ANR,[30] Kim's method,[21] the proposed approach, NARM,[25] GPR,[29] and Shan's method,[9] respectively.

Table 2 Comparisons on peak signal-to-noise ratio (PSNR) values (dB) and structural similarity (SSIM) values (PSNR/SSIM) for 3× magnification.

	ANR[30]	GPR[29]	Kim[21]	The proposed	NARM[25]	Shan[9]
Starfish	22.626/0.711	22.137/0.685	22.581/0.713	21.709/0.676	22.336/0.703	19.239/0.522
Butterfly	19.413/0.763	19.055/0.751	19.441/0.784	19.069/0.761	19.894/0.761	16.248/0.631
Fence	19.785/0.537	19.574/0.511	19.870/0.547	19.077/0.502	19.497/0.500	17.318/0.380
Foreman	29.471/0.869	29.217/0.865	29.313/0.869	28.573/0.854	27.641/0.859	26.098/0.797
House	25.287/0.779	25.311/0.775	25.023/0.777	24.451/0.757	25.352/0.774	21.780/0.674.
Girl	32.181/0.800	31.503/0.775	32.240/0.802	31.382/0.794	28.918/0.677	28.879/0.652
Barbara	25.310/0.768	24.662/0.735	25.127/0.767	25.455/0.774	24.106/0.710	22.658/0.653
Flower	23.436/0.694	23.462/0.673	23.266/0.693	22.625/0.657	23.321/0.687	20.460/0.517
Pepper	28.452/0.841	28.090/0.834	28.486/0.842	27.518/0.816	26.928/0.820	26.614/0.805
Hat	26.545/0.786	26.179/0.778	26.510/0.792	25.953/0.778	25.315/0.769	23.886/0.707
Plants	25.307/0.765	25.379/0.762	25.122/0.764	24.629/0.740	26.287/0.774	22.346/0.635
Average	25.258/0.756	24.961/0.740	25.180/0.759	24.586/0.737	24.509/0.730	22.320/0.634

4.4 Discussion on the Derivative Order

Here, we conduct an empirical study on the effect of derivative orders on the recovery of complex details. Four different derivative order settings in the nonsmooth patches are adopted for comparison, such as the zero-order estimation, the first-order estimation, the second-order estimation, and the adaptive order estimation used in the proposed method. The zero-order derivative estimation refers to assigning $f'(\cdot) = \mathbf{0}$, $f''(\cdot) = \mathbf{0}$, and $\mathbf{r}_m = \mathbf{0}$ in Eq. (7), and a similar notion goes for the first-order derivative estimation by assigning $f''(\cdot) = \mathbf{0}$. As indicated in Fig. 9, higher-order derivative estimation helps to produce fine textural details.
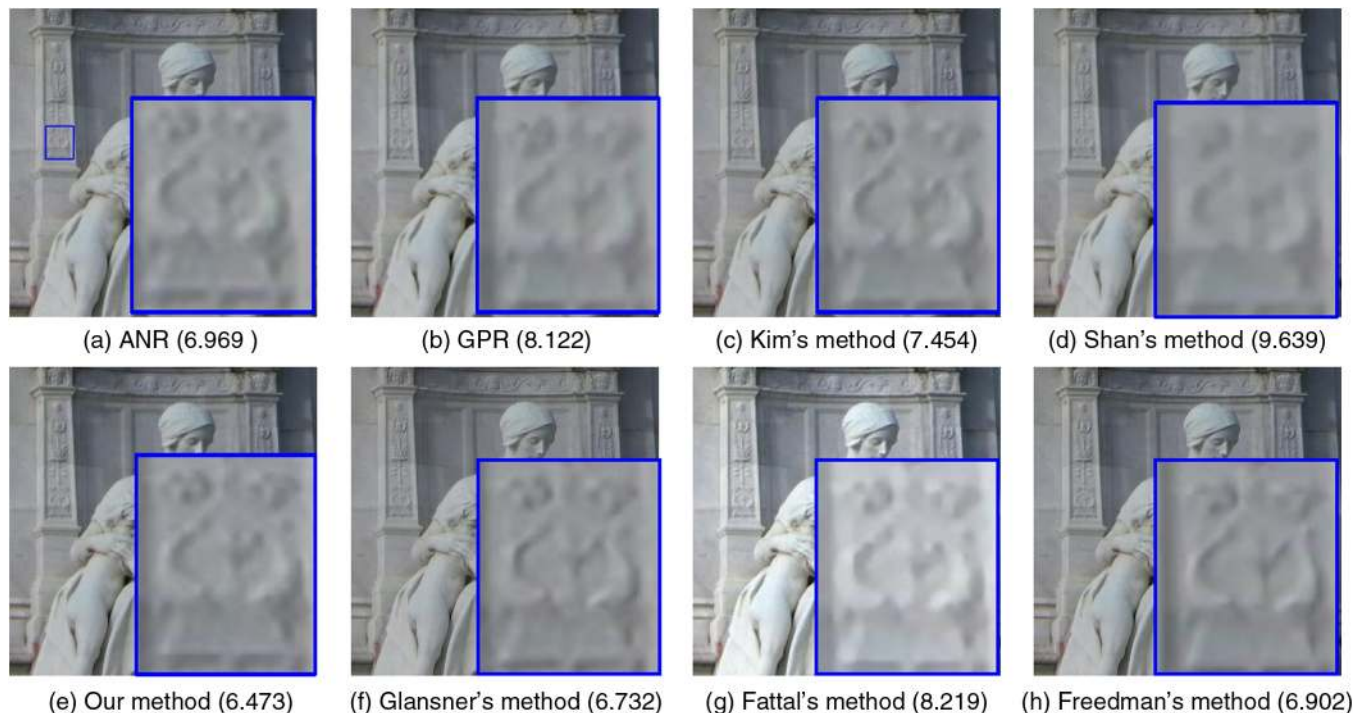

