Are Recent SISR Techniques Suitable for Industrial Applications at Low Magnification?

Wei Jia, *Member, IEEE,* Yang Zhao, *Member, IEEE,* Ronggang Wang, *Member, IEEE,* Hai Min, Shujie Li, and Xiaoping Liu

***Abstract*—In recent years, many deep-network-based super-resolution techniques have been proposed and achieved impressive results for $2\times$ and higher magnification factors. However, lower magnification factors encountered in some industrial applications have not received special attention, such as 720P-to-1080P ($1.5\times$ magnification). Compared to traditional $2\times$ or higher magnification factors, these lower magnifications are much simpler, but reconstructions of high-definition images are time consuming and computationally complex. Hence, in this paper, a fast image upsampling method is designed specifically for industrial applications at low magnification. In the proposed method, edge and nonedge areas are first distinguished and then reconstructed via different fast approaches. For the edge area, a local edge pattern encoding-based method is presented to recover sharp edges. For the nonedge area, a global iterative reconstruction with texture constraint is utilized. Moreover, some acceleration strategies are also presented to further reduce the complexity. The experimental results demonstrate that the proposed method can obtain performance comparable to that of some state-of-the-art methods for 720P-to-1080P magnification, but the computational cost is much lower.**

***Index Terms*—Super-resolution, low magnification, upsampling**

I. Introduction

The development of video display technology has undergone a series of changes: from black and white to color, from 2D to 3D, and from standard definition to high definition. With the rapid development of display screens, the resolutions of recent televisions and mobile devices are becoming increasingly higher. Unfortunately, many current videos are still low-resolution (LR) rather than full-high-definition (FHD) or ultra-high-definition (UHD). These LR resources cannot satisfy the demands of high-resolution (HR) devices. To overcome this problem, super-resolution (SR) has become a fundamental and important technique for computer vision, image processing, and related industrial applications. The goal of SR is to recover HR frames from LR frames, which is a typical ill-posed problem since the details are missing in LR images. At present, high-performance and low-cost SR techniques are still in high demand by the video industry.

Many SR approaches have been proposed over the past decades [1]. Many early works are based on multiframe SR [2]. To avoid the artifacts caused by the difficult alignment process in multiframe SR, single-image SR (SISR) methods have received more attention. Recent SISR methods can be roughly divided into three categories, *i.e.*, interpolation-based methods, reconstruction-based methods, and learning-based methods. Interpolation [3] is the most basic SISR method, where the missing pixels are filled by fixed interpolated kernels, but this approach often generates unnatural artifacts. The reconstruction-based methods were originally designed for multiframe SR tasks [2], and such methods utilize the similarity constraint that the upsampled result should be consistent with the LR input [4], [5]. However, the missing high-frequency (HF) information can hardly be reproduced with these two types of methods. Learning-based methods [6] have thus become the most popular approach to learn from inner or extra exemplars. Various typical learning-based SISR models have been proposed, such as neighbor-embedding-based methods [7], [8], sparse-representation-based methods [9]–[11], and local self-similarity-based methods [12]. However, these learning-based SISR methods are often time consuming because the weights or coefficients are optimized for each local patch. Aiming to avoid the time-consuming patch-by-patch optimization, many efficient SR techniques have been proposed, *e.g.*, anchored neighbor regression (ANR)-based methods [13], [14], random-forest-based methods [15], local-classification-based methods [16], and deep neural network (DNN)-based methods [17].

Over the past several years, increasingly more DNN-based SISR methods have been reported [18]–[27]. Due to their deeper architectures and larger datasets, current DNN models have a powerful fitting ability for nonlinear regression tasks. Hence, these DNN-based methods have achieved impressive performance, *e.g.*, the LeNet-5-model-based SRCNN [17], the VGGNet-model-based VDSR [18], the residual network (ResNet)-based SR-ResNet [19], and the efficient pixel shuffling network [20]. In 2017, an influential SISR competition was held [21], and many outstanding ResNet-based methods have since been proposed, such as EDSR [22], denoising-

This work is supported by the grants of the National Natural Science Foundation of China (Nos. 61673157,61877016), and the Fundamental Research Funds for the Central Universities (Nos. JZ2017HGTB0189, PA2018GDQT0014, JZ2018HGTA0215).(*Corresponding author:Yang Zhao*)

W. Jia, Y. Zhao, H. Min, S. Li, and X. Liu are with the School of Computer and Information, Hefei University of Technology, 193 Tunxi Road, Hefei 230009, China (e-mail: yzhao@hfut.edu.cn; jiawei@hfut.edu.cn;minhai361@aliyun.com;lisjhfut@hfut.edu.cn; lxp@hfut.edu.cn).

R. Wang is with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, 2199 Lishui Road, Shenzhen 518055, China (e-mail: rgwang@pkusz.edu.cn).

Y. Zhao and R. Wang are also with the Peng Cheng Laboratory, Shenzhen 518055, China.

W. Jia, S. Li, and X. Liu are also with Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei 230009, China.

network-based SR [23], the deep Laplacian pyramid network [24], the deep backprojection net [25], the gradual upscaling network [26], [36], and the residual dense network [27]. However, the main purpose of many SR industrial applications is to generate high-quality images with a low cost. Recent state-of-the-art DNN methods may have some weaknesses for specific industrial applications [28], [29], and these drawbacks are summarized as follows:

1) These networks are implemented in an end-to-end mode, but in practice, the LR inputs are often transmitted line-by-line and patch-by-patch rather than directly obtaining an entire frame.
2) The input image is processed simultaneously in multiple channels of the network. Hence, the memory cost of DNN-based methods is related to the number of channels and the resolution of the inputs. For example, SRCNN [17] employs 64 channels, and therefore, 64 frame buffers of 4K/2K resolution are required for UHD/FHD image SR.
3) Although these DNN-based methods have achieved much higher PSNRs than traditional methods, the subjective quality has not significantly improved for small magnification factors. In addition, the frames in the video are constantly moving; thus, the visual quality differences become even more difficult to perceive.

Current SISR methods mainly focus on $2\times$ or higher magnification factors, *i.e.*, $3\times$, $4\times$, and $8\times$. However, in some industrial applications, magnification factors may be much lower, such as $1.5\times$ (720P-to-1080P). For video industry, many 720P videos, which do not have corresponding 2K/4K sources, need to be upsampled via SR techniques. For display device industry, 720P video resources also need to be magnified to fit current FHD or FHD+ screens. This paper proposes a simple, efficient, and low-cost SISR method designed specifically for SR industrial demands with low magnification, which is named the FAst Super-resolution Technique (FAST). The proposed method does not focus on pursuing a higher PSNR value but tends to meet the requirements of related industrial applications. The experimental results demonstrate that the proposed method can achieve visual quality comparable to that of some state-of-the-art methods with a much lower cost.