Fig. 5 Superresolution (4×) comparison on image Sculpture. (a)–(h) are the reconstructed HR results by ANR,[30] GPR,[29] Kim's method,[21] Shan's method,[9] the proposed approach, Glasner's method,[22] Fattal's method,[10] and Freedman's method,[23] respectively. The numbers in bracket denote NIQE[40] values.

(a) ANR (6.969) (b) GPR (8.122) (c) Kim's method (7.454) (d) Shan's method (9.639)
(e) Our method (6.473) (f) Glansner's method (6.732) (g) Fattal's method (8.219) (h) Freedman's method (6.902)
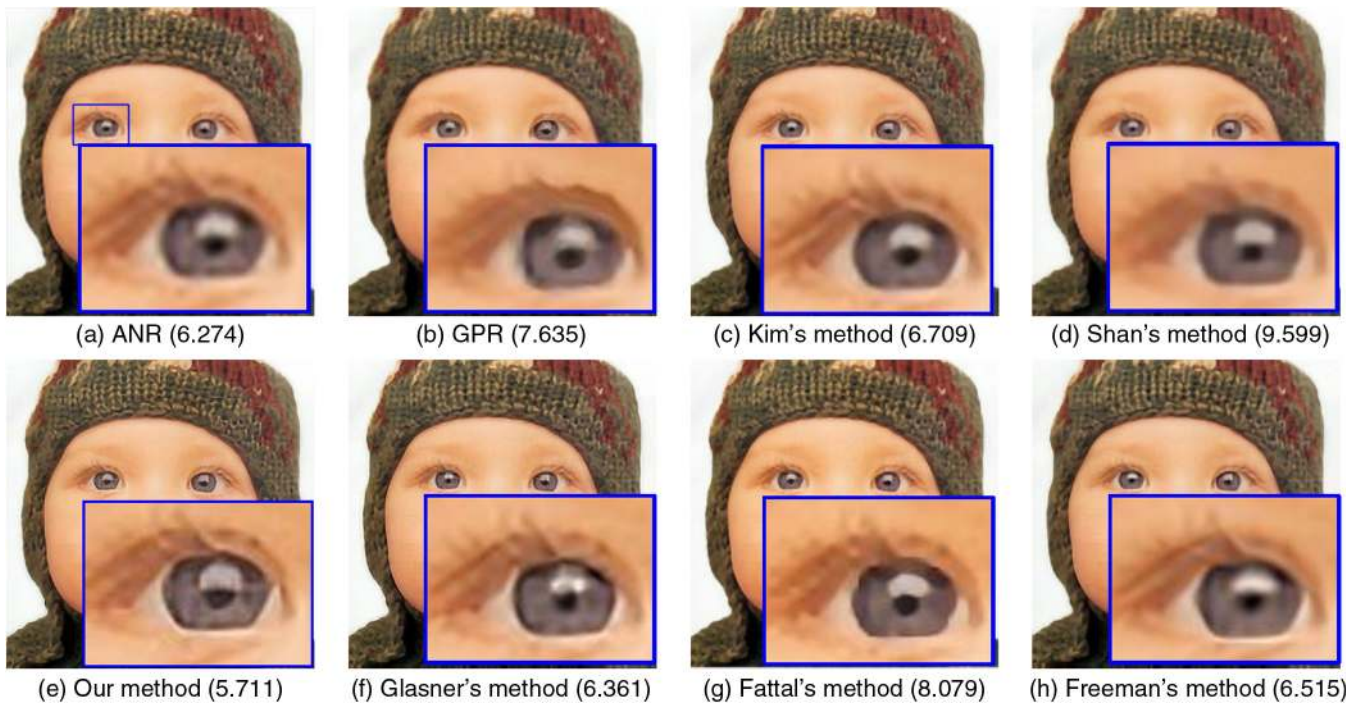
Fig. 6 Superresolution (4×) comparison on image Kid. (a)–(h) are the reconstructed HR results by ANR,[30] GPR,[29] Kim's method,[21] Shan's method,[9] the proposed approach, Glasner's method,[22] Fattal's method,[10] and Freedman's method,[23] respectively. The numbers in brackets denote NIQE[40] values.