The remainder of this paper is organized as follows. Section II analyzes the motivation for the proposed method. Section III presents the proposed FAST in detail, and the experimental results are discussed in Section IV. Finally, Section V concludes the paper.

II. The "Crime and Punishment" of PSNR

PSNR is the most commonly used and basic objective image quality metric, and it is widely applied in many computer vision and image processing scenarios. However, does PSNR really work well for SISR? In many SISR approaches [9], [11], [31], researchers have observed that PSNR values are often inconsistent with the subjective quality. In [37], it is mathematically proven that the common distortion measures are at odds with perceptual quality. Unfortunately, an accurate quality assessment for SR results is lacking. Consequently, PSNR is still widely adopted in SR methods and is even the most important measure for evaluating an SR method.

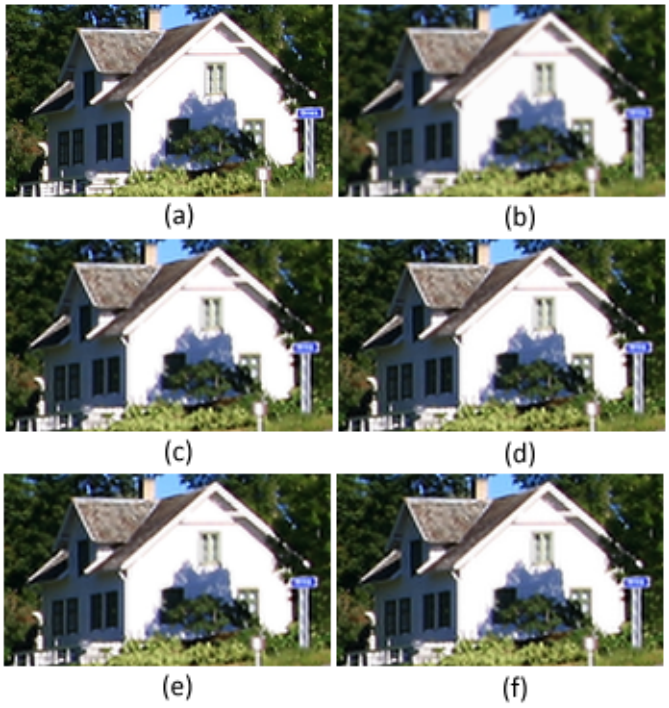

Fig. 1. The $1.5\times$ upsampled results of different methods. (a) Original HR image, (b) bicubic, (c) LPE [16], (d) SRCNN [17], (e) VDSR [18], and (f) ResNet.

Recent DNN-based SR methods keep increasing the record PSNR values [21].This is partially because the l_1/l_2-loss naturally satisfies the PSNR assessment since the training phase is directly designed to increase the PSNR. However, the increase in PSNR values does not correspond to a significant improvement in visual quality, especially for small magnification factors. Fig. 1 presents the $1.5\times$ upsampled results of some learning-based methods in 720P-to-1080P applications. When quickly examining Figs. 1 (d), (e), and (f), it is difficult to distinguish the differences among the results of various DNN-based methods, although these methods obtain clearly different PSNR values. Subjective testing has been performed for $1.5\times$ magnification of several 2K images, and the average mean opinion scores $(0 \sim 5)$ of bicubic, LPE, SRCNN, VDSR, and ResNet are 4.47, 4.94, 4.90, 4.93, and 4.95[1], respectively. This testing also reveals that the viewers cannot exactly distinguish the differences in visual quality of these methods when the upsampling factor is too small.

Therefore, the following question is raised. The DNN-based methods have high computational complexity and are highly dependent on GPU hardware, but the improvement in visual quality is not as conspicuous as the increase in the PSNR values. Are these complex networks really needed for SISR applications with small magnification factors, such as $1.5\times$? This question is the direct motivation for the fast and low-cost FAST method proposed in this paper.

[1]The implementation details and more discussions of the subjective testing can be found in the supplementary materials: http://yzhaocv.weebly.com/projectpage/fast.

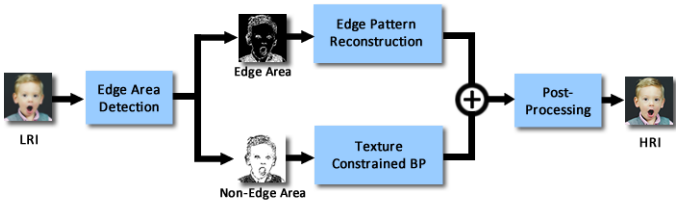

Fig. 2. The framework of the proposed method.

III. The Proposed Method

As illustrated in Fig. 3, the proposed FAST method consists of two streams of processes, *i.e.*, the edge pattern reconstruction for edge areas and texture-constrained backprojection (BP) for nonedge areas. For edge areas, an efficient local encoding process is utilized to encode different local edge patterns and then reconstruct each type of edge pattern with a learned projection matrix. For nonedge areas, which constitute most of a natural image, a simple and fast global BP method is introduced to refine the textural details and avoid noises in flat areas. Finally, an additional postprocessing is applied to remove the slight artifacts caused by combining two different types of methods. In the following, each component is discussed in detail.

A. Focusing on the Edge Patterns

In the patch-by-patch reconstruction process of traditional learning-based SR methods [9], [13], [16], [30], different types of local patches are treated as being equally important. However, the human visual system (HVS) has been proven to be more sensitive to sharp edge patterns [32], and these sharp edge patterns occupy only a small percentage of an image. Conversely, although flat and texture patterns account for most of a natural image, the reconstruction of these patches does not significantly improve the visual quality. Hence, in this paper, the input LR images are divided into edge and nonedge areas, and learning-based refinement is applied on only the edge patches.