For example, the zero-order estimation causes very severe block effects. On the other hand, the other three derivative estimation types successfully restore the complex details. Moreover, achieving second-order derivative estimation even preserves the subtle structure changes in kid's eyelashes, as shown in Figs. 9(c) and 9(d). Although there are no apparent visual differences between Figs. 9(c) and 9(d), the adaptive estimation strategy runs 1.6× faster than estimating second-order derivative on all detailed patches.

To sum up, the adaptive order strategy not only synthesizes complex details, but is also time efficient.

4.5 Discussion on Parameters

For the proposed method, it is important to choose proper parameters so as to obtain good SR performance. In general, there are six parameters to be determined: the patch size, the overlapping step, the deviation of Gaussian filter, the gradual
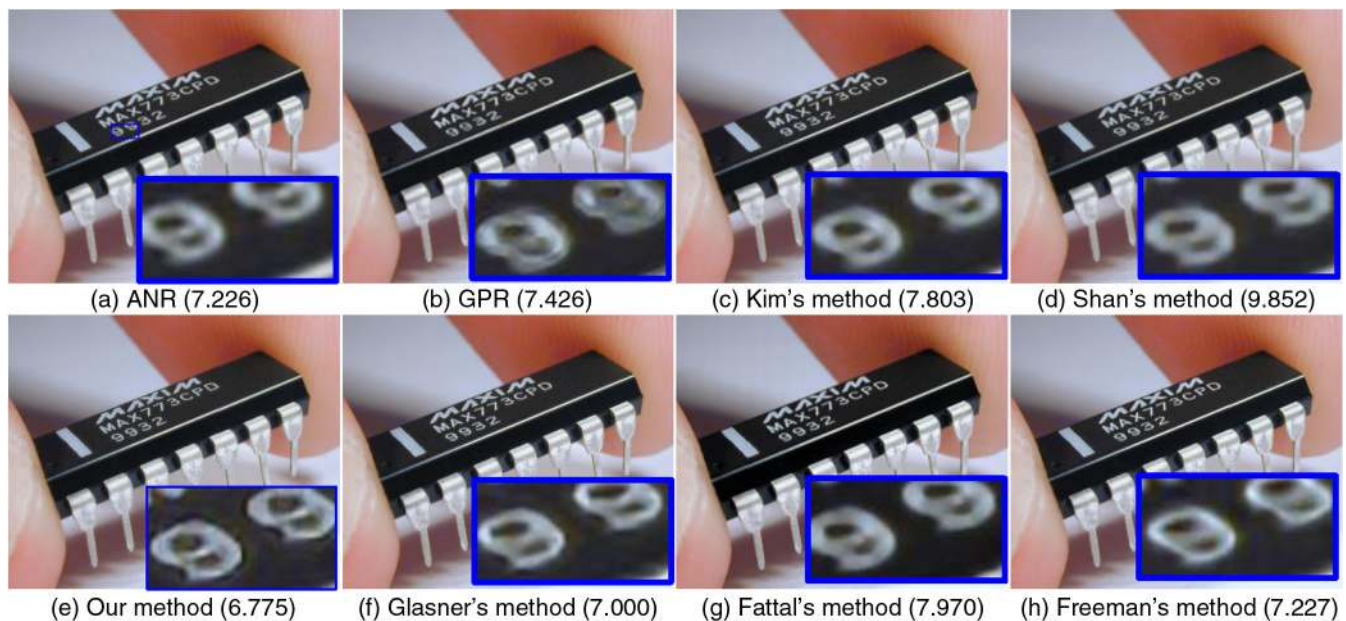