The local gray-value difference is used to measure the degree of local variation, and it can be calculated as

$$s_{LD} = \frac{1}{9}\sum_{i=1}^{9}(g_i - \bar{g}), \tag{1}$$

where $g_i (i = 1, 2, ..., 9)$ denotes the pixels on a local 3×3 patch and $\bar{g} = \frac{1}{9}\sum_{i=1}^{9} g_i$. Note that this simple local gray-value difference s_{LD} is also reused in the following texture reconstruction and postprocessing to avoid extra computational cost. After s_{LD} of each pixel is computed, the average local gray-value variance of the entire image is then denoted as $s_{\bar{LD}}$. The edge patterns are accordingly defined as the patches that have a clearly larger local difference than other patches, as follows:

$$s_{LD} > \lambda s_{\bar{LD}}, \tag{2}$$

where λ is an artificial parameter to adjust the number of edge patterns. The percentages of edge patterns on the DIV2K dataset [21] with different λ are listed in Table I. For intuitively illustrating the results of edge pattern detection, an example is shown in Fig. 4. From Table I and Fig. 4, we find that a small λ leads to more edge patterns, whereas a large λ (*i.e.*, $\lambda = 5$) merely extracts the sharpest edges and omits other edges. To balance the computational cost and the coverage of detected edge patterns, the parameter λ is experimentally set as 2 in this paper.

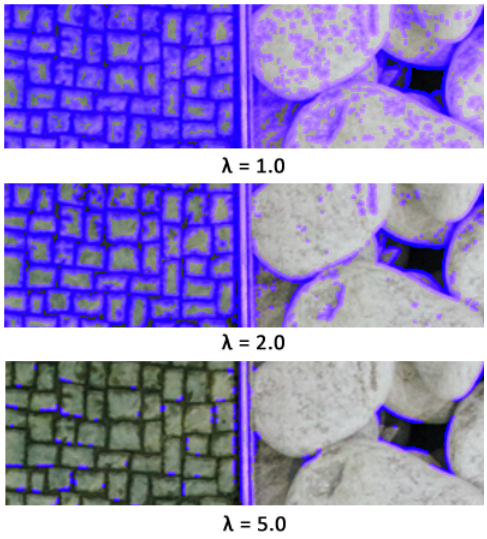

Fig. 3. Detected edge patterns with different λ (these edge patterns are marked in purple).

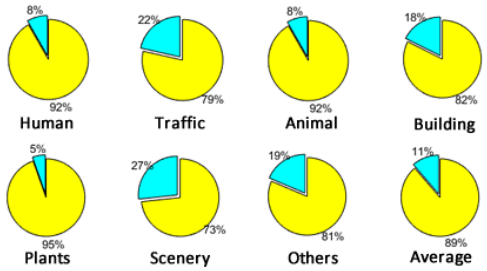

Fig. 4. The percentages of edge patterns on different types of 2K images.

TABLE I
Percentages of Edge Patterns with Different Values of Parameter λ (%)

λ	1.0	1.5	2.0	2.5	3.0	5.0	10.0
Percent	19.86	14.51	11.32	9.05	7.31	2.66	0.04

To illustrate the percentages of edge patterns on different types of images, several 2K images are classified into six common categories, *i.e.*, faces, traffic, animals, buildings, plants, view, and others. The percentages of edge patterns for each category are illustrated in Fig. 5. As shown, the edge patterns are often less than 20% of the whole image, and thus, focusing on the reconstruction of edge patterns can naturally accelerate the SR process.

B. Edge Pattern Encoding and Reconstruction

The local patch encoding (LPE)-based method [16] can effectively distinguish different local patches and learn the corresponding projection matrix for each class. However, the LPE method has some disadvantages for SISR applications. First, there are too many precomputed projection matrices (PPMs) that need to be calculated in LPE. For instance, 12-bit and 17-bit LPEs request 2^{12} and 2^{17} PPMs, respectively. Second, these PPMs are learned equally and do not focus on primary patterns. Third, the size of the PPMs is decided by the size of patches. For example, to magnify an $m \times m$ LR patch to an $n \times n$ HR patch, the PPM has a size of $m^2 \times n^2$. The complexity of reconstruction is thus $O(m^4 n^2)$. Hence, some methods [16] enforce the smallest 3×3 patch to accelerate the reconstruction process.

Based on these analyses, the proposed edge pattern reconstruction process adopts the following strategies to reduce the computational complexity and memory cost. First, the PPMs are only calculated for the edge patterns, which are often less than 20% in natural images. Second, a 9-bit code is introduced in the proposed FAST, and the pre-computed PPMs are thus $2^3 \sim 2^8$ times less than that of the LPE methods. Third, a small number of edge patterns are further manually selected according to their importance. Finally, to reduce the size of PPMs, the patch size is set as 3×3. This size is also naturally suitable for the proposed 9-bit encoding. Due to these effective strategies, the computational cost of the projection reconstruction process in FAST is much less than that of the LPE methods.

Motivated by the famous local texture descriptor local binary pattern, the simple 9-bit local encoding is computed as

$$c = \frac{1}{9}\sum_{i=1}^{9} s(g_i - \bar{g})2^{i-1}, s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

where g_i and $\bar{g}$ have been defined in Eqn. (1).

Since the encoding is calculated on a local 3×3 patch, there are a total of $2^9(512)$ types of codes. However, note that these patterns are not equally important. As shown in Fig. 6, the frequencies of various local edge patterns are counted on the DIV2K set [21]. Some principal patterns with a high frequency of occurrence can be found, while other patterns rarely appear in the edge area. In the proposed method, 55 types of the most frequent edge patterns are first manually selected from the total of 512 types of patterns. Although these patterns are merely one-eleventh of the total categories, their frequencies of occurrence account for more than 93.4% of the total number. There are mainly two reasons for why only 55 types of principal patterns are utilized. First, fewer edge patterns can further reduce the number of PPMs and the time cost of the projection reconstruction. Second, a PPM computed with insufficient exemplars may lead to reconstruction errors if the PPMs are computed for all the edge patterns. Fig. 7 illustrates the most frequent edge patterns, and these patterns contain several typical edge structures, such as tiny lines, spots, and corners. Note that the proposed 9-bit code is encoded in a zigzag mode, from top to bottom and left to right.

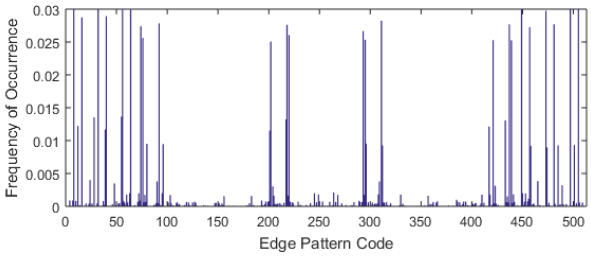

Fig. 5. The frequency histogram of various local edge patterns.