Fig. 7 Superresolution (4×) comparison on image Chip. (a)–(h) are the reconstructed HR results by ANR,[30] GPR,[29] Kim's method,[21] Shan's method,[9] the proposed approach, Glasner's method,[22] Fattal's method,[10] and Freedman's method,[23] respectively. The numbers in brackets denote NIQE[40] values.
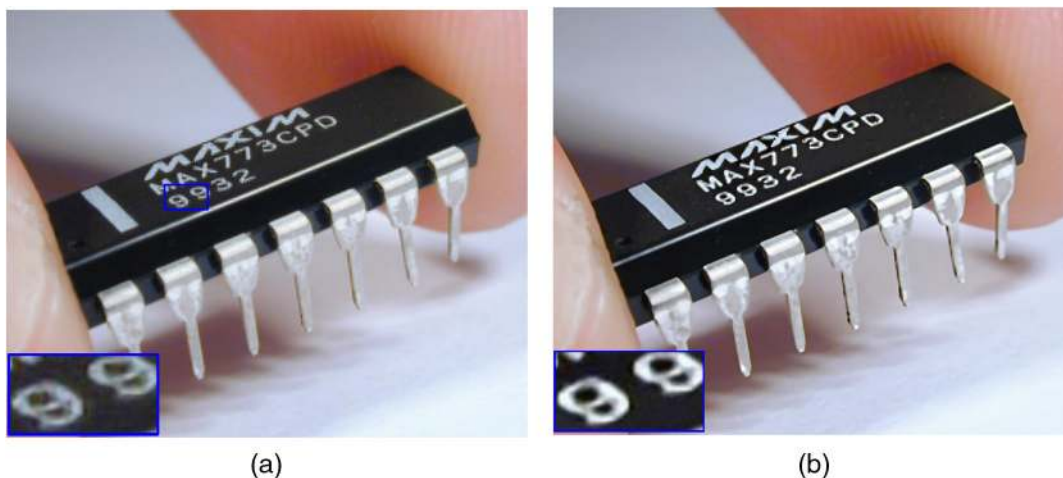
Fig. 8 Comparison of images. (a) after back-projection and (b) before back-projection.

Table 3 Comparisons of different down-sampling methods in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

	Down-sampling methods			
	Nearest	Bilinear	Bicubic	Lanczos
SSIM	0.630	0.703	0.680	0.695
PSNR (dB)	19.958	22.870	21.058	22.214

scaling factor, the size of search region, and the smoothing parameter h in Eq. (9) to compute similarity. We found that our algorithm is insensitive to the last three parameters in a reasonable range. In the following, we will elaborate the influences of the first three parameters, respectively.

4.5.1 *Influence of patch size and overlapping step*

Intuitively, using a large patch size tends to produce a smooth result and using a small patch size may introduce unwanted artifacts in flat areas. In the proposed method, the patch size also determines the number of derivative coefficients that need to be estimated. For example, if the patch size is a little large, say 9×9, for a second-order derivative estimation, we have to find at least 162 similar patches in multiscale versions of the input image. Finding such a large number of patches with high similarity is unrealistic. In addition, the ones that are not the first few nearest patches are very likely to be quite different from the target patch. In light of this, patch size cannot be very large in the proposed method. Therefore, we only discuss the effects caused by different patch sizes of 3×3, 5×5, and 7×7, and show their reconstructed results on image Girl and Hat under a 2× magnification in Fig. 10.