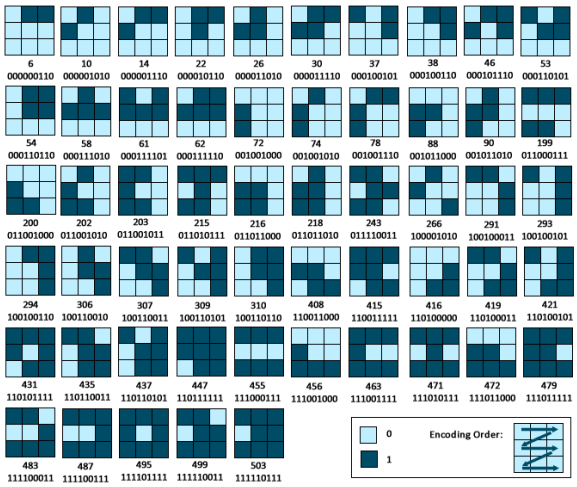

Fig. 6. The 55 types of most frequent edge patterns.

After the local encoding process, the edge patterns can be classified into a total of 55 categories according to their codes. The other low-frequency patterns are simply upsampled with bicubic interpolation. Then, a projection matrix can be computed for each class of edge patterns by utilizing traditional ridge regression or other embedding methods [13], [16], [38]–[43] . In this paper, $K(K = 4096)$ exemplars are randomly chosen from each class of edge patterns, and the LR patch $\boldsymbol{y}$ can be represented by means of the K LR samples $\boldsymbol{L}_K^c$ with the same code c,

$$\arg\min_{\alpha} ||\boldsymbol{y} - \alpha \boldsymbol{L}_K^c||_2^2 + \beta||\alpha||_2^2, \tag{4}$$

where β is a weighting factor to stabilize the solution and experimentally set as 0.1 in this paper. The closed-form solution of Eqn. (4) can be calculated as

$$\alpha = (\boldsymbol{L}_K^{c\,T}\boldsymbol{L}_K^c + \beta I)^{-1}\boldsymbol{L}_K^{c\,T}\boldsymbol{y}. \tag{5}$$

The HR patch $\boldsymbol{x}$ can then be estimated by the optimized weight α and corresponding K HR samples $\boldsymbol{H}_K^c$ of the c-th class, as follows:

$$\boldsymbol{x} = \boldsymbol{H}_K^c\alpha = \boldsymbol{H}_K^c(\boldsymbol{L}_K^{c\,T}\boldsymbol{L}_K^c + \beta I)^{-1}\boldsymbol{L}_K^{c\,T}\boldsymbol{y}. \tag{6}$$

The PPM $\boldsymbol{P}^c$ of the c-th class can thus be defined as

$$\boldsymbol{P}^c = \boldsymbol{H}_K^c(\boldsymbol{L}_K^{c\,T}\boldsymbol{L}_K^c + \beta I)^{-1}\boldsymbol{L}_K^{c\,T}. \tag{7}$$

In the training phase, a total of 55 PPMs $\{\boldsymbol{P}^c\}_{c=1}^{55}$ are precomputed by means of the randomly selected sample-pairs

in each class. In the SISR reconstruction phase, an LR edge patch $\boldsymbol{y}_{edge}^{c}$ only needs to be encoded to obtain its code c by means of Eqn. (3) and then multiplied by the corresponding PPM $\boldsymbol{P}^{c}$ to generate the HR edge patch $\boldsymbol{x}_{edge}^{c}$,

$$\boldsymbol{x}_{edge}^{c} = \boldsymbol{P}^{c}\boldsymbol{y}_{edge}^{c}. \tag{8}$$

C. Texture-Constrained BP for Nonedge Areas

Flat and texture areas occupy the majority of natural images. For flat areas, it is clear that complicated reconstruction is not helpful for improving the visual quality. The main purpose of flat area SR is to avoid noise, banding, and ringing artifacts caused during the process. However, texture areas contains various and abundant changes in local differences. Hence, the SR of texture areas should focus on recovering missing HF components. BP is a traditional and fast method for enhancing textural details, which is mainly based on the global similarity constraint. However, conventional BP also tends to produce undesirable artifacts on edge and flat areas. To inherit the advantages and simultaneously overcome the disadvantages of the traditional BP method, an additional texture structure constraint is presented to constrain the HF details to be consistent with the original image texture, as follows:

$$\boldsymbol{X}^{*} = \arg\min_{\boldsymbol{X}} ||DH\boldsymbol{X} - \boldsymbol{Y}||_2^2 + \lambda_T ||TDH\boldsymbol{X} - T\boldsymbol{Y}||_2^2, \tag{9}$$

where D denotes a downsampling operator; H is a blurring filtering operator; $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote HR image and LR image, respectively; T denotes a local gray-value difference extraction operator;, and λ_T is a parameter to balance the global similarity and texture stability. In this paper, the local gray-value difference of each pixel is also calculated by means of Eqn. (1) such that the precomputed s_{LD} can be reused in this BP process without introducing extra computational cost. Eqn. (9) can then be estimated by updating $\boldsymbol{X}$ using an iterative formula as

$$\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \lambda_T H^T U \boldsymbol{T}_{\boldsymbol{Y}}(DH\boldsymbol{X}_t - \boldsymbol{Y}), \tag{10}$$

where $\boldsymbol{X}_t$ is the estimated HR image after the t-th iteration and U denotes the upsampling operator. $\boldsymbol{T}_{\boldsymbol{Y}}$ represents the local gray-value difference matrix of $\boldsymbol{Y}$, in which each element is calculated by Eqn. (1). $\boldsymbol{T}_{\boldsymbol{Y}}$ can be regarded as a texture coefficient matrix. In flat areas, the coefficients in $\boldsymbol{T}_{\boldsymbol{Y}}$ are small, and these small coefficients can suppress noises. In nonflat texture areas with large local differences, the coefficients are large to enhance the HF details. The texture areas can then be reproduced by performing Eqn. (10) within appropriate iterations. In our experiments, λ_T is typically set as $0.1/\bar{m}$, where $\bar{m}$ denotes the mean value of coefficient matrix $\boldsymbol{T}_{\boldsymbol{Y}}$.

Fig. 8 illustrates the SR results of traditional BP and the proposed BP with texture constraint. By comparing the close-ups of texture areas (Figs. 8(b), (c)), it can be found that the proposed BP can recover clearer texture than the traditional BP and bicubic methods. Figs. 8 (e) and (f) show examples of upsampled flat areas, where we can find that traditional BP enhances some noises in flat areas, while the proposed BP can effectively suppress them by applying texture structure as an extra constraint.

Fig. 7. Upscaled results of different BP methods. The first row shows close-ups of texture areas with different methods: (a) bicubic, (b) traditional BP, (c) the proposed BP. The second row shows close-ups of flat areas with different methods: (d) bicubic, (e) traditional BP, (f) the proposed BP.

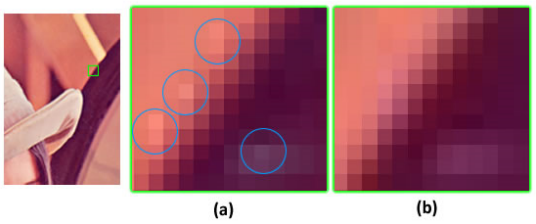

Fig. 8. Close-ups of boundary of edge and nonedge areas: (a) without the postprocessing, (b) final result with the post-processing.

D. Postprocessing

In the proposed hybrid framework, the edge and nonedge areas are processed with different methods. Consequently, the direct combination of these two areas may cause some nonsmooth and discontinuous artifacts around the combination boundary. Postprocessing is thus utilized to remove these potential artifacts.

In the postprocessing, an additional BP is used as the global similarity constraint to suppress the nonsmooth artifacts. However, the global iterative process also damages the reconstructed edges. Hence, the additional BP process is not applied around edges. The iterative formula is described as follows:

$$\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + rH^T U[M(1-\boldsymbol{E})(DH\boldsymbol{X}_t - \boldsymbol{Y})], \tag{11}$$

where $\boldsymbol{E}$ is the binary map of the edge area, M denotes morphological eroding operation, and the iterative parameter r is set as 0.01. Indeed, the nonsmooth artifacts are quite inconspicuous if we do not magnify the local details. The fast postprocessing only iterates a few times ($2 \sim 5$ times), and it can effectively eliminate these slight artifacts. In Fig. 9(a), some nonsmooth pixels are marked with blue circles. These

artifacts can be removed by the postprocessing, as illustrated in Fig. 9(b).

IV. 720P-to-1080P SISR Experiments

A. Test Image Sets and Compared Methods

Current SISR methods are mainly tested on commonly used datasets for $2\times$, $3\times$, and $4\times$ magnifications, *i.e.*, Set5, Set14 [10], and B100 [13]. The proposed method is specifically designed for industrial applications with lower magnification, which is often neglected in traditional SISR methods. Hence, the experiments are implemented for a 720P-to-1080P ($1.5\times$) task. Several 2K images are randomly selected from the DIV2K validation set [21] for testing. These 2K images are first downsampled to 720P to generate LR input and then upsampled with different SISR methods. For color images, only the Y channel is magnified with these methods, and the Cb and Cr channels are simply upsampled via bicubic interpolation.

To validate the effectiveness of the proposed method, FAST is compared with several recent state-of-the-art methods, *i.e.*, ANR [13], LPE and LPE_P_2 [16], SRCNN [17], VDSR [18], and the ResNet-based method [22]. Note that these learning-based methods are designed for $2\times$ or larger magnification, and thus, these methods have been retrained on DIV2K. The DNN-based methods SRCNN, VDSR, and ResNet are already trained for the $2\times$ task and then fine-tuned with $1.5\times$ samples. The structure of the residual blocks of ResNet is set as the EDSR [22] rather than ordinary SR-ResNet [19], but note that this ResNet is implemented without the pixel-shuffling process utilized in EDSR. Thus, the input 720P image is first interpolated to 1080P with bicubic interpolation, and then the network is trained to directly reconstruct the 1080P input.

B. Subjective Testing

As mentioned previously, the PSNR cannot accurately characterize the visual quality of the SR results. Hence, subjective testing is also applied to evaluate the performance. In this paper, the subjective testing is implemented in two ways, *i.e.*, direct rating and pairwise comparison. A total of 20 subjects are invited to participate in the experiment, and the tested images are displayed on a 72-inch UHDTV with NVIDIA Titan X GPUs[2].

For direct rating, the original 2K images are provided as references. The viewers are then required to rate the images from 1 to 5 by comparing with the ground truth. The mean opinion scores are shown in Fig. 10, from which we can find that although the ground truth is also provided, most of the viewers still cannot distinguish the $1.5\times$ upsampled results of the state-of-the-art methods and the original HR images. The proposed FAST, the recent LPE, and the DNN-based methods have all been scored with near full marks. This result indicates that the subjective quality of these methods is good enough for $1.5\times$ magnification.

Fig. 11 presents the upsampling results of various methods, from which we can obtain the following observations. First,

[2]More details can be found in the supplementary materials.

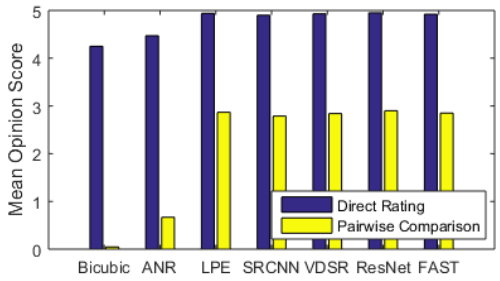

Fig. 9. The mean opinion scores of different SISR methods.

although bicubic interpolation produces noticeable blur and jagged artifacts for $2\times$ and higher magnifications, the $1.5\times$ magnified result of bicubic interpolation appears to be satisfactory. This result is because the $1.5\times$ task is much easier than traditional $2\times$, $3\times$, and $4\times$ magnifications. Second, LPE, DNN-based methods, and the proposed method can reproduce clearer and sharper results than the bicubic and ANR methods. Unfortunately, it is quite difficult for the viewers to indicate which of the results is better than the others.

Fig. 12 shows the SR results on sharp edge areas. By comparing the upsampled characters, it can be found that the bicubic interpolation blurs the edges, and the LPE method, the DNN-based methods, and the FAST method still recover sharp and natural results. Note that ResNet tends to sharpen tiny lines, and the reconstructed characters appear to be even sharper than the original HR ground truth. By comparing the HR ground truth and the upsampled results with the FAST method, the proposed method can obtain high fidelity results.