As shown in Fig. 10, the freckles in image Girl are well preserved by using patch size 3×3, whereas the other two patch size settings are prone to blurring such details. Nevertheless, some unwanted artifacts are introduced in image Hat by using patch size 3×3, such as the black dots along the hat edge, while the results obtained by other patch sizes are more pleasant. Besides, we have found that applying the patch size of 7×7 in our approach greatly increases the running time, which takes 10 times more than what is taken by running with 3×3 patches. Therefore, to balance the reconstruction quality and the time efficiency, we suggest using the patch size of 5×5 in the proposed method.
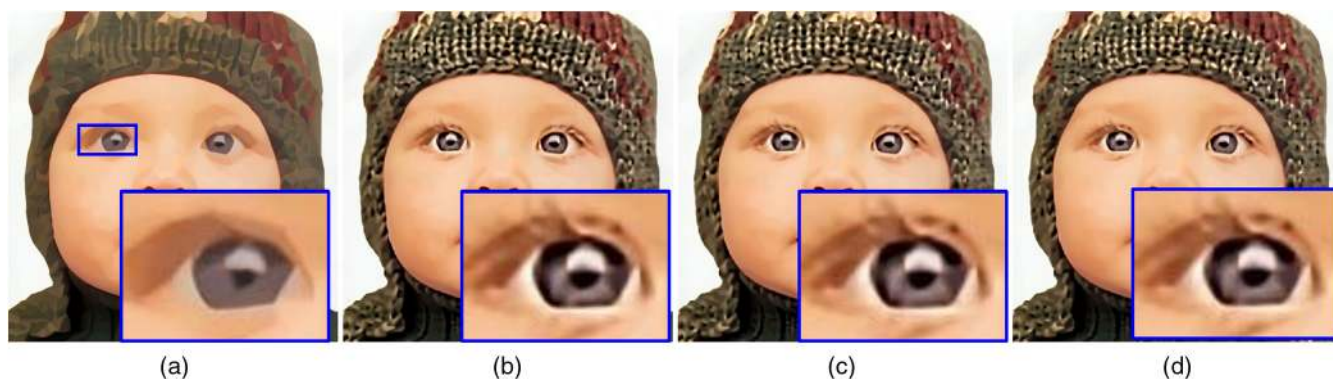

Fig. 9 The SR results (before back-projection) obtained by different derivative orders. (a) Zero order. (b) First order. (c) Adaptive order. (d) Second order. Note that we make the overestimate in (b)–(d) on purpose since the subsequent back-projection step weakens the details.
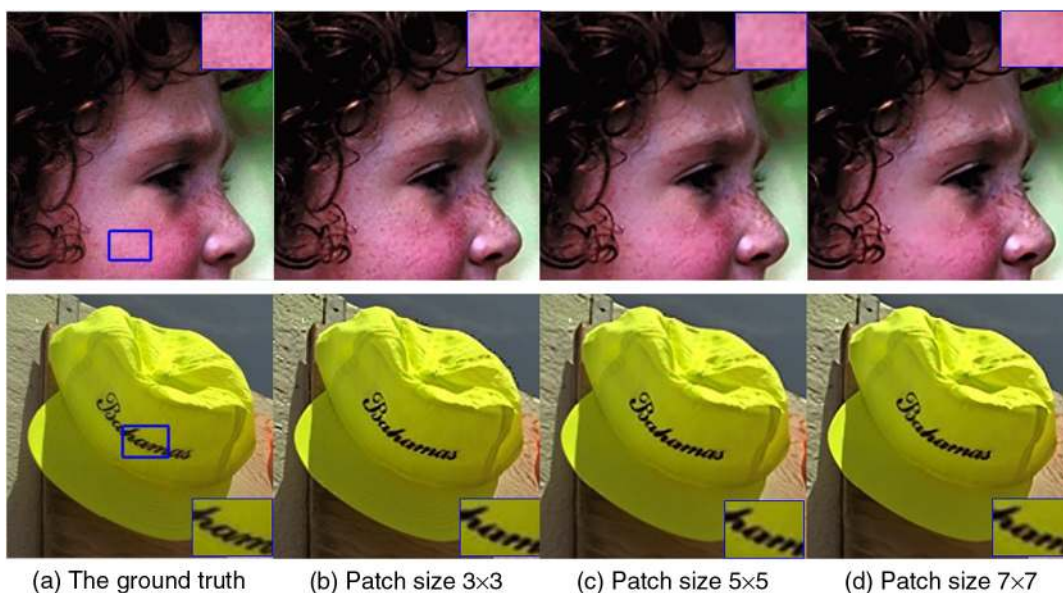
(a) The ground truth (b) Patch size 3×3 (c) Patch size 5×5 (d) Patch size 7×7