In the paired comparison testing, the upsampled results of two different methods are simultaneously displayed on the screen, and the viewer only needs to mark which one is visually better. Each method is compared with all other methods, and then the compared results are accumulated to calculate the scores, as follows:

$$S_j = \sum_{i=1, i\neq j}^{N_m} C(a_j, a_i),\ C(a,b) = \begin{cases} 1, & if\ a\ is\ better \\ 0, & else \end{cases} \tag{12}$$

where N_m denotes the total number of tested methods, a_j is the tested method, and a_i denotes another method. Note that some unreasonable cases may exist, particularly when the tested images have similar quality. For instance, if a_{j1} is better than a_{j2} and a_{j2} is better than a_{j3}, then a_{j1} should be better than a_{j3}. When a_{j3} is scored higher than a_{j1}, an unreasonable circular triad case occurs. In this paper, the Kendall coefficient of consistency (KCC) [33] is used to measure the consistency of the subjective scores, which is calculated as

$$\delta = 1 - \frac{24N_c}{N_m(N_m^2 - 1)}, \tag{13}$$

where N_c is the number of unreasonable cases. In the experiment, we found that most of the failure cases appear during the comparisons of the LPE, FAST, and DNN-based methods. This result also demonstrates that these methods have comparable visual quality. The final score can then be

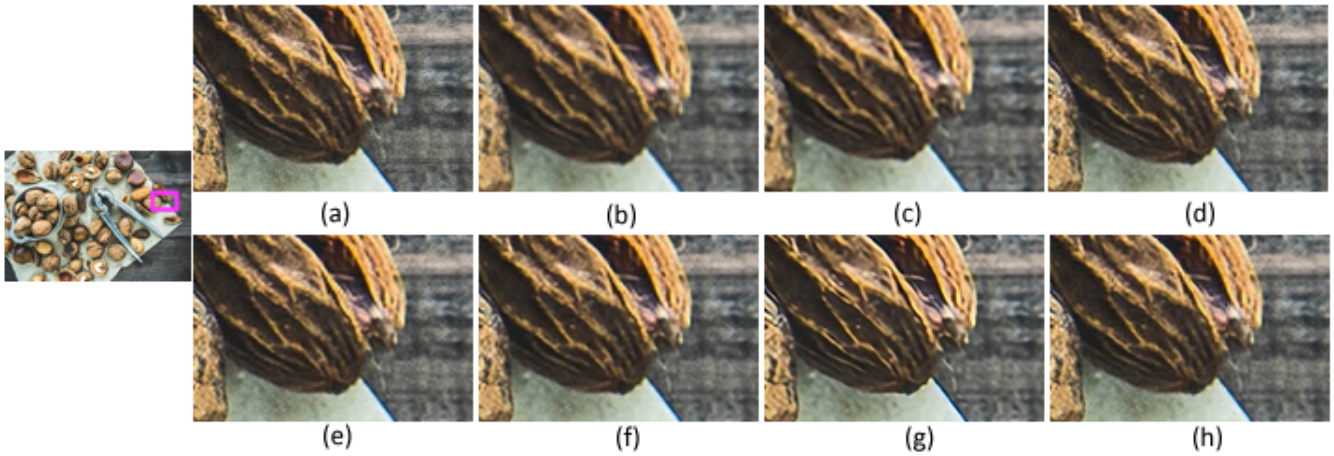

Fig. 10. The $1.5\times$ upsampled results of different methods. (a) The HR images, (b) bicubic, (c) ANR [13], (d) LPE [16], (e) SRCNN [17], (f) VDSR [18], (g) ResNet, (h) the proposed FAST.

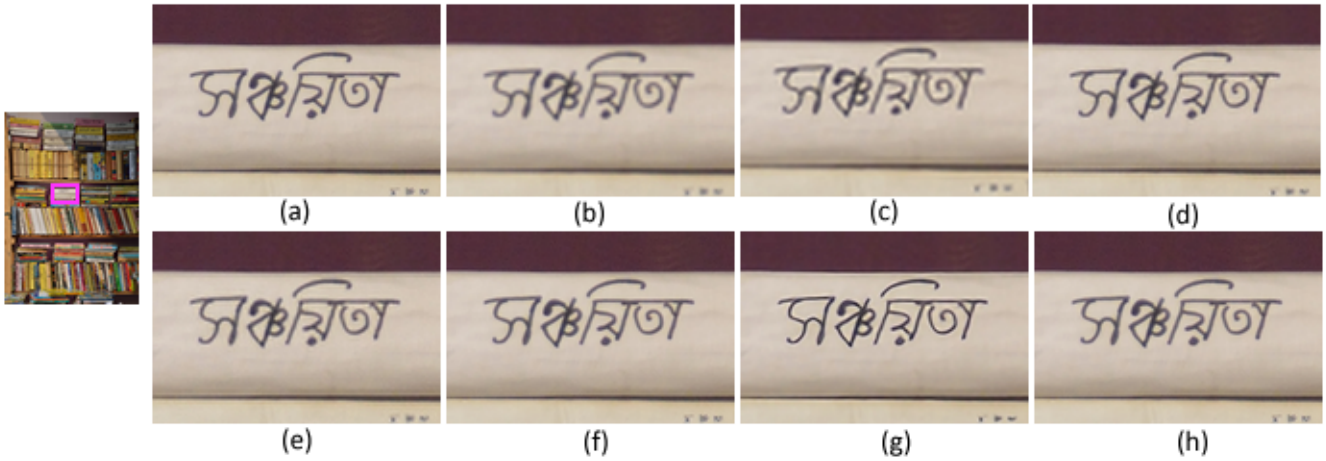

Fig. 11. The $1.5\times$ upsampled results of different methods. (a) The HR images, (b) bicubic, (c) ANR [13], (d) LPE [16], (e) SRCNN [17], (f) VDSR [18], (g) ResNet, (h) the proposed FAST.

computed as the mean score of all the viewers, and each score has been weighted by the KCC δ of the corresponding viewer. The weighted subjective scores are also shown in Fig. 10, and the proposed FAST can achieve visual quality comparable with that of these state-of-the-art methods.

C. Objective Testing

Table II shows the PSNR results of these methods for $1.5\times$ magnification. Note that another commonly used assessment, SSIM, is not listed in this paper because the SSIM values of different $1.5\times$ results are too similar. For example, the average SSIMs for bicubic, FAST, and VDSR are 0.9993, 0.9995, and 0.9996, respectively. By comparing the PSNR values in Table II, the following findings can be obtained. First, the learning-based method can achieve a much higher PSNR than the bicubic method. Second, the FAST and LPE methods obtain comparable PSNR values. Third, although the subjective quality is still similar to other DNN-based methods, ResNet clearly outperforms other methods.

Table III lists the average reconstruction times of different methods. The VDSR and ResNet methods are executed on an Intel Core i7 PC with NVIDIA Titan X GPUs and the MatConvNet package [34], while the other methods are implemented on an Intel Core i5 laptop PC with MATLAB. The computational time of the SISR method depends on the resolution of the input image. Although these methods are either a fast SR method or real-time method on GPU hardware, the time cost for 720P-to-1080P reconstruction is still high. It can be found that the proposed FAST is faster than the other methods because of its multiple acceleration strategies. Moreover, the proposed FAST still has much room for acceleration, such as implementation with C or CUDA.

Note that there are other ways for further accelerating the proposed method. For example, the filtering process [35] runs faster than local patch projection. However, it is more difficult

to solve the optimal filters from LR patches to HR patches. In the future, we tend to utilize the deep network as a solver rather than an end-to-end SR technique, so that the optimal filter or even filter banks can be calculated. The filtering process can further accelerate the local patch reconstruction.

TABLE II
THE PSNR RESULTS OF DIFFERENT METHODS (720P-TO-1080P)

	Bicubic	ANR [13]	LPE [16]	LPE_P_2 [16]	SRCNN [17]	VDSR [18]	ResNet	FAST
0803	45.17	46.83	48.47	48.45	49.17	49.21	**49.35**	47.96
0807	25.75	26.35	27.08	27.07	27.36	27.51	**27.76**	27.15
0808	32.63	33.27	34.24	34.23	34.49	34.67	**34.84**	34.24
0809	42.22	42.78	45.15	45.15	44.68	45.13	**45.21**	44.58
0817	39.11	39.65	40.93	40.92	40.86	41.14	**41.34**	40.75
0818	36.05	37.28	36.86	36.84	39.55	39.80	**39.90**	38.49
0820	31.55	32.77	34.75	34.76	35.12	35.41	**35.56**	34.10
0821	36.88	39.40	40.02	40.01	40.92	41.73	**41.85**	41.01
0833	40.22	41.20	43.59	43.59	43.27	43.83	**44.03**	42.64
0836	31.64	32.43	34.38	34.37	34.61	34.77	**34.95**	33.70
0839	35.19	35.66	36.64	36.63	36.78	36.92	**37.01**	36.59
0843	45.13	46.16	48.25	48.24	47.81	48.14	**48.26**	47.12
0844	49.26	50.68	52.11	52.13	52.02	52.20	**52.25**	51.15
0845	28.92	29.86	31.43	31.46	31.58	31.73	**31.94**	30.81
0855	35.63	36.04	37.19	37.17	37.16	37.42	**37.66**	37.12
0857	37.92	38.00	38.35	38.36	38.23	38.41	**38.76**	38.42
0863	39.45	39.85	41.14	41.12	41.09	41.27	**41.46**	40.97
0867	35.43	36.34	37.93	37.92	38.22	38.43	**38.65**	37.62
0872	30.52	31.43	32.70	32.70	33.00	33.19	**33.28**	32.64
0882	40.18	41.06	42.02	42.00	42.26	42.45	**42.65**	41.95
0896	43.76	44.80	45.52	45.51	45.66	46.01	**46.21**	45.51
Average	37.27	38.18	39.46	39.46	39.71	39.97	**40.14**	39.26

TABLE III
RECONSTRUCTION TIME OF DIFFERENT METHODS (720P-TO-1080P)

Method	ANR (CPU)	SRCNN (CPU)	LPE (CPU)	LPE_P_2 (CPU)	VDSR (GPU)	ResNet (GPU)	FAST (CPU)
Time(s)	348.4	192.0	227.9	229.2	64.2	78.3	**29.7**

V. CONCLUSIONS

Traditional SISR methods focus on $2\times$ or higher magnification factors, but lower magnification factors have not been thoroughly discussed. This paper proposed a fast and low-cost SISR method for special $1.5\times$ industrial applications. For reducing the computational complexity, the proposed method utilizes the following strategies. First, the edge and nonedge areas are reconstructed individually. The edge patterns are recovered by means of a learning-based method, and the nonedge areas are upsampled via a fast global backprojection method with an additional texture constraint. Second, finite edge patterns are encoded via a local encoding process, and these patterns are further manually selected according to their frequency of occurrence. Third, the local difference descriptor has been reused in several steps of the proposed method. The experimental results of 720P-to-1080P SR demonstrate that the proposed method can obtain satisfactory visual quality and has a considerably lower computational cost than the compared methods.

REFERENCES

[1] K. Nasrollahi, T. B. Moeslund, "Super-resolution: a comprehensive survey," *Mach. vision appl.*, vol. 25, no. 6, pp. 1423-1468, 2014.

[2] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, no. 4, pp. 324-355, 1993.

[3] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049-1075, Nov. 1999.

[4] J. Sun, Z. Xu, and H. Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1-8.

[5] Y. Zhao, R. Wang, W. Wang, W. Gao, "High resolution local structure-constrained image upsampling," *IEEE Trans. Image Process.*, vol.24, no.11, pp.4394-4407, Nov. 2015.

[6] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25-47, Jun. 2000.

[7] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, vol. 1, pp. 275-282.

[8] Y. Zhao, R. Wang, W. Wang, W. Gao, "Multilevel modified finite radon transform network for image upsampling," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 26, no. 12, pp. 2189-2199, 2016.

[9] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861-2873, Nov. 2010.

[10] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," *Curv. Surfaces*, pp. 711- 730, 2010.

[11] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang, "Bilevel sparse coding for coupled feature spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2360-2367.

[12] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 349-356.

[13] R. Timofte, V. D. Smet, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920-1927.

[14] R. Timofte, V. D. Smet, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian Conf. Comput. Vis.*, 2014, pp. 1-15.

[15] S. Schulter, C. Leistner, H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp.3791-3799.

[16] Y. Zhao, R. Wang, W. Jia, J. Yang, W. Wang , W. Gao, "Local patch encoding-based method for single image super-resolution," *Information Sciences*, vol.433, pp.292-305, 2018.

[17] C. Dong, C. Loy, K. He, X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.38, no.2, pp.295-307, Feb. 2016.

[18] J. Kim, J. Lee, and K, Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, vol. 1, pp. 1646-1654.

[19] C. Ledig, L. Theis, F. Huszar, et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *arXiv preprint*, arXiv:1609.04802, 2016.