Fig. 10 The SR results (before back-projection) obtained by different patch sizes. (a) The ground truth. (b) Patch size is 3×3. (c) Patch size is 5×5. (d) Patch size is 7×7.

(a) (b) (c) (d) (e)

Fig. 11 The SR results (before back-projection) obtained by different Gaussian deviations: (a) 0.45, (b) 0.5, (c) 0.55, (d) 0.6, and (e) 0.65.

In addition, when patch size is fixed, a larger overlapping step leads to a better SR result. To this end, we use an overlapping step of four in the proposed method, as suggested in the SC-based SR[19] method.

4.5.2 *Influence of deviation of Gaussian filter*

In the proposed method, the Gaussian filter is used to generate the blurred image $\mathbf{L}_n$. A large deviation of the Gaussian filter leads to a blurrier $\mathbf{L}_n$, and consequently presents a more obvious HF enhancement according to Eq. (7). A supported example is shown in Fig. 11, in which a higher Gaussian derivation is prone to overestimation. However, considering that the subsequent back-projection will weaken the image details, we recommend a large Gaussian derivation in the first SR process.

5 Conclusion

In this paper, we present a new learning-based SR method for single image. Without any external training database, the proposed method establishes the relations between self LR-HR patches by leveraging the local regression model and the multiscale nonlocal patch redundancy. To better guarantee the cross-scale similarity, a gradual up-scaling scheme is also adopted, and previous estimates are accumulated for the next derivative estimation. In addition, an adaptive patch-based processing strategy is employed to reduce time consumption. Extensive experimental results, further, show how this proposed method improves over many existing learning-based SR approaches. In general, benefiting from a closed-form solution to high-order derivative estimation, the proposed method yields more photorealistic HR images with sharp details.

In the future, we plan to extend the proposed method in the following ways. (1) Adaptive patch size. As illustrated in Sec. 4.5.1, a large patch size is appropriate in smooth regions and a small patch size helps to recover subtle image details. Therefore, applying these two complementary properties in a combined adaptive patch size policy will further improve the proposed method. (2) Suitable patch features. Several works have pointed out that the way to represent the image patches also affects SR algorithms. In the proposed method, the patch itself is directly used as a basic feature. Other feature descriptors can also be employed, such as contrast-normalized patch and the patch derivative. (3) Contaminated image. Currently, the proposed method could only deal with clean images, and directly applying it to contaminated images

would either introduce unwanted artifacts, e.g., the noise amplification effect in noisy images, or no significant improvement in blurry images. For noisy images, the assumption in Eq. (6) that $\mathbf{p} = f(\mathbf{m})$ is no longer valid; for blurry images, the blur kernel used in generating $\mathbf{L}_n$ should be refined.[41] One possible alternative is to extend the nonlocal patch redundancy to image denoising and image deblurring, which has been very effective in many image processing works.[34,41–43]