[20] W. Shi, J. Caballer, F. Huszar, et al., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, vol. 1, pp. 1874-1883.

[21] R. Timofte, E. Agustsson, et al., "Ntire 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, vol. 1, pp. 1110-1121.

[22] B. Lim, S. Son, H. Kim, et al., "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, vol. 1.

[23] K. Zhang, W. Zuo, L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN based image denoising," *IEEE Trans. Image Process.*, 2018.

[24] W. Lai, J. Huang, N. Ahuja, et al., "Deep laplacian pyramid networks for fast and accurate superresolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol.2, no.3, 2017.

[25] M. Haris, G. Shakhnarovich, N. Ukita, "Deep backprojection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018.

[26] M. Sharma, R. Mukhopadhyay, A. Upadhyay, et al., "Irgun: Improved residue based gradual up-scaling network for single image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, vol.8.

[27] Y. Zhang, Y. Tian, Y. Kong, et al., "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018.

[28] J. Choi and M. Kim, "Super-Interpolation With Edge-Orientation-Based Mapping Kernels for Low Complex 2x Upscaling," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp.469-483, 2016.

[29] J. Choi and M. Kim, "Single image super-resolution using global regression based on multiple local linear mappings," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp.1300-1314, 2017.

[30] Y. Tang, H. Chen, Z. Liu, B. Song, Q. Wang, "Example-based super-resolution via social images," *Neurocomputing*, vol.172, pp.38-47, 2016.

[31] C. Yang, C. Ma, M. Yang, "Single-image super-resolution: a benchmark," in *Proc. Euro. Conf. Comput. Vis.*, Sep. 2014, pp. 372-386.

[32] D. C. Van Essen, C. H. Anderson, and D. J. Felleman, "Information processing in the primate visual system: An integrated systems perspective," *Science*, vol. 255, no. 5043, pp. 419-423, 1992.

[33] S. Siegel and N. Castellan, "Nonparametric statistics for the behavioral sciences," *McGrall-Hill International*, 1988.

[34] A. Vedaldi and K. Lenc, "MatConvNet - Convolutional Neural Networks for MATLAB," in *Proc. of the ACM Int. Conf. Multimedia*, 2015.

[35] Y. Romano, J. Isidoro, P. Milanfar, "RAISR: rapid and accurate image super resolution," *IEEE Trans. Comput. Imaging*, vol.3, no.1, pp.110-125, 2017.

[36] Y. Zhao, G. Li, W. Xie, W. Jia, H. Min, X. Liu, "GUN: Gradual Upsampling Network for Single Image Super-Resolution," *IEEE Access*, vol.6, no.1, pp.39363-39374, 2018.
[37] Y. Blau, T. Michaeli, "The Perception-Distortion Tradeoff," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018.
[38] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting Coherent Groups in Crowd Scenes by Multiview Clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, DOI: 10.1109/TPAMI.2018.2875002, 2018.
[39] S. Gu, L. Zhang, W. Zuo, X. Feng, "Projective dictionary pair learning for pattern classification" in *Proc. NIPS*, 2014, pp. 793-801.
[40] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical Feature Selection for Random Projection," *IEEE Trans. Neural Netw. Learn. Syst.*, DOI: 10.1109/TNNLS.2018.2868836, 2018.
[41] Y. Zhao, R. Wang, W. Jia, W. Wang, W. Gao, "Iterative projection reconstruction for fast and efficient image upsampling," *Neurocomput.*, vol. 226, pp. 200-211, 2017.
[42] Q. Wang, Z. Qin, F. Nie, and X. Li, "Spectral Embedded Adaptive Neighbors Clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, DOI: 10.1109/TNNLS.2018.2861209, 2018.
[43] Q. Wang, F. Zhang, and X. Li, "Optimal Clustering Framework for Hyperspectral Band Selection," *IEEE Transactions on Geosci. Remote Sens.*, DOI: 10.1109/TGRS.2018.2828161, 2018.

Wei Jia (M'12) received the B.Sc. degree in informatics from Central China Normal University, Wuhan, China, in 1998, the M.Sc. degree in computer science from Hefei University of Technology, Hefei, China, in 2004, and the Ph.D. degree in pattern recognition and intelligence system from University of Science and Technology of China, Hefei, China, in 2008. He has been a research assistant and associate professor in Hefei Institutes of Physical Science, Chinese Academy of Science from 2008 to 2016. He is currently a research associate professor in School of Computer and Information, Hefei University of Technology. His research interests include computer vision, biometrics, pattern recognition, image processing and machine learning.

Yang Zhao (M'16) received the B.E. and the Ph.D. degree from department of automation, University of Science and Technology of China, in 2008 and 2013. From September 2013, he is a Postdoctoral Fellow with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, China. He is currently a research associate professor in School of Computer and Information, Hefei University of Technology. His research interests include image processing and pattern recognition.

Ronggang Wang (M'12) received his Ph.D. Degree from the Institute of Computing Technology, Chinese Academy of Sciences. He was a Research stuff in Orange (France telecom) Labs from 2006 to 2010. He is currently a professor with Peking University Shenzhen Graduate School.

His research interest is on video coding and processing. He has done many technical contributions to ISO/IEC MPEG and China AVS. He leaded the MPEG Internet Video Coding (IVC) standard, served as MPEG IVC AHG co-chair since 2012, and severed as AVS implementation sub-group co-chair since 2015. He has authored more than 60 papers in international journals and conferences, and held more than 40 patents.

Hai Min received the Ph.D. degree in pattern recognition and intelligence system from the University of Science and Technology of China, Hefei, China, in 2014. He is currently an Assistant Professor in Department of Computer Science in Hefei University of Technology.His research interests include image processing and pattern recognition.

Shujie Li received the B.Sc. degree in information and computing science, Hefei University of Technology, China, in 2004, the M.SC. degree in computer science from Hefei University of Technology, Hefei,China, in 2008, and the Ph.D. degree in pattern recognition and intelligence system from University of Science and Technology of China, Hefei, China, in 2012. She is currently an assistant professor at the School of Computer and Information, Hefei University of Technology, China. Her research interests include human motion analysis and pattern recognition.

Xiaoping Liu received the M.Sc. and Ph.D. degree from department of computer science, Hefei University of Technology, China. He is currently a professor at the School of Computer and Information, Hefei University of Technology, China. His research interests include computer graphics and cooperative computing.