Acknowledgments

References

1. T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Adv. Comput. Vis. Image Process.* **1**(2), 317–339 (1984).
2. L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.* **15**(8), 2226–2238 (2006).
3. X. Zhang and X. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.* **17**(6), 887–896 (2008).
4. X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.* **10**(10), 1521–1527 (2001).
5. F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image superresolution using multi-surface fitting," *IEEE Trans. Image Process.* **21**(7), 3312–3318 (2012).
6. H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.* **16**(2), 349–366 (2007).
7. T.-M. Kuo and S.-C. Tai, "On designing efficient superresolution algorithms by regression models," *J. Electron. Imaging*, **22**(3), 033002 (2013).
8. T. Q. Pham, L. J. van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP J. Appl. Signal Process.*, 83268, 1–12 (2006).
9. Q. Shan et al., "Fast image/video upsampling," *ACM Trans. Graph.* **27**(5), 1531–1537 (2008).
10. R. Fattal, "Image upsampling via imposed edge statistics," *ACM Trans. Graph.* **26**(3), 95 (2007).
11. S. Dai et al., "Soft edge smoothness prior for alpha channel super resolution," in *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 17–22, IEEE, Rio de Janeiro, Brazil (2007).
12. F. Liu et al., "Visual-quality optimizing super resolution," *Comput. Graph. Forum* **28**(1), 127–140 (2009).
13. W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.* **22**(2), 56–65 (2002).
14. H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 275–282, IEEE, Washington, DC, USA (2004).
15. J. Yu et al., "A unified learning framework for single image superresolution," *IEEE Trans. Neural Netw. Learn. Syst.* **25**(4), 780–792 (2014).
16. J. Yang et al., "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.* **21**(8), 3467–3478 (2012).
17. S. Ma et al., "Fast image super resolution via local regression," in: *Proc. IEEE International Conference on Pattern Recognition*, pp. 3128–3131, IEEE, Tsukuba, Japan (2012).
18. W. Dong et al., "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.* **20**(7), 1838–1857 (2011).
19. J. Yang et al., "Image super-resolution via sparse representation," *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010).
20. X. Gao et al., "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. Image Process.* **21**(2), 469–480 (2012).
21. K. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(6), 1127–1133 (2010).
22. D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. of IEEE Int. Con. on Computer Vision*, IEEE, Kyoto, Japan (2009).
23. G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.* **30**(2), 1–10 (2011).
24. M. Zontak and M. Irani, "Internal statistics of a single natural image," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 977–984, IEEE, Barcelona, Spain (2011).
25. W. Dong et al., "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.* **22**(4), 1382–1394 (2013).
26. K. Zhang et al., "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. Image Process.* **21**(11), 4544–4555 (2012).
27. K. Zhang et al., "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Image Process.* **24**(10), 1648–1659 (2013).
28. J. Yang, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 1059–1066, IEEE, Portland, Oregon, USA (2013).
29. H. He and W.-C. Siu," Single image super-resolution using Gaussian process regression," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 449–456, IEEE, Colorado, Springs, USA (2011).
30. R. Timofte, V. D. Smet, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. of IEEE Int. Con. on Computer Vision*, IEEE, Sydney, Australia (2013).
31. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technique overview," *IEEE Signal Process. Mag.* **20**, 21–36 (2003).
32. P. Milanfar, *Super Resolution Imaging*, CRC Press, Florida (2011).
33. Z. Lin and H. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(1), 83–97 (2004).
34. A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 60–65, IEEE, San Diego, California, USA (2005).
35. M. Zontak, I. Mosseri, and M. Irani, "Separating signal from noise using patch recurrence across scales," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 1195–1202, IEEE, Portland, Oregon, USA (2013).
36. Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.* **27**(5), 73 (2008).
37. J. Hu and Y. Luo, "Non-local means algorithm with adaptive patch size and bandwidth," *Optik* **124**(22), 5639–5645 (2013).
38. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
39. S.-C. Tai, T.-M. Kuo, and H.-J. Chen, "Artifact-free superresolution algorithm with natural texture preservation," *J. Electron. Imaging* **23**(1), 013020 (2014).
40. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013).
41. T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in *Proc. of IEEE Int. Con. on Computer Vision*, pp. 945–952, IEEE, Sydney, Australia (2013).
42. D. Van De Ville and M. Kocher, "Nonlocal means with dimensionality reduction and SURE-based parameter selection," *IEEE Trans. Image Process.* **20**(9), 2683–2690 (2011).
43. C. A. Deledalle and J. Salmon, "Non-local methods with shape adaptive patches (NLM-SAP)," *J. Math. Imaging Vis.* **43**(2), 103–120 (2012).

Jing Hu received her BE degree from the School of Automation at the University of Electronic Science and Technology of China, Sichuan, in 2009. She is working toward the PhD degree in the Department of Automation at Tsinghua University. Her current research interests include superresolution and image denoising.

Yupin Luo is a professor in the Department of Automation at Tsinghua University. He received his BS degree from Hunan University in 1982, and his EM and PhD degrees from Nagoya Institute of Technology, Japan, in 1987 and 1990, respectively. His current research interest is in the area of image processing, graph theory, and artificial intelligence.